Lecture Notes in Artificial Intelligence 5928

Edited by R. Goebel, J. Siekmann, and W. Wahlster

Subseries of Lecture Notes in Computer Science

Ming Xie Youlun Xiong Caihua Xiong
Honghai Liu Zhencheng Hu (Eds.)

Intelligent Robotics and Applications

Second International Conference, ICIRA 2009
Singapore, December 16-18, 2009
Proceedings

Springer

Series Editors

Randy Goebel, University of Alberta, Edmonton, Canada
Jörg Siekmann, University of Saarland, Saarbrücken, Germany
Wolfgang Wahlster, DFKI and University of Saarland, Saarbrücken, Germany

Volume Editors

Ming Xie
Nanyang Technological University
Singapore 639798
E-mail: mmxie@ntu.edu.sg

Youlun Xiong
Caihua Xiong
Huazhong University of Science and Technology
Wuhan 430074, P. R. China
E-mail: {ylxiong, chxiong}@mail.hust.edu.cn

Honghai Liu
University of Portsmouth
PO1 3QL Portsmouth, UK
E-mail: honghai.liu@port.ac.uk

Zhencheng Hu
Kumamoto University
Kumamoto 860-8555, Japan
E-mail: hu@cs.kumamoto-u.ac.jp

Library of Congress Control Number: 2009940407

CR Subject Classification (1998): I.2.9, J.3, J.7, I.4, I.5, I.2.10, J.1

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN 0302-9743
ISBN-10 3-642-10816-4 Springer Berlin Heidelberg New York
ISBN-13 978-3-642-10816-7 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12814619 06/3180 5 4 3 2 1 0

Preface

The market demands for skills, knowledge and personalities have positioned robotics as an important field in both engineering and science. To meet these challenging demands, robotics has already seen its success in automating many industrial tasks in factories. And, a new era will come for us to see a greater success of robotics in non-industrial environments. In anticipating a wider deployment of intelligent and autonomous robots for tasks such as manufacturing, eldercare, homecare, edutainment, search and rescue, de-mining, surveillance, exploration, and security missions, it is necessary for us to push the frontier of robotics into a new dimension, in which motion and intelligence play equally important roles.

After the success of the inaugural conference, the purpose of the Second International Conference on Intelligent Robotics and Applications was to provide a venue where researchers, scientists, engineers and practitioners throughout the world could come together to present and discuss the latest achievement, future challenges and exciting applications of intelligent and autonomous robots. In particular, the emphasis of this year's conference was on "robot intelligence for achieving digital manufacturing and intelligent automations."

This volume of Springer's *Lecture Notes in Artificial Intelligence* and *Lecture Notes in Computer Science* contains accepted papers presented at ICIRA 2009, held in Singapore, December 16–18, 2009. On the basis of the reviews and recommendations by the international Program Committee members, we decided to accept 128 papers having technical novelty, out of 173 submissions received from different parts of the world.

This volume is organized into chapters covering: Ubiquitous and Cooperative Robots in Smart Space, Advanced Control on Autonomous Vehicles, Intelligent Vehicles: Perception for Safe Navigation, Novel Techniques for Collaborative Driver Support, Robot and Automation in Tunneling (973), Robot Mechanism and Design, Robot Motion Analysis, Robot Motion Control, Visual Perception by Robots, Computational Intelligence by Robots, Robots and Challenging Applications, Robots and Current Investigations. We hope that this volume is a useful source of information to inspire new initiatives and new opportunities of collaboration toward the advancement of intelligent robotics and challenging applications of robots in both industrial and non-industrial environments.

We would like to thank all the contributors and authors for their continuous support of the ICIRA conference series. Also, we would like to thank the international Program Committee for their generous efforts and critical reviews. Thanks also go to the keynote speakers, Sadayuki Tsugawa, Meijo University, Japan, Zhengyou Zhang, Microsoft

Research, USA, and Yongkiang Koh, ST Kinetics Ltd., Singapore, for providing very insightful talks. Finally, we would like to thank those who devoted their time and effort to make ICIRA2009 a successful scientific meeting.

October 2009

Ming Xie
Youlun Xiong
Caihua Xiong
Honghai Liu
Zhencheng Hu

Conference Organization

General Chairs

Ming Xie	Nanyang Technological University, Singapore
Youlun Xiong	Huazhong University of Science and Technology, China

Program Chairs

Caihua Xiong	Huazhong University of Science and Technology, China
Honghai Liu	University of Portsmouth, UK
Zhencheng Hu	Kumamoto University, Japan

Program Co-chairs

Ming Li	National Science Foundation of China, China
Guobiao Wang	National Science Foundation of China, China
Juergen Leopold	Fraunhofer Institute, Germany
Zhaohui Wu	Zhejiang University, China

Publicity Chairs

Haoyong Yu	DSO National Laboratories, Singapore
Zhaowei Zhong	Nanyang Technological University, Singapore
Shiqi Li	Huazhong University of Science and Technology, China
Yongan Huang	Huazhong University of Science and Technology, China

International Advisory Committee

Manuel Armada, Spain
Minoru Asada, Japan
Hegao Cai, China
Rudiger Dillmann, Germany
Jean-Guy Fontaine, Italy
Christian Freksa, Germany
Toshio Fukuda, Japan
Norihiro Hagita, Japan
Huosheng Hu, UK
Oussama Khatib, USA
Mark Lee, UK
Peter Luh, USA
Grigory Panovko, Russia
Shigeki Sugano, Japan
Susumu Tachi, Japan
Atsuo Takanishi, Japan
Gurvinder Virk, New Zealand
Miomir Vukobratovic, Serbia
Michael Wang, China
Tianmiao Wang, China
John Weng, USA
Lotfi Zadeh, USA

International Program Committee

Abdel Aitouche, France
Kaspar Althoefer, UK
Marcelo Ang, Singapore
Mario R.A. Saavedra, Spain
H. Levent Arkin, Turkey
Philippe Bidaud, France
Roland Chapuis, France
Hui Chen, China
Jean-Daniel Dessimoz, Switzerland
Hua Deng, China
Han Ding, China
Frank Dylla, Germany
Gary Feng, China
Luigi Fortuna, Italy
Yili Fu, China
Dongbing Gu, UK
Yongqiang Guo, Singapore
Loulin Huang, New Zealand
Makoto Itoh, Japan
Yunde Jia, China
Rolf Johansson, Sweden
Shin Kato, Japan
Kazuhiko Kawamura, USA
Naoyuki Kubota, Japan
Duanling Li, China
Yangmin Li, China
Youfu Li, China
Kianguan Lim, Singapore
Yonghuai Liu, UK
Benny Lo, UK
Chu Kiong Loo, Malaysia
Jiangzhou Lu, Singapore
Qinggang Meng, UK
Rezia Molfino, Italy
Sergiu Nedevschi, Romania
Aipeng New, Singapore
Lilian Ngiam, Singapore
Rainer Palm, Germany
Jee-Hwan Ryu, South Korea
Andy Russell, Australia
Jon Selig, UK
Yandong Tang, China
Yantao Tian, China
Diedrich Wolter, Germany
Yugeng Xi, China
Erfu Yang, UK
Zhijun Yang, UK
Zhouping Yin, China
Changjiu Zhou, Singapore
Limin Zhu, China

Table of Contents

Ubiquitous and Cooperative Robots in Smart Space

Advanced Control on Autonomous Vehicles

Intelligent Vehicles: Perception for Safe Navigation

Novel Techniques for Collaborative Driver Support

Robot and Automation in Tunneling (973)

Robot Mechanism and Design

Robot Motion Analysis

Robot Motion Control

Visual Perception by Robots

Computational Intelligence by Robots

Robot and Application

Robot and Investigation

A Ubiquitous and Cooperative Service Framework for Network Robot System

Yabo Liu[1] and Jianhua Yang[1,2,*]

[1] College of Computer Science and Technology, Zhejiang University,
Hangzhou 310027, Zhejiang, P.R. China
[2] The Sci-Tech Academy, Zhejiang University,
Hangzhou 310027, Zhejiang, P.R. China
{yabozj,jhyang}@zju.edu.cn

Abstract. Network robot system (NRS) is a new concept that integrates physical autonomous robots, environmental sensors and human-robot interactions through network-based cooperation. The aim of this paper is to provide a ubiquitous and cooperative service framework for NRS. We first present foundational concepts of semantic map and service abstraction for the framework. Then, in order to generate feasible service configurations to fulfill tasks, we propose service configuration and reconfiguration algorithms, which dynamically search the appropriate service configurations for each task. Additionally, we put forward a service reasoning an enabling process to tackle service unavailable problems. The service configuration's cost evaluation function is also proposed to facilitate the choosing against different service configurations. In implementation, we have developed a prototype system comprised of two robots and other various devices. Experiments indicate that the versatile service framework provides self-adaptive capabilities and utilizes available resources efficiently under a range of different scenarios.

Keywords: Network robot system, multi-robot cooperation, ubiquitous robot.

1 Introduction

It seems quite likely that robots, which have been charged with increasingly varied and difficult tasks, can be scattered everywhere and become a ubiquitous part of our day-to-day lives in the future [3]. Multi-robot systems (MRS) [1] are considered to be more robust and more effective than single robot systems. However, MRS cannot accomplish complex tasks solely sometimes due to the limitations of their own onboard sensors and actuators.

Network robot system (NRS) is a new concept that integrates physical autonomous robots, environmental sensors and human-robot interactions through network-based cooperation [11]. NRS can be envisioned as a unified, ubiquitous and cooperative

* Corresponding author.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 1–10, 2009.

network comprised of sensors, actuators, and embedded devices, etc. Hence, the robots' perceptual and actuation capabilities in NRS are substantially enhanced by utilizing environment equipments as their own onboard ones.

Furthermore, there is a growing trend to introduce high-level semantic knowledge into robotic systems to improve their task solving capabilities [4]. Although several facets of the problem of endowing a single robot with the capability of acquiring and using semantic knowledge have been tackled [2, 7], we will focus on using semantic knowledge in NRS to enhance its ubiquitous and cooperative characteristics.

Smart space provides a collaborative infrastructure for resources and devices, and is a promising route to bring NRS into effect. SCUDWare [13] attempts to present a semantic and adaptive middleware platform for smart space through synthesized techniques of multi-agent, context aware, and adaptive component management. A context quality control and management infrastructure [12] is also proposed to provide ubiquitous application with different context criterions and trust-worthiness.

In this paper, we extend the former work to introduce a ubiquitous and cooperative service framework for NRS, which maps available resources in NRS into a dynamic virtual service space. The virtual space will follow the robot to provide task-oriented services, just like its *shadow* in the physical world. By exploiting ambient resources, namely sensors and effectors in its environment, a robot can perform more complex tasks without becoming excessively complex itself. Although the topic of user-centric services [14] is also an interesting problem but outside the scope of this paper.

The rest of this paper is structured as follows. In the next section, we discuss related works. Section 3 briefly introduces foundational concepts of semantic map and service abstraction. Section 4 presents a detailed discussion on the services cooperation. An experimental evaluation of the approach is given in section 5, and we conclude with a summary and outlook in section 6.

2 Related Works

The core of NRS can be in some sense considered as building a virtual robot on the fly by combining software and hardware components residing on different devices, in order to collectively perform a given task in a given situation [5].

Nakamura *et al.* extended a framework for network robot platform (NWR-PF) to take account of both present and past status information, and proposed a service allocation method [8]. However, the level of granularity of services is very coarse-grained, namely a robot corresponding to a service.

Lundh *et al.* proposed a configuration generation algorithm to allocate and connect functionalities among the robots [5]. Nevertheless, they generated a service configuration through expanding hand-coded functionality operators and method schemas instead of dynamic searching and composing, and they did not take concurrent task performance into account. Involving auxiliary services to change the current environment status for enabling crucial service is also not considered.

A service robot that operates autonomously in a sensor-equipped kitchen can perform sophisticated tasks in concert with the network of devices in its environment [10]. The major contributions of this AwareKitchen are interfacing ubiquitous sensing

devices and supporting model learning, while any high-level service reasoning and configuration are not considered.

Semantic knowledge is useful in robotics in several ways [4], among which semantic maps offers many advantages in several aspects. Semantic maps can be built in three phases [7]: scene interpretation, object detection and interpretation, and integration and visualization. Consequently, these semantic maps can improve task planning capability and efficiency [2]. However, we focus on the using of semantic knowledge in NRS to enhance its ubiquitous and cooperative characteristics in this paper.

3 Foundation

3.1 The Semantic Map

A semantic map is a map that contains spatial information about the environment augmented with semantic domain knowledge. The system presented in this paper abstracts away from the automatic acquisition of semantic maps [7] to focus on how it can be used in NRS.

What we concern are places and objects, and their relationships in semantic maps. We employ ontology to provide an explicit conceptual model of the entities in the domain. In practice, the PowerLoom knowledge representation system [9] is used to present and reason semantic knowledge.

3.2 Service

Sensors, actuators, algorithms and other softwares are all considered as resources and abstracted as services. Let *N* be a network system, *E* be the set of entities in *N*, and *S* be set of services that exists in *N*.

Definition 1. A service instance S_i is specified as

$$S_i = \{e^{S_i}, I^{S_i}, O^{S_i}, L^{S_i}, W^{S_i}, Con^{S_i}, Pre^{S_i}, Post^{S_i}, Cost^{S_i}\} \tag{1}$$

- Each instance of a service resides in a specific physical entity $e^{S_i} \in E$.
- $I^{S_i} \subset \Omega$ denotes the set of inputs of S_i, and Ω is the domain of interfaces.
- $O^{S_i} \subset \Omega$ denotes the set of outputs of S_i, and Ω is the domain of interfaces.
- $L^{S_i} = L(e^{S_i})$ denotes the current location of S_i, $L: E \to P$, where P is the places set in semantic map.
- $W^{S_i} = W(e^{S_i})$ denotes where the service can appear, $W: E \to 2^P$.
- $Con^{S_i} \in \Re^+$ denotes the amount of different service configurations in which S_i can be concurrently used.
- Pre^{S_i} denotes the pre-conditions that must be satisfied.
- $Post^{S_i}$ denotes the post-effects that happen.

- $Cost^{S_i}$ denotes the cost of using the service, and a detailed description is given in the following section.

The inputs and outputs of services are in the same domain. For example, a SICK laser range-finder service provides a *RangeScanner2d* interface, which can be fed into a local navigation service.

The pre-conditions and post-effects are status (i.e., propositions) that should be true or false, and PowerLoom provides two primitives to assert and query knowledge: `assert` and `ask`. Taking a clean service for example, it requires that the cleaning robot be in the room to be cleaned, and the post-effect of this service is a clean room.

4 Services Cooperation

4.1 Dynamic Service Configuration

A complex task (or mission) in the environment is modeled as a time-constrained AND/OR task tree [15]: the root is the most abstract description, and the leaves contain basic tasks (or actions) that can be executed directly. However, due to space limitations, the details of task modeling and planning techniques are not included in this paper. We concern how a task can be fulfilled by a set of services in a cooperative manner in this section.

A task is carried out by a service configuration, which comprises a set of services and a graph which specifies how the services are connected. We use a dynamic programming idea to obtain a service configuration instead of process algebra such as Pi-calculus and Petri net on the grounds that a simple algorithm is sufficient here.

Algorithm 1. Service configuration

1. Exclude the services that are not in the place where a task t is to be carried out.
   $\prod = S \setminus \{s \,|\, L^s \neq L(t), s \in S\}$
2. Find a service configuration with the lowest cost constructed from $\prod$ to complete the current task. If found, go to step 5; else, continue.
3. Include the services that can appear in the current place.
   $\prod' = \prod \cup \{s \mid L(t) \in W^s, s \in S\}$
4. Find a service configuration with the lowest cost constructed from $\prod'$ to complete the current tasks. If found, recruit the required services to the desired places and go to step 5; else, report failure.
5. Carry out the task according to the service configuration.

4.2 Service Reconfiguration on Service Conflicts

In general, several tasks might be performed concurrently, and new tasks might emerge dynamically. With such an extension, issues such as service conflicts must also be considered. There are two factors that may cause service conflicts: 1) services required by

the new configuration are currently in use in existing ones; 2) the post-effects of the services required by the new configuration violate the pre-condition of existing ones.

Algorithm 1 reports failure when it cannot find a feasible service configuration. However, re-configurating existing service configurations to release some services when the failure is due to service conflicts would be a better solution. For easy notations, let C be a set of potential service configurations for the new task, and $\mathbb{C}$ be the set of existing service configurations. A greedy algorithm for service reconfiguration is given in Algorithm 2. Additionally, actions with a high priority can suspend other actions with low priorities in order to release crucial services when necessary. The algorithm can also be employed to deal with service failures.

Algorithm 2. Service reconfiguration

1. For each $c \in C$, let $\mathbb{C}' \subset \mathbb{C}$ be the set of service configurations that have conflict with c;
2. For each $c' \in \mathbb{C}'$, let c'' be the feasible alternative service configuration with the lowest cost for c';
3. The feasible service configuration for the current new task is given by $\arg\min_{c \in C} \left(Cost(c) + \sum_{c' \in \mathbb{C}'} \left(Cost(c'') - Cost(c') \right) \right)$;
4. Update corresponding service configurations, and carry out the task according to the latest service configurations.

4.3 Service Reasoning and Enabling

It happens that under some specific environment status, some services are disabled. For example, a service that depends on a camera may behave abnormally when the light condition is very poor, so the camera mounted on the ceiling for object tracking fails at night with the lights turned off; a service that depends on a microphone may also behave abnormally when it's very noisy in the environment, so a microphone mounted on a robot for human-robot interaction fails when the background music is so loud.

Apparently, a light service and a camera service are "*independent*" services with regard to their inputs and outputs. As a result, this service unavailable problem cannot be tackled with service configuration and/or service reconfiguration algorithms. Hence, we introduce Algorithm 3 to cover the shortage.

When preconditions of crucial services of a new service configuration are not satisfied, auxiliary services are picked up to change the current status. However, the issue must be considered is the change of the current status should not disturb other current running services and violate normative constraints. For instance, the light can be turned on to enable a camera service as usual, but it is inappropriate to turn the light on when a person is sleeping in the room.

We can insert conventional rules into the knowledge system, and check whether the current status change violates the rules. For instance,

```
(ask (and (exists ((?x human) (?y bed)) (and (in ?y
(place light1)) (inbed ?x ?y)))(not (exists ((?z
light)) (and (in ?z (place light1)) (on ?z ))))))
```

returns `TRUE` when somebody is in bed and all the lights are turned off in a room where `light1` is located in, and the retrieved result prevents `light1` from being turned on (see Fig. 1).

Algorithm 3. Service reasoning and enabling

1. Find candidate auxiliary services that can enable the current crucial service, marked as S_{cand}.
2. For each candidate auxiliary service $s \in S_{cand}$, check whether it violates the semantic constraints, and mark all the permitted services as S'_{cand}.
3. For each permitted candidate auxiliary service $s' \in S'_{cand}$, pickup the service with the lowest cost.
4. Carry out the task according to the current service configurations.

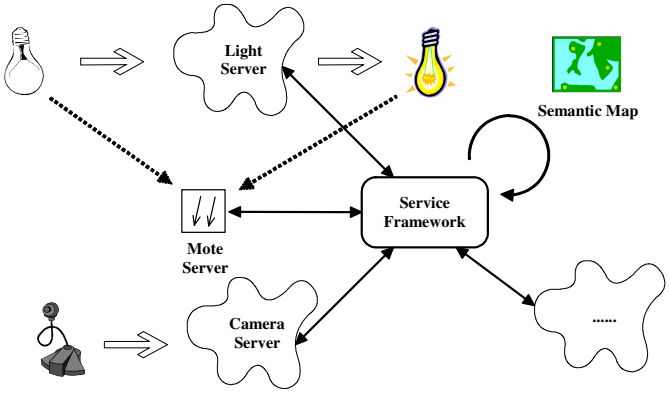


Fig. 1. Service reasoning and enabling

4.4 The Cost of Service Configuration

A generalized form of service configuration's cost evaluation function can be written as

$$C = \sum_1 + \sum_2 + \sum_3 \tag{2}$$

where $\sum_1$, $\sum_2$ and $\sum_3$ are the functions evaluating the costs to enable services, connect services and use services respectively during a service configuration.

The cost required to enable the services in the service configuration (the $\sum_1$ sub-function of Eq. (2)) can be broken down as follows:

$$\sum_1 = a\sum_W + b\sum_P \tag{3}$$

where $\sum_W$ and $\sum_P$ denote the total costs to change the place of service and change the environment status respectively, and a and b are scalar weighting factors. The values of the factors are generally set so that $b > a > 0$, since to change the place is better than to change the status.

The $\sum_2$ sub-function of Eq. (2) evaluating the cost to connect services can be expanded as

$$\sum_2 = \alpha\sum_N + \beta\sum_E \tag{4}$$

where $\sum_N$ and $\sum_E$ are the total numbers of services involved in the service configuration and corresponding entities respectively, and α and β are scalar weighting factors. The values of α and β are generally set to be smaller than a and b, and $\beta > \alpha > 0$ so there is a selective pressure against overmuch participators.

The $\sum_3$ sub-function of Eq. (2) becomes

$$\sum_3 = \delta\sum_C \tag{5}$$

where $\sum_C$ denotes the total cost to use the services in the service configuration and δ is a scalar weighting factor. The cost of a service is actually a balance on quality and expense, which may vary under different status, and the details are not included in this paper due to space limitations.

5 Experiments

5.1 Implementation

A prototype system is implemented and tested based on a component-based software platform [6]. The services are all encapsulated as components, and we manage the services according to algorithms we have proposed.

The experiment setup involves two robots and other various devices, as depicted in Fig. 2. The P3-AT robot is equipped with the SICK LMS200 laser rangefinder, Canon VC-C50i camera, front and rear sonar, compasses and tilt-position sensors; the CCNT robot is equipped with a Logitech QuickCam Pro 4000 camera, infrared distance sensors, compasses, and Crossbow Cricket Mote.

A Canon VC-C50i camera is mounted on the ceiling for object tracking. Additionally, several Cricket beacons are laid out on the ceiling to implement a wireless location system for the entities equipped with a Cricket listener (i.e., a CCNT robot). A number of MICA2 and MICA2DOT Motes are deployed in the environment to construct a wireless measurement system, and a Stargate NetBridge is used to serve as a sensor network gateway.

We use BSWA MPA 416 array microphones, NI 9233 analog input modules and an NI USB-9162 USB single module carrier to construct an acoustic array for source

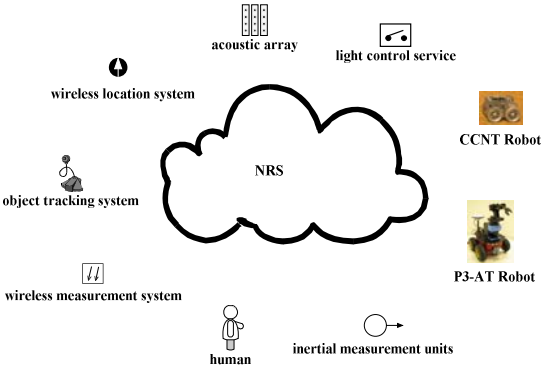


Fig. 2. A prototype system of NRS

detection, localization, and tracking. The person in the environment is equipped with small inertial measurement units (XSens MTx) that provide detailed information about the person's motions.

A wide variety of different wirelessly enabled nodes, ranging from notebooks (Lenovo ThinkPad T61 and X61s) to UMPCs (Fujistu LifeBook U1010) and PDAs (HP iPAQ hx2700), are used to provide access on the network to some of the aforementioned ubiquitous sensing devices.

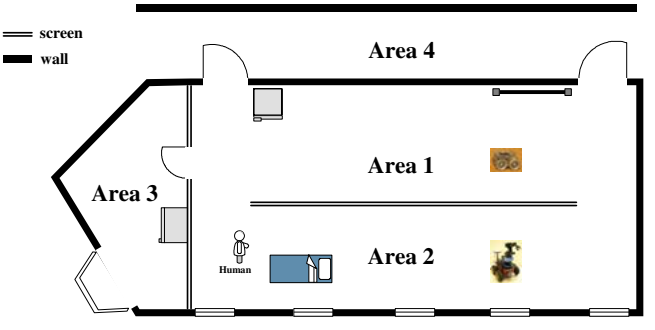


Fig. 3. A sketch of experimental environment in CGB Building

We demonstrate the prototype system in our laboratory room 526 at CGB Building, as shown in Fig. 3. The test environment consists of four areas, namely Areas 1~4. Areas 1~3 are divided by screens, Area 3 has a balcony, and Area 4 is a corridor. Area 1 and Area 3 are connected with a small door that can be only passed through by the CCNT robot. The object tracking system, wireless location system, wireless measurement system and computer-controlled lighting system cover Areas 1~3; and the acoustic array covers Area 3.

5.2 Results

We conducted Experiments 1~5 using the proposed service framework. Experiment 1 focused on the dynamic service configuration. Experiments 2~3 were conducted to evaluate the service reconfiguration under different status. Experiments 4~5 assessed the service reasoning and enabling abilities of the framework.

In *Experiment* 1, the acoustic array detects a suspicious sound in Area 3, issues a task to check Area 3 for security. After service configuration, the CCNT robot is chosen to head for the suspicious area due to the fact that P3-AT robot is too big to pass through the door which connects Area1 and Area 3, and object tracking system helps CCNT robot to locate itself for the convenience of navigation.

Experiment 2 shares the same scenario with Experiment 1 except that it's at night with the lights turned off. The wireless measurement system reports the poor light condition to the semantic map, and the object tracking system's precondition is not satisfied. Without any manual adjustments, after service configuration, the wireless location system is chosen to substitute for the object tracking system.

Experiment 3 shares the same scenario with Experiment 2 except that the Cricket listener on the CCNT robot cannot receive signals from the wireless location system due to noise and signal interference. Without any manual adjustments, after service configuration, the P3-AT robot is chosen to use its odometer, camera and laser range-finder to provide a location service to the CCNT robot.

Experiment 4 shares the same scenario with Experiment 3 except that P3-AT robot is busy on other tasks. After service reasoning, the system involves the light service to turn on the light. The wireless measurement system detects that the light condition changes and reports it to the semantic map. Consequently, the object tracking system's precondition is satisfied, and helps the CCNT robot to locate itself for the convenience of navigation.

Experiment 5 shares the same scenario with Experiment 4 except that a person is sleeping in Area 2. After service reasoning, the system finds no solution to carry out the task under the current restrictions. Since the task is very critical, and P3-AT's task is suspended, the P3-AT robot is chosen to use its odometer, camera and laser range-finder to provide a location service to the CCNT robot.

Through the experiments, we found that the versatile service framework provides self-adaptive capabilities and utilizes available resources to the utmost under various different scenarios.

6 Conclusions

In this paper, we consider a network-based cooperation problem for NRS. A ubiquitous and cooperative service framework has been proposed to enable a group of artificial autonomous entities to make an important use of communication and cooperation through a network in order to fulfill their tasks. The framework is validated by using a series of experiments in a prototype system comprised of two robots and other devices. For the future work, we will enhance our system and complete the remaining

parts of the framework (e.g., the security manager, service execution monitor, and fault recovery) and apply it to more complex scenarios.

Acknowledgments

This work was supported by the Astronautical Supporting Technology Foundation of China (Grant No. 08-3.4), the Ph.D. Programs Foundation of Ministry of Education of China (Grant No. 200803351129), the Zhejiang Provincial Natural Science Foundation of China (Grant No. Y1090449), and the Defense Advanced Research Foundation of the General Armaments Department of the PLA (Grants No. 9140A06060307JW0403, 9140A06050208JW0414, 9140A06050509JW0401, 9140A16070409JW0403, 9140A06050609JW0402).

References

1. Arai, T., Pagello, E., Parker, L.E.: Editorial: advances in multirobot systems. IEEE T. Robotic. Autom. 18, 661–665 (2002)
2. Galindo, C., Fernández-Madrigal, J., González, J., Saffiotti, A.: Robot task planning using semantic maps. Robot. Auton. Syst. 56, 955–966 (2008)
3. Gates, B.: A robot in every home. Sci. Am. 296, 58–65 (2007)
4. Hertzberg, J., Saffiotti, A.: Using semantic knowledge in robotics. Robot. Auton. Syst. 56, 875–877 (2008)
5. Lundh, R., Karlsson, L., Saffiotti, A.: Autonomous functional configuration of a network robot system. Robot. Auton. Syst. 56, 819–830 (2008)
6. Makarenko, A., Brooks, A., Kaupp, T.: Orca: Components for robotics. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Workshop on Robotic Standardization, pp. 163–168 (2006)
7. Nüchter, A., Hertzberg, J.: Towards semantic maps for mobile robots. Robot. Auton. Syst. 56, 915–926 (2008)
8. Nakamura, Y., Machino, T., Motegi, M., Iwata, Y., Miyamoto, T., Iwaki, S., Muto, S., Shimokura, K.: Framework and service allocation for network robot platform and execution of interdependent services. Robot. Auton. Syst. 56, 831–843 (2008)
9. Preece, A., Flett, A., Sleeman, D., Curry, D., Meany, N., Perry, P.: Better knowledge management through knowledge engineering. IEEE Intell. Syst. 16, 36–43 (2001)
10. Rusu, R., Gerkey, B., Beetz, M.: Robots in the kitchen: Exploiting ubiquitous sensing and actuation. Robot. Auton. Syst. 56, 844–856 (2008)
11. Sanfeliu, A., Hagita, N., Saffiotti, A.: Network robot systems. Robot. Auton. Syst. 56, 793–797 (2008)
12. Tang, S., Yang, J., Wu, Z.: A context quality management infrastructure for complex ubiquitous environment. Comput. Syst. Sci. Eng. 24 (2009)
13. Wu, Z., Wu, Q., Cheng, H., Pan, G., Zhao, M., Sun, J.: ScudWare: a semantic and adaptive middleware platform for smart vehicle space. IEEE T. Intell. Transp. 8, 121–132 (2007)
14. Zhang, L., Pan, G., Wu, Z., Li, S., Wang, C.-l.: SmartShadow: modeling a user-centric mobile virtual space. In: IEEE International Conference on Pervasive Computing and Communications (PerCom), pp. 1–4. IEEE Press, Galveston (2009)
15. Zlot, R., Stentz, A.: Market-based multirobot coordination for complex tasks. Int. J. Robot. Res. 25, 73–101 (2006)

The Hand-Bot, a Robot Design for Simultaneous Climbing and Manipulation

Michael Bonani, Stéphane Magnenat, Philippe Rétornaz, and Francesco Mondada

EPFL-LSRO, Station 9, CH-1015 Lausanne
michael.bonani@epfl.ch
http://mobots.epfl.ch/

Abstract. We present a novel approach to mobile object manipulation for service in indoor environments. Current research in service robotics focus on single robots able to move, manipulate objects, and transport them to various locations. Our approach differs by taking a collective robotics perspective: different types of small robots perform different tasks and exploit complementarity by collaborating together. We propose a robot design to solve one of these tasks: climbing vertical structures and manipulating objects. Our robot embeds two manipulators that can grasp both objects or structures. To help climbing, it uses a rope to compensate for the gravity force. This allows it to free one of its manipulators to interact with an object while the other grasps a part of a structure for stabilization. Our robot can launch and retrieve the rope autonomously, allowing multiple ascents. We show the design and the implementation of our robot and demonstrate the successful autonomous retrieval of a book from a shelf.

1 Introduction

Service robotics has a large economic potential [1]. In its applications, if robustness and flexibility are dominant requirements, collective robotics is a promising paradigm [2,3]. Because collective robotics distributes the work to multiple robots, it requires different control, more hardware units, and usually more engineering resources than the approaches based on single robots [4]. Yet the physical redundancy between the robots enables a high availability. Moreover, their modularity provides adaptation to changing needs and situations. These unique features might outstrip the drawbacks of collective robotics eventually.

To implement service robotics tasks successfully, groups of robots must be able to manipulate objects in the environment. These objects can be located in altitude, and robots part of a collective are small. Therefore, these robots must be able to climb the environment to fetch the objects. This paper proposes a design concept and demonstrates a working implementation of a climbing robot that performs manipulation tasks.

In a heterogeneous group, as proposed by the Swarmanoid project [5], our robot is only responsible for manipulation of objects in altitude. While on the

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 11–22, 2009.

floor, it relies on other types of robots for mobility. These robots would have different hardware, in particular, they would be able to self-assemble with our robot and move it on the ground [6]. Such task and hardware specialization allows the best use of the capabilities of each type of robots while self-assembling provides the synergy to compensate for the limitations of each individual type.

2 State of the Art

The ascension of vertical surfaces is a difficult problem: it requires a lightweight robot with strong actuators, which results in a sensitive mechatronic design. For this reason, few systems combine the abilities of climbing and manipulation. Researchers in the field of space manipulators —where robots move in microgravity— have already studied this combination [7].

Some robots require an adaptation of the environment, such as in [8] where the robot needs custom-tailored docks to attach. To remove this limitation, several works have explored vacuum adhesion. For instance, [9] proposes a medium-size robot (diameter of 50 cm) for inspection of large surfaces. A vacuum system provides the adhesion force while three wheels allow the robot to move on the surface. To scan the surroundings of the robot, an inspection sensor is attached to a parallel arm. This system does not allow manipulation of external objects.

The Alicia^3 [10] is a larger robot (length of 1.3 m, weight of 20 kg) which also uses pressure for adhesion. It consists of three attachment units linked by an articulated arm. Each unit contains its own vacuum generator and displacement wheels. Although the design of this robots is focused on climbing, its arm-based structure could support manipulation capabilities.

Several authors have proposed large platforms for outdoor operations on buildings or industrial superstructures [11], often for specific applications [12]. These platforms could support the addition of manipulators; however their size makes them unsuitable for the type of application we consider.

There exist several robots that utilize some form of grasping to climb [13,14,15]. Their grippers are custom-tailored to specific structures. Yet none of these robots can use their grippers to perform object manipulation.

To the best of our knowledge there is no small autonomous robot (few decimeters cube) that implements a combination of climbing and object manipulation.

3 Design and Implementation

We propose a novel approach to mobile object manipulation for service in indoor environments. It takes advantage of synergies between climbing and object manipulation. This translates as the ability to climb common vertical offices structures such as shelves. Within these structures, the robot gets small objects such as lightweight books or compact discs. It then retrieves these objects to the ground and brings them to a specific location. We have implemented this approach in the *hand-bot* (Fig. 1) robot that we present in this article.

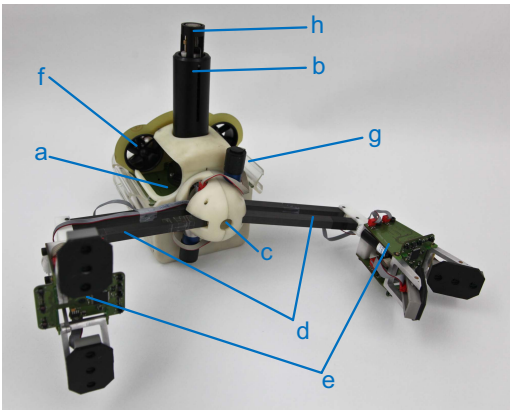
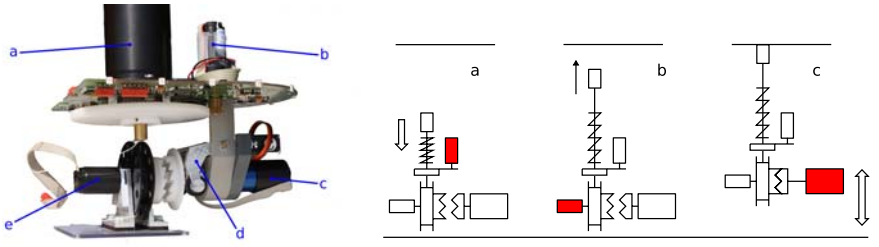


Fig. 1. Overview of the hand-bot. **a.** The body, containing in particular the battery, the rope launcher, and the head rotation motor. **b.** The tube of the rope launcher. **c.** The head, containing the two arms bending motors. **d.** The arms. **e.** The grippers, each containing two motors for rotation and grasping. **f.** Two fan to control the yaw when hand-bot hangs at the rope. **g.** Connection ring. **h.** Switchable magnet.

The biggest constraint for climbing under gravity is to provide the vertical lift force; we thus implement climbing by combining two techniques. Rolling a rope provides the vertical lift force while manipulators provide horizontal operations. The hand-bot fixes the rope to a ceiling and coils it around a reel. This mechanism is simple and can lift a large mass by using a strong motor; as shown by previous work [16]. To be autonomous, the robot must be able to attach its rope to a specific location, use it to climb, and retrieve it afterwards. We propose a launching mechanism based on a strong spring (Fig. 1b) that projects a magnet to the ceiling. Albeit this approach works only in environments with a ferromagnetic ceiling, it is well understood and reliable. Moreover, depending on the type of ceiling, other attachment mechanisms such as plungers are applicable. To be able to detach from the ceiling, the attachment must be switchable. We implement this feature using a magnetic switch that the robot can trigger with a specific infrared transmission.

When attached to the ceiling using the rope, the hand-bot has vertical mobility but is horizontally unstable. To stabilize and position itself on the horizontal plane, it needs manipulators to grasp the structure around it. The hand-bot has two arms, each with a gripper as manipulator (Fig. 1d,e). When using the two arms to climb, it maintains its stability all the time. It can also manipulate an object with one gripper while keeping the other one attached to the structure. That way, the hand-bot can manipulate objects precisely. Once on the ground, other types of robots can assemble with the hand-bot and displace it and the object it carries to a specific location. The other robots [5] attach to the hand-bot by grasping its translucent ring, which also contains 12 RGB LEDs.



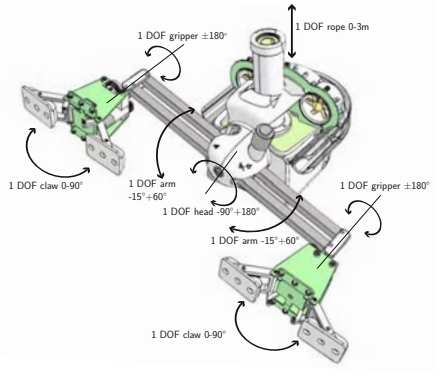
(a) rope launcher (b) rope launching sequence

Fig. 2. (A) Mechanism of the rope launcher. **a.** The launching tube containing a spring. **b.** The motor to compress the spring. **c.** The lifting motor. **d.** The servomotor clutch to engage the lifting motor. **e.** The motor directly connected to the reel, to coil the rope on magnet retrieval and to provide tension control. (B) Schematics of the launching sequence, filled rectangles show the active motor. **a.** A motor compresses the spring. **b.** A fast motor maintains tension in the rope while the spring launches the magnet. **c.** A clutch is engaged to let the strong motor lift the robot.

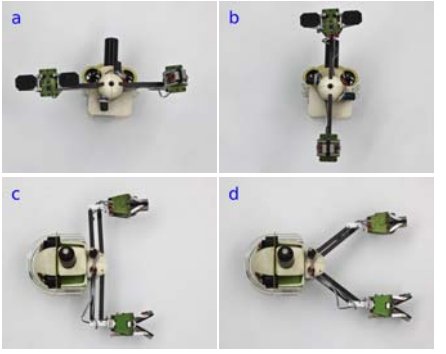
3.1 Rope Launcher

As we explained in the preceding sections, the rope provides the main lifting force of the hand-bot. Fig. 2A shows a photo of the mechanism of the rope launcher and Fig. 2B shows the schematics of the launching sequence. To launch the rope, a motor compresses a spring using a small wagon. This wagon moves inside the launch tube and is driven by a worm gear. When the spring is fully compressed (at that point, it applies a force of 110 N), the wagon hits the bottom of the tube and liberates the spring, which launches the rope, the magnet, and the detachment mechanism up to an altitude of 270 centimeters. Two different motors drive the rope: a strong one provides the main lifting force and a fast one controls the tension of the rope, during launch and retrieval. To switch between these two motors, a servomotor activates a clutch. During launch, the fast motor brakes the reel: this is necessary to prevent the formation of knots in the rope. The launching control program monitors the length of the uncoiled rope in real time. As soon as the magnet reaches the target altitude, the fast motor firmly coils back, which ensures the tension in the rope. If the magnet fails to attach, this action will coil back most of the rope and the hand-bot will know the result of the launch. If the launch succeeds, the servomotor engages the clutch so that the strong motor drives the rope, and provides the main lifting force for the hand-bot. This force is strong enough to lift the robot by itself.

If the hand-bot hangs freely at the rope, because it holds an object or has failed to grasp an element of structure, it can stabilize and orientate itself using an inertial measurement unit and two fans. The fans are located at the top of the body on both side of the rope launcher tube.



(a) kinematic of the hand-bot.



(b) extreme head and arms positions

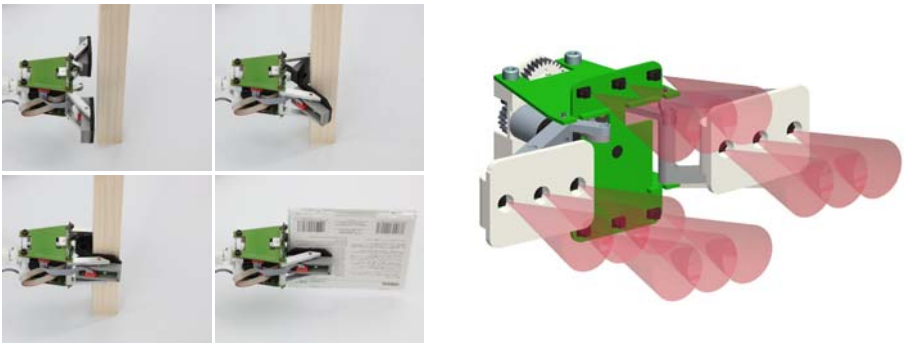
Fig. 3. The degrees of freedom of the hand-bot

3.2 Manipulators

The hand-bot has two arms on its front side. They can bend/extend forward — independently — and rotate with respect to the robot body (Fig. 3). At the end of each arm, the hand-bot has a gripper (Fig. 4A). Each gripper can rotate with respect to its arm and can open and close its claws. To grasp objects and structures of different thicknesses, the gripper claws have a parallel compliance mechanism. When no object is present, a spring maintains a large opening angle between the claws. Once the claws squeeze an object, the points of contact are different than the points of rotation which generates a moment that aligns the claws in parallel with the object. This provides a strong force over a large range of thicknesses. To control grasping, each gripper can detect structures and objects at close range using 12 infrared proximity sensors on its perimeter (Fig. 4B). These sensors have a range of 12 centimeters. In addition, the gripper has a VGA camera in its centner. This camera is capable of applying in hardware a Sobel filter to the image, which eases line and object detection.

3.3 Electronics and Control

The motors of the hand-bot embed a position encoder, which provides control in position and speed. In addition, the driving electronics measure the currents in the motors, which provides control in torque. The electronics is built around modules containing a 16-bit microcontrollers (Fig. 5). Each microcontroller manages the sensors and actuators it is locally connected to. The modules communicate together using asynchronous messages. They exchange them over a CAN bus [17] using the ASEBA architecture [18,19]. The ASEBA architecture distributes the processing locally by embedding a lightweight virtual machine inside each microcontroller. An integrated development environment compiles bytecode for these virtual machines out of a user-friendly scripting language. The development



(a) passive compliance when grasping (b) locations of the infrared sensors

Fig. 4. The gripper can open at 90°and close completely. When closing, its parallel compliance mechanism allows it to grasp objects of different thicknesses. The claws can apply a force up to 4 kg thanks to the high reduction of its worm drive.

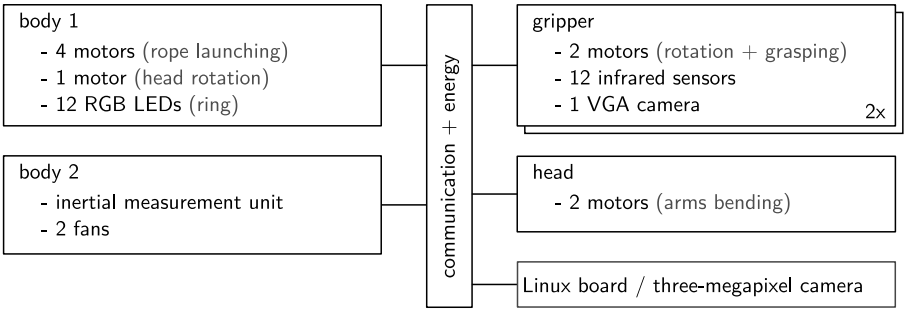


Fig. 5. Distribution of functions between the different microcontrollers. Each box represents a different microcontroller. The microcontrollers communicate together using asynchronous messages that they exchange over a CAN bus using the ASEBA architecture.

environment runs on a remote computer and provides the concurrent development — including debugging — of the programs on all the microcontrollers. Once complete, the program can be flashed into the microcontrollers for autonomous operations.

While the ASEBA architecture is powerful enough to implement any low-level autonomous behavior such as climbing a shelf; high level cognitive tasks such as locating a specific book using vision require more computational power. To fulfill this need, we will equip the hand-bot with an embedded computer running Linux, built around an ARM 11 processor and 128 MB of RAM. This computer will also provides a Wifi connection and a three-megapixel camera located in the centner of the head. The experiments that we discuss in the next section run solely using ASEBA [18]; but the scenarios presented in the introduction will require this embedded computer.

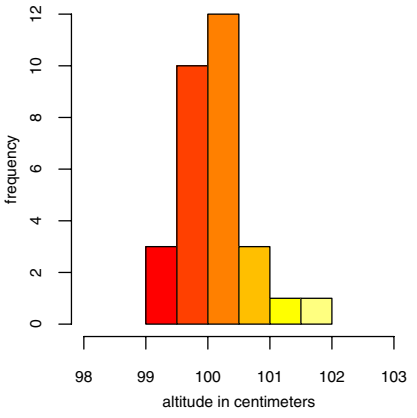
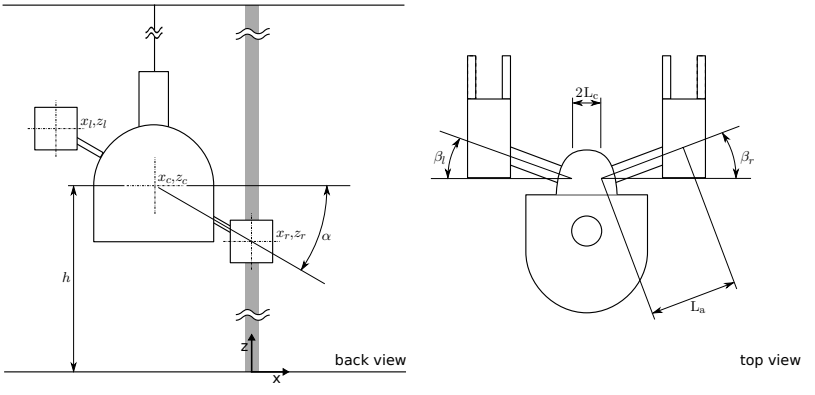
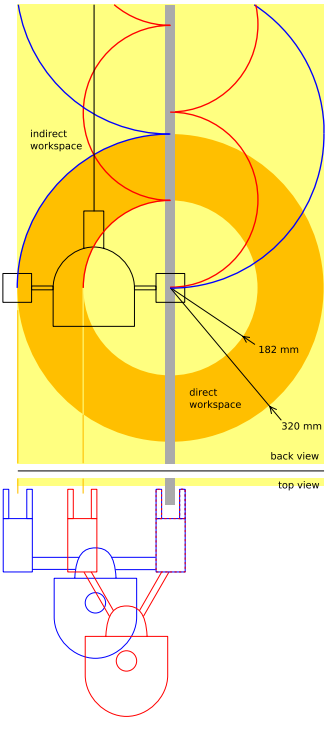


Fig. 6. Distribution of altitudes reached by the hand-bot for a target of 100 cm over 30 runs. The empirical mean is 100.2 cm and empirical standard deviation is 0.6 cm.

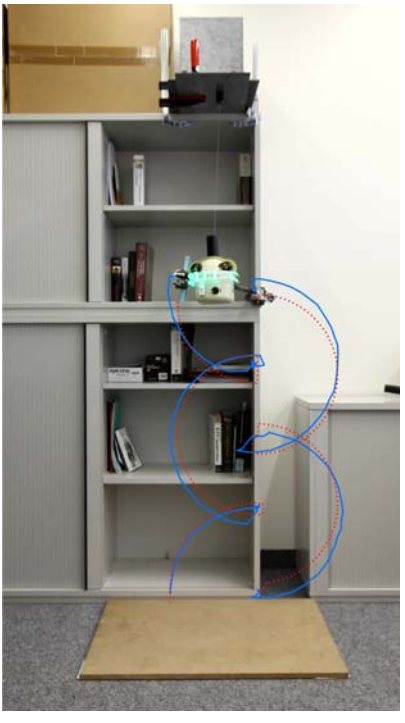


$$l_l = \mathrm{L_c} + \mathrm{L_a}\cos(\beta_l) \qquad l_r = \mathrm{L_c} + \mathrm{L_a}\cos(\beta_r) \qquad x_r = 0 \qquad z_r = h - l_r\sin(\alpha)$$
$$x_l = x_r - (l_l + l_r)\cos(\alpha) \qquad z_l = z_r - (l_l + l_r)\sin(\alpha) \qquad x_c = l_r\cos(\alpha) \qquad z_c = h$$

Fig. 7. Equations of the positions of the grippers given the length of the coiled rope, the head rotation, and the arms extensions. These equations apply when the right gripper is attached; they are symmetric when the left one is attached. The constants are the followings: $\mathrm{L_a} = 145$ mm is the length of an arm. $\mathrm{L_c} = 20$ mm is half the length between the arms attachments points in the head. The variables measured by the robots sensors are: β_l, β_r are the bending of the left, right arm. α is the rotation of the head. h is the altitude of the centner of the head. The variables we look for are x_c,z_c, x_l,z_l, and x_r,z_r: the positions of the center of the head, the left gripper, and the right gripper.



(a) workspace



(b) climbing traces

Fig. 8. (A) Workspace of the hand-bot. The trajectories of the grippers for the two extreme positions of the arms are drawn in blue and red. The dark orange disc shows the direct workspace, which is the locus of the positions where the hand-bot can grasp an object with one gripper fixed. The light yellow zone shows the indirect workspace, which is the locus of all locations where the robot can grasp an object, possibly after climbing up and down (B) Trace of the hand-bot climbing to fetch a book. The blue lines show the real trajectories of the grippers, extracted from Video 2. The red dotted lines show the trajectories of the grippers, as computed by the robot.

3.4 Precision of Altitude Control

The length of the coiled rope is the only information of the altitude of the robot. The experiment that we present in Sec. 4 considers the retrieval of a book located at a predefined altitude; which means that the precision of the altitude control is a determinant factor for its success. The hand-bot counts the rotations of the reel to estimate the length of the coiled rope. However, the rope might coil itself in different ways which might affect the effective radius of the reel. Moreover, the rope is elastic and its coiled length per reel rotation might change from launches to launches.

For these reasons, we characterize the error on the altitude. We let the hand-bot launch its rope and then elevate by coiling the rope until it reaches a target setpoint of 100 centimeters above the ground. We have repeated this sequence

35 times. In five of these trials, the launch was unsuccessful as the magnet did not attach to the ceiling. In the other 30 trials, the launch was successful. This results in a 86 percent success rate. Fig. 6 shows the altitudes that the hand-bot reached in these 30 trials. The mean altitude is 100.2 centimeters and the standard deviation is 0.6 centimeter; the biggest error is 2.0 centimeters. The altitude control using the rope is thus precise enough to climb to a target altitude and retrieve an object.

3.5 Workspace

The hand-bot climbs vertical structures by rotating its arms with one gripper fixed, and by switching grippers after each half turns (Fig. 8A). At the same time, the hand-bot coils the rope which provides the main lifting force. As the hand-bot only needs one gripper attached to provide stability, it can use the other one to manipulate objects. The hand-bot can compute the position of a gripper knowing the length of the coiled rope, the position of the head, and the extension of the arms (Fig. 7). While keeping a gripper attached, the hand-bot can reach objects located in a vertical disk (inner radius 18.2 cm, outer radius 32.0 cm) centered around this gripper. To access objects out of this disk, it must climb up or down, change the extensions of its arms, and then climb back to a position where it can access the object. In this way, the hand-bot can reach any objects located at maximum at 32.0 centimeters away from climbing pole (Fig. 8A).

4 Experiment: Climb of a Shelf and Retrieval of a Book

In this experiment, the hand-bot starts lying on the ground with its right gripper attached to a vertical pole, in this case the border of a shelf. The hand-bot first launches the magnet, and then climbs the shelf by switching grippers alternatively, always keeping at least one gripper attached. To do so, the hand-bot rotates its head slowly and coils the rope accordingly using its elevation motor. The hand-bot actively maintains the grippers in parallel with its body, so that it can grasp the border of the shelf or a book. The hand-bot uses the infrared proximity sensors of the gripper and the orientation of the head to decide when to grasp. When the hand-bot reaches a specific altitude (120 cm), it scans for the book using the proximity sensors of the gripper. When it has grasped the book, it goes down, freely hanging at the rope. Video 1[1] shows this sequence. This experiment lasts for two minutes.

Fig. 8B shows the trajectories of the grippers while the hand-bot climbs the shelf and retrieves the book. The figure shows both the real trajectories extracted from Video 2[2] and the trajectories estimated by the robot. The robot

[1] Hand-bot retrieves a book, video 1: `http://www.youtube.com/watch?v=92bLgE6DO2g`

[2] Hand-bot retrieves a book, video 2: `http://www.youtube.com/watch?v=FFyqfO51sik` In the video, we move slightly the magnet once it is attached, because the shot of the current prototype is not precise enough. This does not affect the autonomy of climbing.

computes them using the length of coiled rope, the head orientation, and the arms extensions. Overall, the two trajectories match well. Both the real trace and the estimated one show a vertical displacement when the robot attaches (or detaches) one of its gripper. The reason is that the arm is a parallel structure, so when the robot retracts (or extends) an arm, the projected position of the gripper on the vertical plane moves away from (or moves closer to) the robot.

There are several causes to the differences between the two traces. First, climbing creates high torques on the grippers' joints, which affects the robot's balance that is not perfectly horizontal anymore. In the future, we will solve this problem by fusing in the control the information from an accelerometer. Second, as the robot climbs it displaces itself on the horizontal axis, which creates discrepancies with respect to the assumptions of our cinematic model which considers a vertical rope. However, the passive compliance of the grippers allows the robot to overcome this difference and to climb successfully anyway. In the future, we plan to take this effect into account in our model. Third, we manually marked the centers of the grippers every 0.5 seconds on the video of the experiment, which introduced errors on the positions. In addition, the projection of the three-dimensional scene on the two dimensional sensor of the camera introduces major distortions on the depth axis, and minor distortions on the other axis.

5 Future Work

The current prototype of the hand-bot is able to autonomously climb a shelf and retrieve a book, provided that the robot knows the approximate position of the book. In the future, we will extend the autonomy of the hand-bot in various directions. First, we will add the main processor board running Linux and the three-megapixel camera to enable the robot to perceive and interact with objects in a more dynamic way. Second, we will explore self-assembling of the hand-bot with ground robots, to provide ground mobility to the hand-bot. This will allow us to implement more complex scenarios and explore deeper scientific questions than what is currently possible with a single hand-bot. Third, we will analyze collaborative tasks involving several hand-bots that manipulate large objects together. Finally, we will perform benchmarks of the collective approach with respect to single-robot solutions, such as humanoids.

6 Conclusion

Albeit the use of collective robotics for service applications imposes larger development and introduction costs then an approach based on single robots, it provides specific advantages. These include robustness, flexibility, and scalability that might prove critical for applications such as large deployments of robots in domestic or industrial environments. However, the relationship between the added value in performances and the additional costs in design and development is still unexplored, especially for complex tasks such as the ones involving manipulation. Robots such as the hand-bot provide an engineering foundation to

explore this scientific question. The prototype of the hand-bot managed to successfully climb a shelf and retrieve a book autonomously. It achieved this thanks to its innovate synergy between its climbing and manipulation subsystems.

Acknowledgements

We thank Basilio Noris who shot all the photos of the exterior of the hand-bot and provided us with material and expertise for filming. This work was supported by the Swarmanoid project, which is funded by the Future and Emerging Technologies program (IST-FET) of the European Community under grant IST-022888. The information provided is the sole responsibility of the authors and does not reflect the Community's opinion. The Community is not responsible for any use that might be made of data appearing in this publication.

References

1. IFR Statistical Department: World Robotics 2008. IFR Statistical Department (2008)
2. Konolige, K., Fox, D., Ortiz, C., Agno, A., Eriksen, M., Limketkai, B., Ko, J., Morisset, B., Schulz, D., Stewart, B., et al.: Centibots: Very large scale distributed robotic teams. In: Experimental Robotics: The 9th International Symposium, Springer Tracts in Advanced Robotics (STAR), pp. 131–140. Springer, Heidelberg (2005)
3. Sahin, E.: Swarm robotics: From sources of inspiration to domains of application. In: Swarm Robotics Workshop: State-of-the-art Survey, pp. 10–20. Springer, Heidelberg (2005)
4. Winfield, A.F.T., Harper, C.J., Nembrini, J.: Towards dependable swarms and a new discipline of swarm engineering. In: Şahin, E., Spears, W.M. (eds.) Swarm Robotics 2004. LNCS, vol. 3342, pp. 126–142. Springer, Heidelberg (2005)
5. Dorigo, M., Gambardella, L., Mondada, F., Floreano, D., Nolfi, S.: Swarmanoid: Towards humanoid robotic swarms, `http://www.swarmanoid.org/`
6. Groß, R., Tuci, E., Dorigo, M., Bonani, M., Mondada, F.: Object transport by modular robots that self-assemble. In: Proceedings of the 2006 IEEE International Conference on Robotics and Automation, pp. 2558–2564 (2006)
7. Staritz, P., Skaff, S., Urmson, C., Whittaker, W.: Skyworker: a robot for assembly, inspection and maintenance of large scale orbital facilities. In: IEEE International Conference on Robotics and Automation, 2001. Proceedings 2001 ICRA, vol. 4, pp. 4180–4185. IEEE Press, Los Alamitos (2001)
8. Balaguer, C., Gimenez, A., Huete, A., Sabatini, A., Topping, M., Bolmsjo, G.: The MATS robot: service climbing robot for personal assistance. IEEE Robotics & Automation Magazine 13(1), 51–58 (2006)
9. Hillenbrand, C., Berns, K.: Inspection of surfaces with a manipulator mounted on a climbing robot. In: 37th International Symposium on Robotics, ISR (2006)
10. Longo, D., Muscato, G.: The alicia3 climbing robot: a three-module robot for automatic wall inspection. IEEE Robotics and Automation Magazine 13(1), 42–50 (2006)

11. Luk, B.L., Cooke, D.S., Galt, S., Collie, A.A., Chen, S.: Intelligent legged climbing service robot for remote maintenance applications in hazardous environments. Robotics and Autonomous Systems 53(2), 142–152 (2005)
12. Zhang, H., Zhang, J., Zong, G., Wang, W., Liu, R.: Sky cleaner 3: a real pneumatic climbing robot for glass-wall cleaning. IEEE Robotics and Automation Magazine 13(1), 32–41 (2006)
13. Aracil, R., Saltaren, R., Reinoso, O.: A climbing parallel robot: a robot to climb along tubular and metallic structures. IEEE Robotics and Automation Magazine 13(1), 16–22 (2006)
14. Vona, M., Detweiler, C., Rus, D.: Shady: Robust Truss Climbing with Mechanical Compliances. In: International Symposium on Experimental Robotics, pp. 431–440. Springer, Heidelberg (2006)
15. Scheidegger, N., Mondada, F., Bonani, M., Siegwart, R.: Bi-pedal Robot for Rescue Operations. In: 9th International Conference on Climbing and Walking Robots, pp. 425–430 (2006)
16. Krishna, M., Bares, J., Mutschler, E.: Tethering system design for Dante II. In: Proceedings of 1997 IEEE International Conference on Robotics and Automation, 1997, vol. 2 (1997)
17. ISO Standard 11898: Road Vehicles Interchange of Digital Information - Controller Area Network - ISO 11898. International Organization for Standardization (1993)
18. Magnenat, S., Retornaz, P., Bonani, M., Longchamp, V., Mondada, F.: Aseba: a modular architecture for event-based control of complex robots (submitted for publication, 2009)
19. Magnenat, S., Mondada, F.: Aseba Meets D-Bus: From the Depths of a Low-Level Event-Based Architecture. In: IEEE TC-Soft Workshop on Event-based Systems in Robotics, EBS-RO (2009)

Human Multi-robots Interaction with High Virtual Reality Abstraction Level

Khelifa Baizid[1,2], Zhao Li[1,2], Nicolas Mollet[1], and Ryad Chellali[1]

[1] Italian Institute of Technology, TERA Dept, Via Morego, 30 Genova, Italy
[2] Genova University, Via Balbi, 5 - 16 Genova, Italy
baizid.khelifa@iit.it
http://www.iit.it

Abstract. Tele-operating robots' team is a way to increase tasks type and complexity, the working space size and to improve the remote interactions robustness. These interesting potentialities have a counterpart in terms of human robots interface. Indeed, we increase the complexity of the system and users must handle heterogeneous entities with different intrinsic mobility and sensing capabilities: tele-operators make integration and prediction efforts, firstly to built the remote world status and secondly to generate the right commands to be sent to the robots. In this paper, we present the platform we are developing to allow multi-robots tele-operation based tasks. The main aim of this platform is to support investigations about Human Multi-robots interfaces. Namely, to conduct studies concerning the integration of virtual and augmented reality technologies to reduce operators mental efforts. A description of this system is given. We present preliminary results in executing a simple scenario allowing to a single operator to supervise a given area.

1 Introduction

Multi-robots concept was introduced in early 2000's to take advantage of the inherent redundancy of this system. This property allows them to improve the system's robustness, to increase versatility and to endow co-operation capabilities. Following that, researchers tackled some challenging robotics applications and developed interesting platforms such as underwater and space exploration, hazardous environment monitoring or service robotics (USAR) systems. Unfortunately, when facing complex tasks and environments, such systems lack perception and cognitive capabilities. Therefore, human intervention is needed at different levels. Indeed, current effective autonomous or semi-autonomous systems still need human presence within the control loop, at least to supervise high level operations. In other words, to cope with complexity, the cognitive workload remains within the operators' part. For tele-operation systems in general, the control and thus the cognitive workload is also supported mainly by operators. More specificaly, these last compensate perception lacking and integrate the remote robot's model in order to generate the right controls. For multi-robots systems, the same situation exists and naturally, it is amplified because

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 23–32, 2009.
Springer-Verlag Berlin Heidelberg 2009

of a greater combinatory: more degrees of freedom, and more feedback sensory information to integrate, leading to a wider solutions' set that has to be handled to achieve a specific task.

To improve tele-operated multi-robots performances, two main directions were investigated:

1. Human robots interfaces
2. Adjustable autonomy

The first item concerns techniques allowing to simplify the interactions among users and robots. Devices, enabling natural or close natural acquisition of commands and controls to be sent to robots set, were developed. On the other hand, people have also developed rendering devices to display sensory feedback information. Both groups of devices aim to pre-process and prepare the information, in order to make it intelligible for robots and humans. The main difficulty remaining in this context is the knowledge about humans: which processes impact his/her decision making and how does he/she integrate stimulus coming from non human agent.

The second item concerns a more general topic, namely robots autonomy. Techniques and concepts are focusing on endowing robots with more capabilities to handle complex situations with a minimal help of humans.

In this paper, we focus on the design of human-robots interfaces. Namely we concentrate on studying the potential contribution of Virtual Reality (VR) technologies to create a useful feedback information flow letting users understanding easily the status of the remote world, including the tele-operated robots. VR in our case is used as a stimuli generator: depending on the description provided by the remote sensors, we create a synthetic description of the remote world from which decision making is natural. The other contribution here concerns control tools. We also use VR technologies in order to facilitate the understanding of the controls and commands users want to send to the remote robots to be executed.

In the first part of the paper we give a short classification of human-robots interfaces approaches. In the second section, we will describe more deeply our platform, namely Virtual Reality for Advanced Teleoperation (ViRAT). Within this part, we will describe the platform's structure as well as its basic building blocks. In the third and last part, we will show through two scenarios how ViRAT is used in order to accomplish collaborative tasks. The first one is a multi-robots' "Hello world" experiment and the second one is an area surveillance-monitoring task from precise viewpoints.

2 Human Robots Interaction: Concepts, Classification and Overview

In this section, we will briefly summarize and classify existing research areas about the interaction between human and group of robots, (Figure 1a) and its interface design from VR perspective. We propose that the interaction between human and robots can be looked from three angles (See figure 1b). In

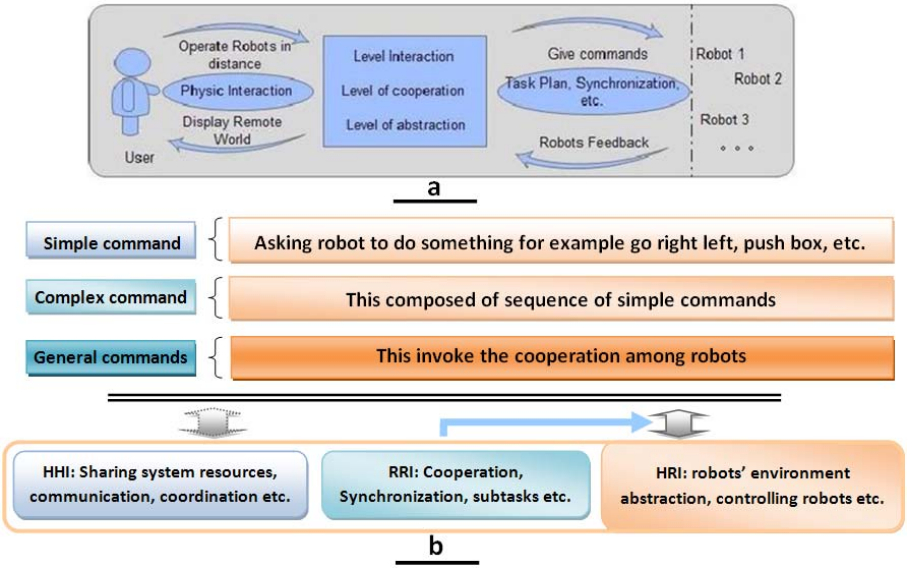


Fig. 1. Teleoperation concept, (a) Teleoperation is not a direct control on an object: it's a chain of interaction mediators. (b) Different human-robots intereaction classes.

case of multiple users, the interaction could be based on how much users have cooperated, communicated, operated (a single or group of robots) and shared system resources together. This issue has been called as Human-Human Interaction (HHI). The concept of interaction based on VR abstraction has been introduced by Nicolas and Ryad [MR08] termed as Collaborative Virtual Environment (CVE). The concept of this environment allows users to personally interact with real and virtual robots. The concept of multi-users interacting with a single virtual robot [AD02] or multi-robots has been given in [MWC05]. On the other hand, complex missions [SFM98] usually need cooperation of more than one robots [CP97]. The organization [XySqDs05] and coordination [BA98] among several robots have been considered in the category of Robot-Robot Interaction (RRI). This may provide a high user-robot interaction level with the group [MR08]. User's perception of robots' feedback and how he/she provides a suitable control to robots based on the understanding of remote world state can be named as Human Robot Interaction (HRI) [Sch03]. This last category is targeted in this work and more precisely we will consider the HRI as the ability of human (user) to positively interact with group of robots.

Our overview of HRI deals with two important aspects: *From Human to Robots* and *From Robots to Human*. In fact, the interaction interface should translate user's intention to remote robots. The figure 1b shows multiple ways in which user can command robots through this interface from lower interaction level (simple commands) to higher interaction level (general commands) [MR08]. The higher level can reduce user's cognitive workload and improve RRI

efficiency. Commands can be defined by a specific communication language as implemented in [JEJM00]. The second aspect is how user easily perceive and understand robots' feedback. This can be related with the interface feature to abstract current useful information and then present it to user. Concerning this aspect, the VR simulation and visualization impact combined with other sensory feedbacks has been proposed in [FLdD+05]. Furthermore, interaction interface should be standardized [MBKR08] where different kinds of robots can be integrated easily. Another important aspect is that how VR can allow user to operate the real robots as well as the virtual ones. In this case, the real robot and its virtual image object should be fully synchronized.

Several concepts about the interaction of human with robots have been suggested, from different VR perspectives. However, important aspects which have been partially discussed are the real-time VR abstraction, real/virtual robots' environments coordination, standardization concepts etc. Compared with cited research works the concept of VR that impacts to decrease human's workload and create a useful information source for robots' feedback to user will be further highlighted. More detaileds about the proposed aspects will be presented within next sections.

3 Human Robot Interaction via the ViRAT Project

We have designed ViRAT platform (Figure 2) to support several heterogeneous robots. As a standardized platform, the integration of such kind of robots is easier because of ViRAT modularity feature. As mentioned earlier, the HRI interaction should implement two main interaction concepts ***Human to Robot (Control)*** and ***Robot to Human (Feedback)***. In fact, ViRAT platform provides different kinds of control layers to user. User can operate a robot directly by simple commands and can also operate a group of robots in higher control layer [MCB09]. The platform has the real time feedback mechanism. This feedback is an important information about the system (e.g. of video of supervised area). Because of feedback, the VR based interaction feature allows user to abstract the remote environment in a higher level of interaction. This feature abstracts the correct and useful information from the real environment and present it on the virtual environment. In this case, VR is acting as commands input and information output console to and from the robots system. Augmented Reality (AR) features functions have also been provided through the platform interface where user can see the real environment using Head Mounted Display (HMD).

3.1 Technical Description

As we can see from the figure 2 ViRAT makes the transition between several users and group of robots. It's designed as follows:

1. *ViRAT Human Machine Interfaces* provide high adaptive mechanisms to create personal and adapted interfaces. ViRAT interfaces support multiple users to operate at the same time even if the users are physically at different

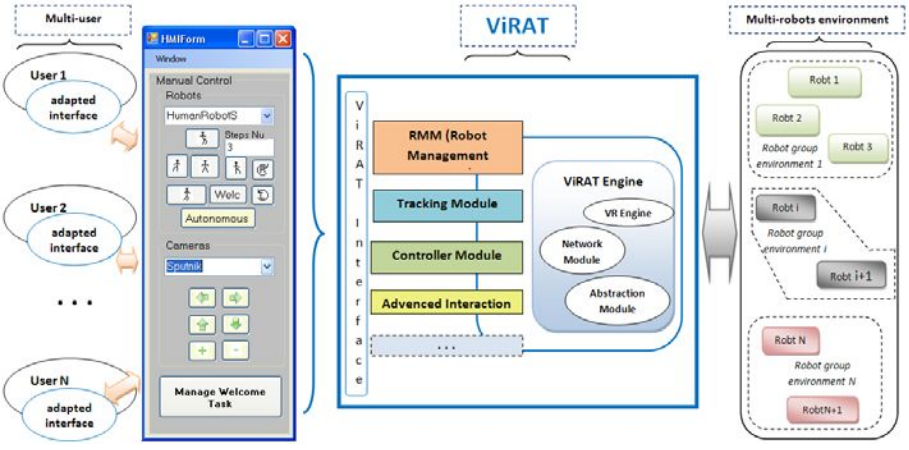


Fig. 2. ViRAT design

places. It offers innovative metaphors, GUI and integrated devices such as Joystick or HMD.

2. *Set of Plug-in Modules*. These modules are presented as follows:
   - Robot Management Module (RMM) gets information from the ViRAT interface and tracking module and then outputs simple commands to the control module.
   - Tracking Module (TM) is implemented to get current states of real environment and robots. This module also outputs current states to abstraction module.
   - Control Module (CM) gets simple or complex commands from the ViRAT interface and RMM. Then it would translates them into robots' language to send to the specific robot.
   - Advance Interaction Module (AIM) enables user to operate in the virtual environment directly and output commands to other module like RMM and CM.
3. *ViRAT Engine Module* is composed of a VR engine module, an abstraction module and a network module. VR engine module focuses on VR technologies such as: rendering, 3D interactions, devices drivers, physics engines in VR world, etc. VR abstraction module gets the current state from the tracking module and then it abstracts the useful information, that are used by the RMM and VR Engine Module. Network Module handles communication protocols, both for users and robots.

3.2 How ViRAT Platform Works

When a user gives some commands to ViRAT using his/her adapted interface, the standardized commands are sent to the RMM. Internal computations of this

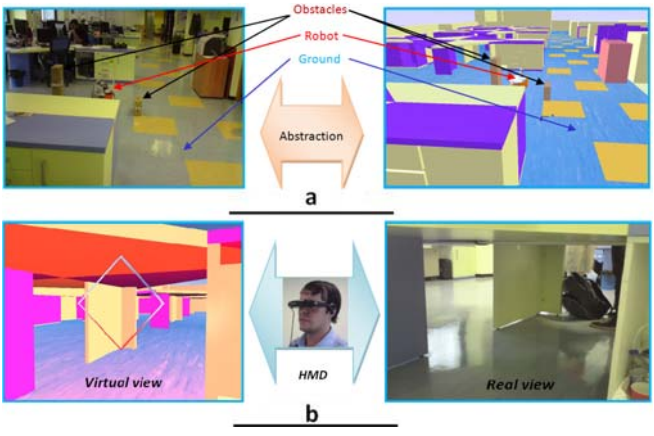


Fig. 3. VR abstraction, (a) Virtual environment interaction tool. (b) One example of some Metaphors given through VR tools.

last module generates simple commands for the CM. During the running process, the TM gets the current state of the real environment and sends the state to the Abstraction Module, which abstracts the useful information in VIRAT's internal models of representation and abstraction. Based on this information, VR engine module updates the 3D environment presented to the user. RMM re-adapts its commands according to users' interactions and requests.

ViRAT project has a lot of objectives [MBCF09], but if we focus on the HRI case there are two main objectives that interest us particularly for this paper:

1. *Robot to Human*
   (a) Abstract the real environment into the virtual environment: This will simplify the environment for the user. Ignorance of useless objects makes the operation process efficient. In the abstraction process, if we use a predefine virtual environment (Figure 3a), it will be initialize when the application starts running. Otherwise we should construct the new virtual environment with user's help. The same situation happens when we use ViRAT to explore an unknown area for example. After construction of virtual environment in accordance with the real environment, we can reuse the virtual environment whenever needed. Sometimes for the same real environment, we have to use different virtual abstracted environments. Thus the virtual environment must be adaptable to different applications.
   (b) Take feedback from the real environment and reflect on the virtual environment: ViRAT has an independent subsystem to get the current state information from real environment termed as 'tracking module' in the previous section. The operator makes decisions based on the information pecepted from the virtual environment. Because the operator does not need all the information from the tracking module so abstraction module will optimize and present the state information in real-time to user.

2. *Human to Robot.* Transfer commands from the virtual environment into the real world: Often user uses other interfaces to give commands to robots. However, ViRAT provides a mechanism to the user to operate the real robots from the virtual environment and an easily understandable way to get the information that user needs to know from the reality. For the user, it is analogous to 'what you want is what you see'.

4 Experiments and Results

4.1 Description, Concept and Goals

The main objective in this section is to present our approach though ViRAT platform, to propose new solutions for real tele-operation missions and to evaluate human in such interaction context. For this purpose, two example applications have been developed and tested with the ViRAT platform. These applications show the possibilities of cooperation between two robots in a 'Hello World' robotics task (hereafter called as 'Welcome task') and an example of a high level interaction tasks through VR. The features of those applications include inter-robots cooperation, coordination between real and virtual environments, and VR based interaction. These example applications serves two purposes:

- *Cooperative Tasks.* Welcome task consists of welcoming someone in any location in our lab area. The task involves two robots Sputnik[1] robot and Erectus[2] robot. The latter is a small Humanoid robot. Sputnik brings Erectus to a desired place in order to welcome the guest. The cooperation here is considered as an assistance to Humanoid robot because its locomotion is very slow, so it will take lot of time to reach the guest place. In the same time, Sputnik robot doesn't have a camera to provide user the real view of the guest, so Humanoid robot can accomplish this task using its camera, which is the goal of the second example. In order to manage these example applications, several subtasks showed in figure 5 have been defined. These subtasks have been synchronized each other to construct a mission scenario, through a scenario language [MBCF09].
- *VR based interaction.* The user can give general command (For example, the GoNear command) to the robots and then RMM will generate the subtasks for this command. Also, the General Command "welcome the visitor" runs the whole scenario automatically. During the welcome mission, user may interact with the group, showing the path and the targets to Sputnik and defining the requested view from the VR environment. The robot path should be defined in the VR world by users, or generate with a path planning module, because Sputnik is simulated has having problems with its distance sensors, so it blocks the robot to navigate through the environment's obstacles. This shows the idea of user robots cooperation and assistance since

[1] Dr Robot.
[2] Advanced Robotics and Intelligent Control Centre.

Sputnik helps Humanoid to reach the desired target, while Humanoid helps the user to welcome his/her guest or to obtain a local subjective real live view.

4.2 Details of Running Scenarios and VR Benefits

Welcome task scenario. Step by step (not through the general command "welcome" that runs everything automatically), first the user gives a general request to Sputnik robot to go near Humanoid robot (figure 4a). Then the system invokes several modules to accomplish this subtask. The TM provides the position and the orientation of the Sputnik while the initial position of Humanoid is predefined. The RMM generates the movements subtask to reach the Humanoid. Generally three kinds of movements are generated if there is no collision prediction of Sputnik with Humanoid: *Rotation*, *Translation* and then *Rotation*. Sputnik will arrive at the position only 15cm further from Humanoid. In this position (Figure 4a) it is easy for Humanoid to climb on Sputnik, which has the capability to transport it to the target location. When the user clicks on 'Manage Welcome Tasks' button on the interface, the Humanoid starts to prepare for climbing on Sputnik. After finishing this task, Sputnik moves and puts its supports under Humanoid's arms. Finally, Humanoid flexes its arms on the supports, and bends the legs (Figure 4b-3). At this time user intervenes through virtual environment to define Sputnik's path, since this robot does not know guest's location. At this time the RMM and CM will be invoked to carry out this sequence of movements, while TM provides the current state of robots. A *correction method* is regularly invoked during this time and is combined with the VR engine module and abstraction module to update the current real Sputnik's position in the virtual environment. When Humanoid gets off from Sputnik, the Sputnik should move a bit so that the Humanoid can welcome the guest conveniently (Figure 4b-9).

Advanced interaction task scenario and VR's metaphors. The scenario is an evolution of the welcome task, and aims to show the interest and potentiality of VR metaphors and abstraction. However, the user neither need to ask Sputnik to go near Humanoid, nor to define the Sputnik path in the virtual environment. All mission's subtasks are auto-generated by the system, including path planning. The user just has to wear HMD to see the real view corresponding to the virtual one. In fact, the user interacts with the entire system.

The interaction via VR to tele-operate a group of robots provides several metaphors. The first metaphor deals with the definition of a virtual robots' path to follow. Another metaphor is how user can obtain a real view through a virtual one (Figure 3b). In this case, user navigates easily and freely into the virtual environment and sets the area he wants to supervise (so to *really observe*), then the system organize the robots and their camera and finally provides the live-view image from the defined area, through Humanoid's camera. Since HMD and Humanoid's head are synchronized, therefore user can move freely and naturally his/her head to feel immersed and present through the Humanoid's robot.

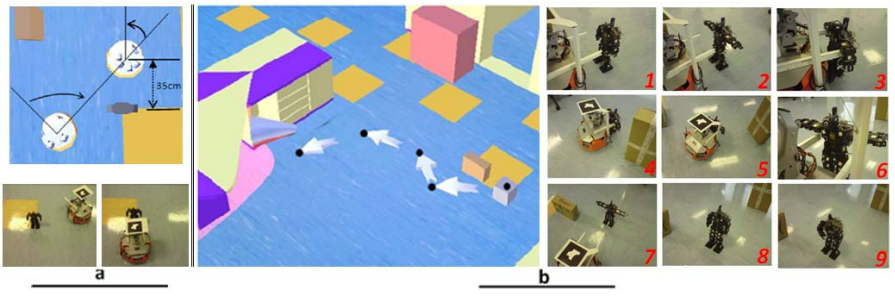


Fig. 4. Interaction through VR environment. (a) Go Near sub-task. (b) Welcome scenario.

Fig. 5. Subtasks defined for Welcome Task

5 Conclusion and Future Work

In this paper HRI concepts have been presented and proved. We have proposed the concept of different classes and levels of interactions among human and robots. This concept can vary between higher and lower interactive levels. The main objective is to show how humans can interact easily through existing interactions tools. Advanced interactive concepts based on VR technology have been proposed and tested. The impact of VR abstraction on human cognitive and perception capabilities has been evaluated. Vital issues like Human robots cooperation and inter-robot coordination have been discussed. Keeping in view these objectives, we have proposed and implemented ViRAT platform. This platform provides capability to interact with multiple robots simultaneously. Multiple remote users can interact with robots in parallel using multi-interfaces including GUI, HMD, VR etc. Using these interfaces, human can give commands and supervise states of robots. Two important application examples have been proposed and tested to prove the concept. These examples suggested the capability of ViRAT modules to interact with one other to achieve a desired mission. VR's metaphors provided by the implementation of VR technology have also been discussed. The proposed ViRAT platform finds its applications in disasters management, rescue in case of building collapse and earthquakes etc with slight modifications. In near future, we are going to implement ViRAT platform to quantify HRI performance parameters and analyze various concepts including tele-operation efficiency, multi-robot management, cognition etc.

References

[AD02] Monferrer, A., Bonyuet, D.: Cooperative robot teleoperation through virtual reality interfaces. In: 6th IEEE International Conference on Information Visualisation, Computer Society, July 2002, pp. 243–248 (2002)

[BA98] Balch, T., Arkin, R.C.: Behavior-based formation control for multirobot teams. IEEE Transactions on Robotics and Automation 14(6), 926–939 (1998)

[CP97] Christensen, H.I., Pirjanian, P.: Theoretical methods for planning and control in mobile robotics. In: Conference on Knowledge-Based Intelligent Electronic Systems, vol. 1, pp. 81–86 (May 1997)

[FLdD+05] Crison, F., Lecuyer, A., d'Huart, D.M., Burkhardt, J.-M., Michel, G., Dautin, J.L.: Virtual technical trainer: learning how to use milling machines with multi-sensory feedback in virtual reality. In: IEEE Conference on Virtual Reality, March 2005, pp. 139–145 (2005)

[JEJM00] Romano, J.M.G., Camacho, E.F., Ortega, J.G., Bonilla, M.T.: A generic natural language interface for task planning application to a mobile robot. Control Engineering Practice 8, 1119–1133 (2000)

[MBCF09] Mollet, N., Brayda, L.G., Chellali, R., Fontaine, J.-G.: Virtual environments and scenario languages for advanced teleoperation of groups of real robots: Real case application. In: Second International Conferences on Advances in Computer-Human Interactions, February 2009, pp. 310–316 (2009)

[MBKR08] Mollet, N., Brayda, L.G., Khelifa, B., Ryad, C.: Standardization and integration in robotics: case of virtual reality tools. In: IEEE International Conference on Cyberworlds, September 2008, pp. 525–530 (2008)

[MCB09] Mollet, N., Chellali, R., Brayda, L.: Virtual and augmented reality tools for teleoperation: improving distant immersion and perception, pp. 135–159 (2009)

[MR08] Mollet, N., Ryad, C.: Virtual and augmented reality with head-tracking for teteoperation of group of robots. In: IEEE International Conference on Cyberworlds, September 2008, pp. 102–108 (2008)

[MWC05] Castillo-Effen, M., Alvis, W., Castillo, C.: Modeling and visualization of multiple autonomous heterogeneous vehicles. In: IEEE International Conference on Systems, Man and Cybernetics, October 2005, vol. 3, pp. 2001–2007 (2005)

[Sch03] Scholtz, J.: Theory and evaluation of human robot interactions. In: 36th Annual Hawaii International Conference on System Sciences (January 2003)

[SFM98] Schneider-Fontan, M., Mataric, M.J.: Territorial multi-robot task division. IEEE Transactions on Robotics and Automation 14(5), 815–822 (1998)

[XySqDs05] Xian-yi, C., Shu-qin, L., De-shen, X.: Study of self-organization model of multiple mobile robot. Advanced Robotic Systems Journal 2(3), 1729–8806 (2005)

Laser Based People Following Behaviour in an Emergency Environment

Jose Maria Martínez-Otzeta, Aitor Ibarguren, Ander Ansuategi, and Loreto Susperregi

Fundación Tekniker, Av. Otaola 20, 20600 Eibar, Gipuzkoa, Spain
jmmartinez@tekniker.es

Abstract. Rescue robots have a large application potential in rescue tasks, minimizing risks and improving the human action in this kind of situations. Given the characteristics of the environment in which a rescue robot has to work, sensors may suffer damage and severe malfunctioning. This paper presents a backup system able to follow a person when camera readings are not available, but the laser sensor is still working correctly. A probabilistic model of a leg shape is implemented, along with a Kalman filter for robust tracking. This system can be useful when the robot has suffered some damage that requires it to be returned to the base for repairing.

1 Introduction

Rescue robots have a large application potential in rescue tasks, minimizing risks and improving the human action in this kind of situations. For an overview of potential tasks of rescue robots and the related research in general see for example [7].

One of the main challenges in using robots in search and rescue missions is to find a good trade-off between completely remotely operated devices and full autonomy. The complexity of search and rescue operations makes it difficult, if not impossible, to use fully autonomous devices. On the other hand, the amount of data and the drawbacks of limited communication possibilities make it undesirable if not unfeasible to put the full control of the robot into the hands of a human operator.

Human-robot collaboration has significant potential to improve rescue missions. Specifically, by enabling humans and robots to work together in the field. The mission productivity can be greatly increased while reducing cost, particularly for surface operations such as material transport, survey, sampling, and in-situ site characterization.

Given the characteristics of the environment in which a rescue robot has to work, sensors may suffer damage and severe malfunctioning. There are several approaches in the literature that combine vision with other sensors (laser, sonar) to provide a reliable people following behaviour. Our approach is to build a backup system able to follow a person when camera readings are not available, maybe due to some hardware failure, but the laser sensor is still working

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 33–42, 2009.

correctly. This system can be useful when the robot has suffered some damage that requires it to be returned to the base for repairing.

In order to accomplish the people following behaviour, this paper presents a probabilistic approach for human leg detection based on data provided by a laser scan, and developed within the ROBAUCO project. Besides the theoretical aspects of the leg detection, the proposed approach integrates tracking techniques as Kalman filters [5], [9] to endow the system with a error recovering tool to be used in a real enviroment. Finally, to avoid the robot to get stuck when an erroneous detection has been performed, after a predefined time span without movement, the robot looks for other possible targets.

The paper is organized as follows. Section 2 briefly describes the ROBAUCO project where this research is located. Section 3 gives information about previous research activities related to this paper. In Section 4 the hardware used in the experiments is described, including the mobile robot and laser scan. Section 5 presents the proposed approach for leg detection and tracking. Section 6 and Section 7 describe the experimental set-up and experimental results of the proposed architecture. Finally Section 8 presents the conclusions obtained and the future work to be done.

2 ROBAUCO Project

"ROBAUCO - mobile, autonomous and collaborative robots" project's main objective is the generation of the technologies necessary for the development of mobile robots able to carry out complex tasks with a high degree of autonomy and capacity for collaboration. These robots, moreover, have to share tasks with people in the most friendly and natural way possible.

ROBAUCO expects to materialise all these developments in a terrestrial robot prototype which, in all probability, will be a test bank for solutions to emergency situations such as forest fires, rescues, etc. In order to know the peculiarities and skills these tasks require and thereby to orientate the prototype accordingly, contacts have been made with SOS Deiak (the Basque Emergency Rescue Service) and it is also expected to know other viewpoints from other autonomous emergency services.

The project is one of six which, at a Spanish state-wide level, is being financed by the State Office for Small and Medium Enterprises of the Ministry of Industry, Tourism and Trade, through the programme of partnered projects designed to stimulate a synergic effect from the collaboration of various technological centres.

3 Related Work

People detection and tracking is a popular topic in the computer science community. Several approaches have been tested, most of them based in some kind of vision sensors. For example, color vision has been used [8], as well stereo [3] or infrarred cameras [2].

Multisensor approaches have also been developed, for example fusing stereo and thermal images [4] or vision and sonar readings [6].

In [1] a person is detected and tracked in two different ways depending on the sensor: the vision sensor detects the face and the laser detects the legs. A fusion is made between these two modules. Though some papers deal only with leg detection [10], without relying in other sensors, they are comparatively rare in the literature.

This paper describes a system for detecting legs and follow a person only with laser readings. To increase the reliability of the detector, a Kalman filter is implemented, as well as a procedure to recover when being stuck by a false detection.

4 Hardware

The next section describes the hardware used for the experiments, including the mobile robot and laser scan.

4.1 Mobile Robot

During the experiments a robuLAB80[1] mobile robot has been used, with dimensions of 772x590x475 mm and a weight of 125kg, as seen in Fig. 1. This mobile robot, compatible with Microsoft Robotics Studio, is equipped with a Pentium 1.4 GHz, 512 MB RAM and multiple analog and digital inputs and outputs.

Fig. 1. Left: robuLAB80 mobile robot used in the experiments Right: Hokuyo laser

Besides the described hardware, the robuLAB80 has also been equiped with different sensors and actuators such as bumpers, Hokuyo laser scan (described later), security light and an ultrasound belt to fullfill the needs of ROBAUCO project.

4.2 Laser Scan

The laser scan chosen for the leg detection task is a Hokuyo URG-04LX[2], as seen in Fig. 1. This laser scan provides a measuring area of 240 angular degrees, from 60 to 4095 mm in depth and 625 readings per scan.

[1] http://www.robosoft.fr/eng/
[2] http://www.hokuyo-aut.jp/

5 Propossed Approach

The proposed approach presents an architecture comprising several modules: a leg detection unit, a Kalman filter unit and a control unit that receives results from the other two modules and combines them to produce the final result. An scheme of the architecture is presented in Fig. 2.

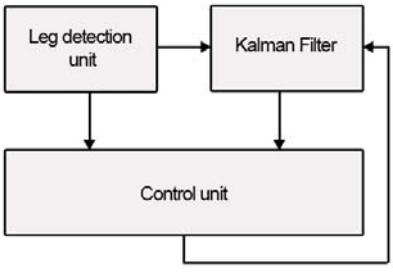


Fig. 2. System architecture

The leg detector has some similarity to the one developed by [1]. In its paper, given a set of laser readings in an instant, a leg pattern is defined as a sequence of maximums and minimums, where there are some constraints between them. In Fig. 3, from the paper above mentioned, the points P_a, P_b, P_c, P_d and P_e define a leg. The constraints they use are of the type $P_x - P_y > K$, where P_x and P_y are two of the points that define the pattern, and K is a constant. The idea is to reflect the physical constraints (distance between feet, between a foot and the background, etc.) that occur in the real world.

In this work, instead of implementing a leg detection system that only returns *detected* and *not detected* states, the above mentioned constraints have been extended to implement a fuzzy detector. The difference between marks $P_x - P_y$ is associated to the range of values permitted, as well as an optimum value that it is considered to maximize the probability of being part of a leg. Constraints are now of the form

$$\begin{aligned} &\textit{if } P_x - P_y = \textit{out of range then prob(detected)} = 0 \\ &\textit{else prob(detected)} = \Phi(abs((P_x - P_y) - \textit{optimum})) \end{aligned}$$

where Φ is a function that assigns maximum value (one) when the difference between marks reaches its optimum, and decreases when the value drifts from the optimum, until it reaches a minimum of cero when the value is out of range.

Due to the exclusive use of laser readings (no other sensor is used), a Kalman filter has been implemented to add reliability to the system.

To avoid the robot get stuck following something in the environment erroneously identified as a leg, when the robot is not moving for a predefined time span, it starts to spin on its axis looking for other possible targets.

5.1 Leg Detection Unit

The proposed system has implemented a leg model as a sequence of $max \rightarrow min \rightarrow max \rightarrow min \rightarrow max$, given the laser readings, as in [1]. The main difference lies in the computation over those values. While in their work they implement rules of *all or nothing*, we have chosen a fuzzy approach, where the rules represent the likelihood of a characterist of a leg. Therefore, to estimate the probability of a set of laser readings to be a leg, we have implemented several measures over the readings at P_a, P_b, P_c and P_d and P_e, as seen in Fig. 3.

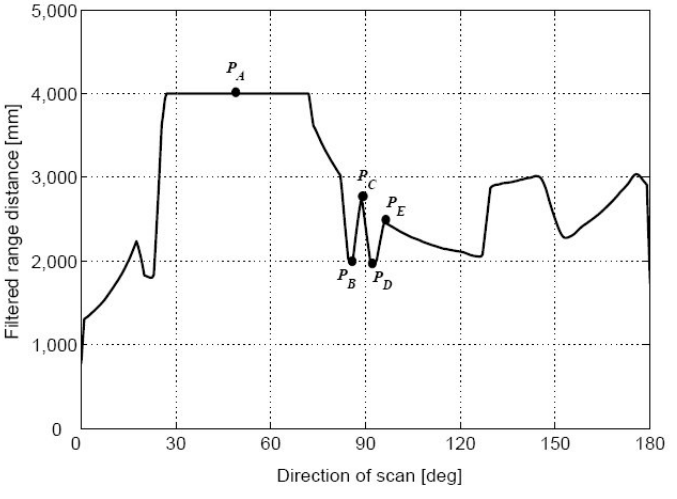


Fig. 3. Leg pattern (Bellotto et al.)

Measures over the laser readings:

1. Likelihood of the distance between P_b and P_d supposed it is part of a leg (distance between feet)
2. Likelihood of the distance between P_b and P_c, and between P_c and P_d (distance between legs and the background between them)
3. Likelihood of the distance between P_a and P_b (distance between right foot and previous background)
4. Likelihood of the distance between P_d and P_e (distance between left foot and posterior background)

The function Φ introduced above has been implemented for every measure in the following way:

For measure 1: there is a value that is considered *optimum*, and a *minimum* and *maximum*, both of them defining the range lenght. Therefore,

$$\Phi(x)=\begin{cases} 0 & \text{if } x <= minimum \\ (x - minimum)/(optimum - minimum) & \text{if } x > minimum \wedge x < optimum \\ 1 & \text{if } x = optimum \\ (maximum - x)/(maximum - optimum) & \text{if } x > optimum \wedge x < maximum \\ 0 & \text{if } x >= maximum \end{cases}$$

For measure 2, 3 and 4: there is a *threshold* value, for which the values above it are asigned a maximum probability, and a *minimum* value, with the same meaning that in the previous measure.

$$\Phi(x)=\begin{cases} 0 & \text{if } x <= minimum \\ (x - minimum)/(threshold - minimum) & \text{if } x > minimum \wedge x < threshold \\ 1 & \text{if } x >= threshold \end{cases}$$

The overall probability of a leg is computed as a combination of these four measures, each of them ranging between 0 and 1. So far only an arithmetic average has been tested, taking into account that when one of the component values is zero, the probability of the resulting combination is zero too.

5.2 Kalman Filter Unit

The next quation shows the known Kalman filter formulas:

$$\begin{aligned} \mathbf{x(k)} &= \mathbf{Ax(k-1)} + \mathbf{w(k-1)} \\ \mathbf{y(k)} &= \mathbf{Cx(k)} + \mathbf{v(k)} \end{aligned}$$

where $\mathbf{x(k)}$, $\mathbf{x(k-1)}$ are the state vectors at time k and $k-1$, $\mathbf{y(k)}$ is the observation vector at time k, $\mathbf{w}$ and $\mathbf{v}$ are the noises on the state and the observation respectively. $\mathbf{A}$ and $\mathbf{C}$ are matrices specifying the state and measurement models for the considered system. The Kalman theory gives the equations for the optimal estimate $\widehat{\mathbf{x}}(\mathbf{k+1}|\mathbf{k})$ given the statistics of the system and observation noises.

In the case of the leg tracking addressed in this paper, the next state vector $\mathbf{x} = [\ \alpha \quad \omega\ ]^{\mathbf{T}}$ and observation vector $\mathbf{y} = [\ \alpha\ ]^{\mathbf{T}}$ are defined where α is the angle of the detected legs and ω defines the angular velocity of the leg position. The covariance of the state and measurement noises are estimated directly from data.

During leg tracking, a Kalman filter is instantiated when legs are detected and it is used to predict leg's position in the next time step. This Kalman filter is iteratively updated with the new leg position provided by the *Leg detection unit*. On the other hand, when the *Leg detection unit* provides no output (legs are lost) the *Control unit* feeds the Kalman filter with its own (Kalman filter's) predictions to update it, allowing further predictions.

5.3 Control Unit

Finally the *Control unit* combines the information provided by both the *Leg detection unit* and the *Kalman filter unit*. Initially it compares the difference between the detected leg position (α_{det}) and the Kalman filter's prediction (α_{Kalm}). If the difference is below a threshold K, the output from the *Leg detection unit* is accepted. On the other hand, if the difference is greater than the threshold K, the detected leg position is analysed. If the positions' confidence rate ($Conf_{det}$) is greater than threshold P, the detected leg position is accepted, otherwise the

Kalman filter's prediction is the chosen position. A summary of the process is shown in formula (1).

$$\alpha_{returned} = \begin{cases} \alpha_{det} & \text{if } |\alpha_{det} - \alpha_{Kalm}| <= K \\ \alpha_{det} & \text{if } |\alpha_{det} - \alpha_{Kalm}| > K \wedge Conf_{det} > P \\ \alpha_{Kalm} & \text{if } |\alpha_{det} - \alpha_{Kalm}| > K \wedge Conf_{det} < P \end{cases} \quad (1)$$

Apart from the previous information fusion, the *Control unit* also feeds and updates the *Kalman filter* when no leg is detected. For security reasons, if no leg is detected for a period of time (around 1-2 seconds) the *Control unit* stops updating and predicting by means of the Kalman filter and starts a new leg search process.

6 Experimental Set-Up

To assess the performance of the built system, an experimental setup in a manufacturing scenario has been devised as it is shown in Fig. 4.

The manufacturing plant is a real manufacturing shop floor where machines and humans share the space in performing production activities. With regard to the exploration purposes the shop floor can be characterised as an industrial environment, with multiple machines, tools and materials in unpredictable arrangement.

The experimental method comprises several runs along a path of about 30 meters with different ambient conditions (given by the daily changing activity), pace of walking and different people.

The laser readings have been restricted to the range between $-45\,^{\circ}$ and $45\,^{\circ}$, to avoid calculations over readings that are far away from the natural path of the robot. Denoting the definition of the Φ function (as defined in the previous section) corresponding to the measure number i as Φ_i, the following values have been taken in our experiments:

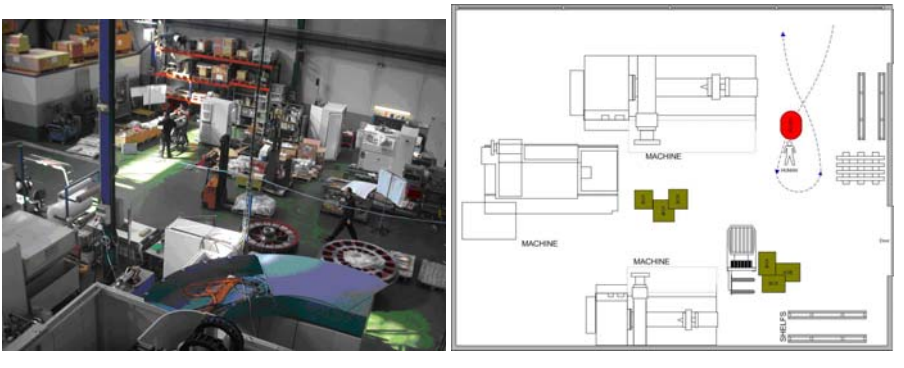


Fig. 4. Manufacturing plant and path covered in the experiments

- Φ_1: Optimum: 50 cms Minimum: 5 cms Maximum: 150 cms
- Φ_2: Threshold: 75 cms Minimum: 5 cms
- Φ_3: Threshold: 75 cms Minimum: 5 cms
- Φ_4: Threshold: 75 cms Minimum: 5 cms

With respect to the thresholds K and P defined in the *Control unit* section, a value of 5 angular degrees has been selected for K, and a confidence rate (probability of a leg) of 0.40 has been selected as value for threshold P.

Once a person (their legs) is detected, the robot moves in its direction, not approaching more than a predefined security distance. In these experiments, the distance has been of 100 cms.

7 Experimental Results

Fig. 5 shows the graph corresponding to one of the experiments (all the experiments show a similar pattern). The number of times the path has been covered has been of five, and the number of people involved in the experiment of three. The graphs show time in milliseconds against distance and angle from the robot to the person. This distance has been measured from the laser readings. As it can be observed, the robot tries to keep the distance around that value, maintaining it in a range of 1-1.5 meters most of the time. The times the distance

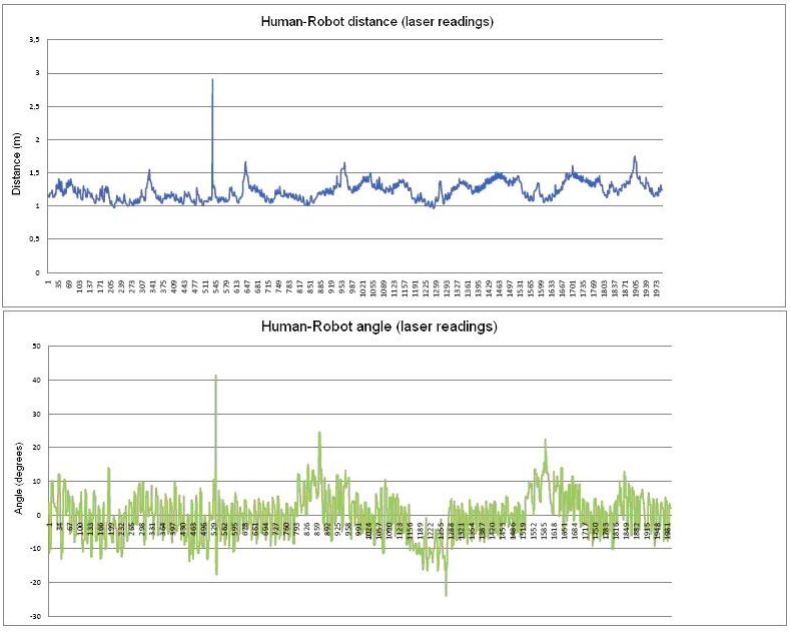


Fig. 5. Graphs depicting the distance and angle the robot maintains through an experiment

is bigger is due to the human walking too fast for the robot to follow him/her. In the same way, it is observed that the measured angle lies between $-10\,^\circ$ and $10\,^\circ$ most of the time, which represents the lateral movements of the robot when the person is not moving straight forward. As expected, the laser readings are not 100% reliable, as is shown by the appearance of some peak noise around the 500 milliseconds mark.

The system has shown to be robust, losing its track only in cases when the person walks so fast that the person disappears from its sight, or when the person walks close to glasses, which are difficult for the laser to deal with. Even in these adverse conditions, the system is able to recover itself pretty fast when the person locates in front of it again.

The distance between the person and the robot, as measured by the laser readings, always lies over the minimum security distance, and even when the robot loses the person's track and the subject has to put him/herself in front of it, the robot is able to recover the track before colliding.

8 Conclusions and Future Work

In this paper a prototype of a robotic system to assist emergency personnel in rescue tasks has been presented. Specifically a leg detection architecture has been developed, including a tracking system to improve its reliability. Experiments have also been carried out to test the leg detection system, yielding good results.

A robotic mobile platform integrating this leg detection architecture has been developed, in which the robot movement will be directed by the user. Current experiments show a good performance in a real industrial shop floor.

As further work, there are two interesting paths to follow. On one hand, leg detection could be improved adding new features to track as well as adjusting the probabilistic framework used. On the other hand, the tracking system could also the modified, studying the use of other estimation algorithms as extended Kalman filters or particle filters to observe their efficiency in the posed problem.

This research has been supported by the PROFIT project ROBAUCO FIT-170200-2007-1, funded by the Spanish Government, and by the ETORTEK project TEReTRANS, funded by the Basque Government.

References

1. Bellotto, N., Hu, H.: Multisensor integration for human-robot interaction. The IEEE Journal of Intelligent Cybernetic Systems 1 (July 2005)
2. Dai, C., Zheng, Y., Li, X.: Pedestrian detection and tracking in infrared imagery using shape and appearance. Comput. Vis. Image Underst. 106(2-3), 288–299 (2007)
3. Gavrila, D.M., Davis, L.S.: 3-D model-based tracking of humans in action: a multi-view approach. In: CVPR 1996: Proceedings of the 1996 Conference on Computer Vision and Pattern Recognition (CVPR 1996), Washington, DC, USA, p. 73. IEEE Computer Society, Los Alamitos (1996)

4. Guan, F., Li, L.Y., Ge, S.S., Loh, A.P.: Robust human detection and identification by using stereo and thermal images in human-robot interaction. International Journal of Information Acquisition 4(2), 1–22 (2007)
5. Kalman, R.E.: A new approach to linear filtering and prediction problems. Transactions of the ASME–Journal of Basic Engineering 82(Series D), 35–45 (1960)
6. Kalyan, B., Balasuriya, A., Ura, T., Wijesoma, S.: Sonar and vision based navigation schemes for autonomous underwater vehicles. In: International Conference on Automation, Robotics and Computer Vision, vol. I, pp. 437–442 (2004)
7. Nourbakhsh, I.: On the study of human-robot collaboration. In: 9th IEEE International Workshop on Advanced Motion Control, pp. 9–11 (2006)
8. Oren, M., Papageorgiou, C., Sinha, P., Osuna, E., Poggio, T.: Pedestrian detection using wavelet templates. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 193–199 (1997)
9. Welch, G., Bishop, G.: An introduction to the Kalman filter. Technical Report TR95-041, University of North Carolina at Chapel Hill (1995)
10. Xavier, J., Pacheco, M., Castro, D., Ruano, A., Nunes, U.: Fast line, arc/circle and leg detection from laser scan data in a Player driver. In: IEEE International Conference on Robotics and Automation, ICRA, pp. 3930–3935 (2005)

RENS – Enabling a Robot to Identify a Person

Xin Yan[1], Sabina Jeschke[2], Amit Dubey[4], Marc Wilke[1], and Hinrich Schütze[3]

[1] Institute for IT Service Technologies, University of Stuttgart
[2] Center for Learning and Knowledge Management, RWTH Aachen University
[3] Institute for Natural Language Processing, University of Stuttgart
[4] Institute for Communicating and Collaborative Systems, University of Edinburgh

Abstract. We outline a web personal information mining system that enables robots or devices like mobile phones which possess a visual perception system to discover a person's identity and his personal information (such as phone number, email, address, etc.) by using NLP methods based on the result of the visual perception. At the core of the system lies a rule based personal information extraction algorithm that does not require any supervision or manual annotation, and can easily be applied to other domains such as travel or books. This first implementation was used as a proof of concept and experimental results showed that our annotation-free method is promising and compares favorably to supervised approaches.

1 Introduction

This paper outlines our RENS personal information mining system. RENS was originally inspired by the scenario of a receptionist robot who determines the identity of a person by using a facial recognition process coupled to a web search process. To be able to focus on the aspect of personal information mining, we assume the existence of a functioning facial recognition system capable of collecting, for a given target person, *URL-image pairs*: pairs of a URL and an image such that the URL's page contains at least one *identifying image*, i.e., an image of the person's face. A URL-image pair indicates the mapping between a web page and its containing identifying image and therefore enables us to combine web mining and facial recognition technologies. After analyzing these URL-image pairs and their retrieved web pages, the system searches the web for the person's information and generates a business card for him as output.

The RENS system attempts to address the following three issues:

1. How to ascertain an unknown person's identity using given URL-image pairs
2. How to extract personal information from an HTML page
3. How to select the right personal information for a particular person

For the purposes of our system, the *identity* of a person is defined to include his name and his organization. The *personal information* of a person is defined by the typical information shown on a business card, including address, email, telephone, fax number, title and position. We define a *personal information record*

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 43–54, 2009.

(or simply, *record*) as an area with high density of personal information about a particular person on a web page.

The RENS system has three main components: Personal Identity Ascertainment (henceforth referred to as *PIA*), Records Extraction (*RE*), and Records Selection (*RS*). Each of the three components addresses one of the problems listed above. The PIA component ascertains a person's name and organization by applying a named entity recognizer to the relevant text contents of web pages that contain identifying images. The RE component moves through the DOM tree of a web page and extracts records by applying a rule-based algorithm. The RS component selects the best records matching a particular person by calculating and sorting the records' confidence scores based on cosine similarity.

Empirical evaluations were conducted on web pages related to people working in academia as their personal information is often freely available online. Experimental results show that the methods proposed in this paper are promising. Our two main contributions are: (1) an investigation of the concept of personal identity ascertainment with given URL-image pairs and (2) the development of a simple but powerful rule-based records extraction algorithm. This paper is organized as follows: Sec. 2 reviews related work. A description of the RENS system is given in Sec. 3. Results of the evaluation are discussed in Sec. 4. Sec. 5 provides conclusions and an outlook on future research.

2 Related Work

Prominent work in personal information mining includes the work of Tang et al.[1] and [2]. Their system ArnetMiner aims at extracting and mining academic social networks. The ArnetMiner system focuses on extracting researcher profiles from the Web automatically, integrating the publication data from existing digital libraries, modeling the entire academic network and providing search services on this academic network. Although extracting researcher profiles is only a component of ArnetMiner, it does similar tasks as the RENS system using a different approach. It first collects and identifies a person's homepage from the Web, then uses a unified approach to extract the profile properties from the identified document [1]. As it supports search for experts, which is similar to search for persons, it is taken as the baseline system in the evaluation of RENS. Yu et al.[4] discuss extracting personal information from résumés in a two step process: first, segmenting a résumé into different types of blocks and, second, extracting detailed information such as address and email from the identified blocks.

In addition to work directly concerning personal information mining, it is worth discussing research related to the underlying techniques used by RENS. RENS extracts information by walking through the nodes of a DOM tree. Such a DOM tree based extraction approach was first introduced by Gupta et al.[3]. Their basic idea was to use the links-to-text ratio and remove nodes with a high ratio in order to extract general web content from different domains. Prasad et al.[5] used a similar DOM-based heuristic applied to news stories. Kim et al.[6] suggested extracting information from DOM trees using tree edit distance.

Another aspect of the RENS system is its use of *wrappers*. A wrapper extracts data (including unstructured, semi-structured and structured data) from web pages and turns the data into a self-described structured representation for further processing. Liu et al.[9] proposed a novel algorithm that is able to discover noncontinuous data records and uses partial tree alignment to integrate data records. In another paper, Liu et al.[8] also proposed a two-step extraction approach by first identifying data records without extracting data and then aligning data items to a schema.

3 System Description

Figure 1 shows an overview of RENS. In addition to the PIA, RE and RS components mentioned in Section 1, there are 3 smaller components: preprocessing, the information source builder (*ISB*) and card generation. We describe these 6 components in turn.

3.1 Preprocessing

The preprocessing step takes a file containing several URL-image pairs as input, and retrieves the corresponding web pages. We call these pages "seed pages", which we use as a starting point to discover the identity and relevant personal information of a person.

3.2 Personal Identity Ascertainment (PIA)

The PIA component determines a person's name and information related to his organization by applying the Stanford Named Entity Recognizer(SNER) to the relevant text contents on each seed page respectively. The texts of a seed page are relevant when they describe the page's identifying image, like the image name, the text adjacent to the image, the image's alt text, the page meta information, title and so on. Because SNER is case-sensitive, espacially with person names, and because the texts online are often informally edited, we need to ensure the capitalization of letters is correct in order to increase the accuracy of the named entity recognition. We used a very simple heuristic: if a token is found having a capitalized first letter, then all the occurences of this token will be enforced to have their first letter capitalized. SNER tags proper nouns as PERSON, ORGANIZATION or LOCATION. Anything tagged as a PERSON could possibly be the person name we are interested in. We found that simply taking the most frequently occurring PERSON name resulted in poor results. In our case, a name having occurred 10 times but only in one document(here a seed page) is often less important than another name with an occurence of one time in 7 documents respectively. The high document frequency suggests a high global accordance to the initial information provided by seed pages. Thus, we picked the entity with the highest score according to the formula: $0.99 \times df + 0.01 \times cf$ where df refers to the document frequency (where a document is defined as

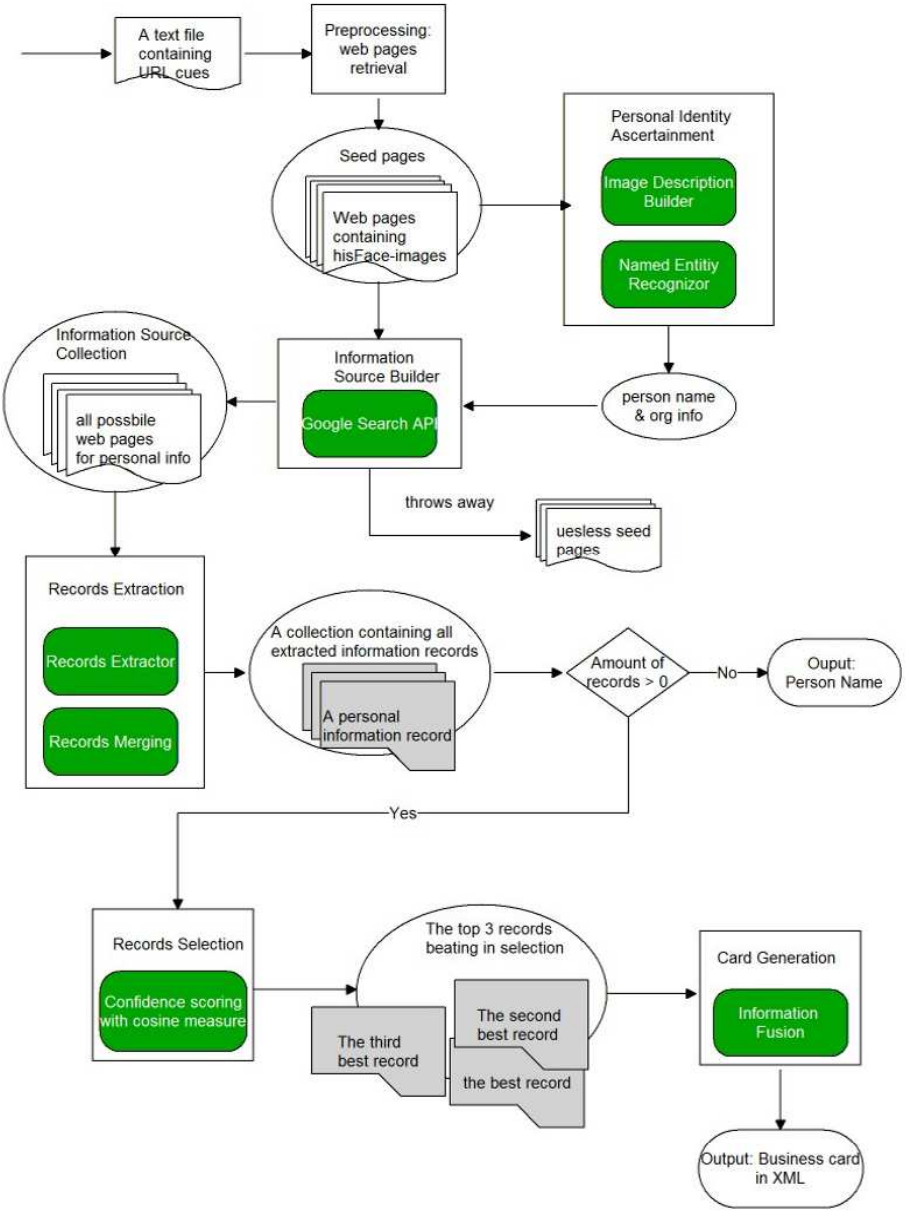


Fig. 1. RENS System Architecture. The 6 main components are highlighted in blue. If there is no record found, system output is the identity of the person. If records are found, the system outputs a business card in XML format.

the text content of a seed page), or number of documents in which a name occurred, and cf refers to collection frequency, or the number of times a name occurs among all the documents. Having attained the person name, one can use pointwise mutual information to find the best matching organization phrase corresponding to the person name. Organization-type named entities are indexed as bigram and trigram phrases; we do this because in a unigram representation, the weighting formula would give too much weight to stop words like "of" and "for" (which are semantically empty, but occur frequently in ORG names). In addition, the key words of organizational names are often phrases of length 2 or 3. The mutual information between person and organization phrases is calculated in the following manner:

$$MI(p, o) = \log \frac{P(p, o)}{P(p)P(o)} \tag{1}$$

where p stands for person name and o for organization phrase. $P(p)$ is the normalized document frequency of p, $P(o)$ is the normalized document frequency of o, and $P(p, o)$ is normalized document frequency of joint occurrences of p and o.

3.3 Information Source Builder (ISB)

This component has two purposes. First, it uses two search queries (one is the person's name, where possible the full name, another is the combination of the person's name and the organizational association that was also found in the PIA component) and the Google AJAX search API to get the top 10 ranked pages respectively for each query. The seed pages and new pages found during the search represent the information source collection. Second, the ISB removes repetitive or useless (not containing the person name found in the PIA) web pages from the collection.

3.4 Records Extraction (RE)

This component uses the information source collection, namely the output of the ISB, as input. For each page in the collection, it traverses the corresponding DOM tree, annotating all nodes with the personal information features of a particular language, like the ones for US. english mentioned in Table 1. It then uses a local extraction strategy to extract personal information records. The output of this component is a collection of all possible records that are detected from the pages in the information source collection. A recursive bottom-up extraction algorithm we developed in this project is given in Figure 2. The extraction process starts from the `<html>` node which is the root of an HTML DOM tree, and computes a weight for each node inside the DOM tree. This weight is the ratio of the text which can be classified as personal information compared to all the text in the node. When this ratio exceeds a predetermined threshold (empirically determined to be 0.13 for our task), we classify the node as being a potential personal information record.

Table 1. Personal information features: Attributes and Indicators of US. English

Attributes	Indicators
Email:	email, netmail e-mail, mailto ···
Telephone:	telephone, tel, call, mobile, phone, cellphone ···
Fax:	fax, telexfax, facsimile ···
Address:	department, avenue, Ave., building, room, office ···
Website:	homepage, website, url ···
Title:	professor, Ph.d ···
Position:	CEO, CFO, dean, chief, coach ···

```
PROCEDURE recordsExtractor(aNode, threshold)

    new Set attributes
    new Set indicators

    add the indicators detected from
    aNode into Set indicators

    add the attributes detected from
    aNode into Set attributes

    List children = get children of aNode

    FOR EACH aChildNode in children
        recordsExtractor(aChildNode, threshold)
        add attributes of aChildNode into Set attributes
        add indicators of aChildNode into Set indicators
    END

    IF amount of attributes > 1 THEN
        weight = proportion of indicators in the node text
            IF weight > threshold THEN
                aNode is a record.
            END IF
        RETURN attributes and indicators of aNode;
    END IF

END PROCEDURE
```

Fig. 2. The algorithm of the personal information records extractor

Computing this ratio therefore depends upon being able to classify text as containing personal information or not. Recall that personal information is the typical information shown on a business card. A unique property of this kind of information is that there are often obvious words and patterns which strongly indicate its presence. We manually developed a set of these 'indicator words' for US English, a subset of which is listed in Table 1.

Indicators are not limited to words, but also include regular expressions that identify personal information including email addresses, zip codes, telephone and fax numbers. These regular expressions and indicators are not only used to compute the above-mentioned ratio, but also to annotate a node as containing a particular kind of personal information. If an indicator occurs in a node (including its children), the node is annotated with this indicator and the indicator's corresponding attribute. The weight of a node is thus formalized to eq. 2. t is a node inside a DOM tree.

$$weight(t) = \frac{||\text{indicators in t}||}{||\text{ words in t}||} \tag{2}$$

In order to prevent a node with very high weight but only one attribute from being taken as a record, we require that a record have at least 2 attributes. However, not all attributes strongly indicate a personal information record. Attributes like position or title can occur within any node in a DOM tree, because their indicators like "Professor" or "CEO" could be mentioned anywhere on a page with a person name. On the other hand, particular email, fax, telephone or address patterns are very suggestive (in particular, the ZIP code pattern), so they are good attributes to identify a record. Thus, besides the weight, another precondition to be a record is that a node should have at least 2 good attributes. In addition, we count the attributes and indicators in a boolean model which means no attribute gains any additional weight beyond its first occurrence. We do this to dampen the weight of large nodes like `<html>` which may have more than one record as child nodes. If nodes that have 2 good attributes and exceed a predefined threshold are all chosen for records, we would have much redundancy, due to the nested structure of HTML pages. By eliminating the nested records, we finally attain the mutually exclusive personal information records on a page.

3.5 Records Selection (RS)

The records selection component calculates and sorts the confidence scores of the records, and outputs them in sorted order. Scoring uses cosine similarity:

$$\text{SIM}(R, P) = \frac{\boldsymbol{r} \cdot \boldsymbol{p}}{|\boldsymbol{r}||\boldsymbol{p}|} \tag{3}$$

In eq. 3, the vectors $\boldsymbol{r}$ and $\boldsymbol{p}$ are TF-IDF representations of the record R and the seed page P, respectively. The cosine measures the similarity between $\boldsymbol{r}$ and $\boldsymbol{p}$, and therefore shows how likely the record R relates to the person described on the seed page P. As there are multiple seed pages $(P_1, P_2, P_3, \ldots, P_n)$, the confidence score of the record R is the overall similarity and is calculated using:

$$score(R) = \sum_{i=1}^{n} \text{SIM}(R, P_i) \tag{4}$$

3.6 Business Card Generation (BCG)

The card generation component takes records as input and generates business cards in XML format as output. We predefined a business card template that is coposed of 5 slots:

person name, fax, telephone, address, academic title or position.

The person name slot is filled with the name we have found in personal identity ascertainment component. For the other slots, we use pattern matching and heuristic methods as annotating a DOM tree node mentioned in Section 3.4. In the end, the business cards are generated in XML format with JDOM.

3.7 Assumptions and Preconsiderations

As the system does not include an actual face recognition system, we have to set certain limitations on the test set used as input. Existing face recognition techniques are not perfect. To account for this deficiency and to simulate a real world scenario, we assume there are a few misleading URL-image pairs, containing information of "wrong" persons. Thus, our first, arbitrary assumption is that the error rate of the input URL-image pairs is 30%. The experiments are performed on web pages related to academics, a useful limitation as their personal information can be easily found online. To reduce complexity, we experimented only on web pages, not including files of other formats like pdf or MS Word. In future , there will be more investigations on these types of files. Finally, we assume that the person we are searching for has only one unique social identity. It is still unclear how to deal with people who have multiple social identities (a mathematician can also be a musician), and who have different personal information during different periods of time. We left further discussions of this problem to future work.

4 Empirical Evaluations

The evaluation consists of 3 tests:

1. *Records Extraction Test*: Given a web page, RENS decides whether the page contains records. If the page contains at least one record, it extracts all detected records from the page.
2. *Personal Identity Ascertainment Test*: Given a set of URL-image pairs, RENS ascertains the name of the person who the set points to. If no full name exists, it finds the first or last name.
3. *Evaluation of the* RENS *System*: Given a set of URL-image pairs, RENS finds the records best matching the target person.

The evaluation was designed to measure the accuracy of the RENS system in the framework of the tests defined above.

The records extraction test was performed on 815 web pages. 15 of these do not contain any records directly but have links to contact pages containing records. If RENS detects the records from the contact page of such a test page, we score this instance 1, else 0. In the other 800 web pages that do not have a linked contact page, 500 of them contained at least one record (most of them containing exactly one record). If records of a page are returned, we assign a score of 1 else 0. The other 300 test pages do not contain records. In this case, when the RENS system (correctly) returns no records, we score the instance with 1, else with 0.

For the personal identity ascertainment test and the RENS system test, we used 100 test sets. Each set is composed of 9 URL-image pairs, 3 of which are related to wrong persons, according to the 30% input error rate. In the personal identity ascertainment test, we checked manually whether the output person's name corresponds to the target person. If correct, accuracy is 1, else 0. In the last test, the RENS system test, we check how accurately RENS could find personal information on that particular person. The evaluation metric for this test is given below.

4.1 Metrics

The metrics used to evaluate our system are fine-grained accuracy and coarse-grained accuracy, both of which take a value between 0.0 and 1.0. The coarse-grained accuracy is computed by taking the ceiling of the fine-grained accuracy. The fine-grained accuracy is computed as follows:

- **Case 1** there are information records available for a particular person. If the best record is returned at the first place, the fine-grained accuracy is 1.0, second place 0.8, third 0.6, fourth 0.4, fifth 0.2. After 5th place, the fine-grained accuracy is scored 0.0.
- **Case 2** no information record is provided for a particular person. If no cards are returned by the RENS system, fine-grained accuracy is 1.0, else 0.0.

4.2 Baseline

As a personal information mining system, the RENS system was compared with the ArnetMiner system's expert search component. The ArnetMiner implements the process in three steps: relevant page identification, preprocessing, and extraction. Given a researcher name, they get a list of web pages by a search engine (we use the Google API) and then identify the homepage/introducing page of the researcher and in the end they extract personal information by using machine learning methods[1].

4.3 Results

Records Extraction Test. For the 500 pages containing records, RENS has an accuracy of 91.2%, for the 300 pages without records 93.33% and for the 15 embedded contact pages 80%. It reaches an average accuracy of 91.4% on the 815

test pages. *Personal Identity Ascertainment Test* The test result showed an accuracy of 96% for the PIA component. RENS *System Test* If the best record could be found in one of the seed pages, the result of the RENS system is exceptional with a fine-grained accuracy of 89.6% and a coarse-grained accuracy of 92%. As a comparison, the result of ARNETMINER is 81.6% and 92.0% respectively. If no record of the person is given in the seed pages, the performance of RENS drops down to a fine-grained accuracy of 72.0% and coarse-grained accuracy of 80.0%. In this case, ARNETMINER has 93.6% fine-grained accuracy and 96.0% coarse-accuracy. The average performance of RENS is 80.8% of fine-grained accuracy and 86.0% of coarse-grained accuracy, while ARNETMINER has 87.6% and 94% respectively.

4.4 Discussion

Records Extraction. The extraction test failed for 8.8% of the 500 pages that contain records, mainly because the shortcomings of the local extraction strategy cause false negatives. If a node contains many other text elements besides all the right personal information we need, its weight becomes too small to pass the threshold test. As a result, this node will not be classified as a record. A possible remedy for this weakness is to take the change rate of personal information into account. Inside a node, when entering the area that contains personal information, the number of indicators increases very quickly; upon leaving, the increase rate slows down and eventually approaches 0.

Of the 300 pages that did not contain any records, 7.67% were classified incorrectly. These errors were often numbers with a pattern identical to phone and fax numbers. This is a direct result of the use of regular expressions in the annotation of personal information. Additionally, some people have a separate contact information page that is linked from the main page and contains most of the personal information. To address this problem, RENS uses Google search to acquire additional information beyond the seed pages. The pages found in this manner usually contain the required information or, at worst, link directly to them. In the second case we could use simple regular expressions to extract the contact links. The test result was 12 out of 15 contact pages detected and extracted correctly with an accuracy of 80%. Our approach of records extraction needs many improvements to get a better performance. In many cases, it can not extract all the personal information at one time but requires post processing steps. However, the test result still indicates it is a simple but reasonable way to extract personal information.

Personal Identity Ascertainment. The error rate of 4% proves the high performance of the Stanford named entity recognizer and also the efficiency of our method. The two exceptions that were not found by our method are both Italian names. For the Asian names within our test set, the Stanford named entity recognizer shows a very high accuracy of 100% in recognition.

Rens System. If the seed pages contained the best record already, RENS had a slightly better result in fine-grained accuracy. In some cases, ARNETMINER does

not find any correctpersonal information, mainly because its strategy is based on finding a person's homepage or profiling first and then extracting his information. On those people who do not have a valid homepage or never published their personal information right on their homepage or whose personal information is in an embedded contact page, ArnetMiner does not perform very well. rens, in contrast, does not select the homepages as its only source for extraction, thus performing better in the same situation.

However, rens's performance was lower when the right personal information record was not included in the seed pages. This is probably because the search term is not good enough, or in many cases, false positive. If a person does not have his personal information available online, but a related person does, the personal information of this related person will be returned. We have not found a satisfying solution to this problem yet. ArnetMiner performed very well in this case. As a mature academic search engine project, it receives its search term by user input, providing an advantage at the level of search terms and its machine learning extraction approach is often more accurate on a large scale corpus. In the future, we can also apply our automatic annotation methods to prepare a corpus for machine learning approach. Although ArnetMiner has an advantage in search term correctness, in contrast to rens it requires name disambiguation as a large-scale academic search engine. Thus we consider our comparison to be fair.

5 Conclusion

We have presented a methodology for combining facial recognition and web mining technologies enabling a robot to determine a person's identity and his personal information based on visual perception. We have also implemented a simple, yet modular algorithm to extract data records like personal information from web pages. We have tested and compared the resulting, fully automatic system based on heuristics against ArnetMiner, which uses a machine learning approach and needs large labeled training sets. Our simple rule-based approach has shortcomings in accuracy, but delivers a good approximation and shows that our proof of concept is successful. There are a lot of potential future directions of this work. Name disambiguation is crucial to the performance of the system for large scale mining. Further interesting avenues for research are the discovery and interaction of different social contexts, like a person's information as a mathematician vs. as a musician and ensuring that the information mined is up to date.

References

1. Tang, J., Hong, M., Zhang, J., Liang, B., Li, J.: ArnetMiner: Extraction and Mining of Academic Social Networks. In: Proceedings of the Fourteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD 2008), pp. 990–998 (2008)
2. Tang, J., Hong, M., Zhang, J., Liang, B., Li, J.: A New Approach to Personal Network Search based on Information Extraction. Demo paper. In: Proc. of ASWC 2006 (2006)

3. Gupta, S., Kaiser, G., Grimm, P., Chiang, M., Starren, J.: Automating Content Extraction of HTML Documents, pp. 179–224. Kluwer Academic Publishers, Dordrecht (2004)
4. Yu, K., Guan, G., Zhou, M.: Resume information extraction with cascaded hybrid model. In: IACL 2005: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, pp. 499–506 (2005)
5. Prasad, J., Paepcke, A.: Coreex: content extraction from online news articles. In: CIKM 2008: Proceeding of the 17th ACM conference on Information and knowledge management, pp. 1391–1392 (2004)
6. Kim, Y., Park, J., Kim, T., Choi, J.: ArnetMiner: Web Information Extraction by HTML Tree Edit Distance Matching. In: International Conference on Convergence Information Technology, pp. 2455–2460 (2007)
7. Gomez, C., Puertas: Named Entity Recognition for Web Content Filtering. Natural Language Processing and Information Systems, 286–297 (2005)
8. Zhai, Y., Liu, B.: Mining data records in web pages. In: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), pp. 601–606 (2003)
9. Zhai, Y., Liu, B., Grossman, R.: Mining web pages for data records. IEEE Intell. Syst., 49–55 (November/December 2004)

Classifying 3D Human Motions by Mixing Fuzzy Gaussian Inference with Genetic Programming

Mehdi Khoury and Honghai Liu

Institute of Industrial Research, University of Portsmouth, Portsmouth, PO1 3QL, United Kingdom
{mehdi.khoury,honghai.liu}@port.ac.uk

Abstract. This paper combines the novel concept of Fuzzy Gaussian Inference(FGI) with Genetic Programming (GP) in order to accurately classify real natural 3d human Motion Capture data. FGI builds Fuzzy Membership Functions that map to hidden Probability Distributions underlying human motions, providing a suitable modelling paradigm for such noisy data. Genetic Programming (GP) is used to make a time dependent and context aware filter that improves the qualitative output of the classifier. Results show that FGI outperforms a GMM-based classifier when recognizing seven different boxing stances simultaneously, and that the addition of the GP based filter improves the accuracy of the FGI classifier significantly.

1 Introduction

The process of behaviour understanding is usually performed by comparing observations to models inferred from examples using different learning algorithms. Such techniques presented in [1] and [25] can be used either in the context of template matching [4], state-spaces approaches [27], or semantic description [21]. Our application domain is focused on sport, and more precisely, boxing. We have discarded template matching as it is generally more susceptible to noise, variations of the time intervals of the movements, and is viewpoint dependent [25]. We are not interested in a pure semantic description as we need to analyse and evaluate a boxing motion in a relatively detailed way. We therefore focus on identifying static states during a motion (state-spaces approach). Conventionally, machine learning techniques in use for solving such problems vary from dynamic Time Warping [3], to Hidden Markov Models [20], Neural Networks [10], Principal Component Analysis [26], or variations of these techniques. This study presents a novel machine learning technique tested in the application domain of behaviour understanding, that is to say the recognition and description of actions and activities from the observation of human motions. It introduces a different method that allows us to build from learning samples fuzzy qualitative models corresponding to different states. An automated way to generate fuzzy membership function is proposed[13]. It is applicable to biologically "imprecise" human motion, by mapping an estimation of centroid and range from a cumulative normal distribution to a membership function. In order to introduce a time

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 55–66, 2009.

dimension, instead of using a Mangmani-typed rule-based like in [2], we use Genetic Programming to build fuzzy rules that filter and refine the qualitative output. First the human skeletal representation in use will be described, then the process by which stances are recognized (Guard, Jab, Cross, Lower Cross, Lower Jab, Right Hook, Left Hook, Lower Left Hook, and Right Uppercut) with fuzzy membership functions, then some mathematical properties of this technique, and finally, experimental results will be presented and discussed.

2 Human Skeletal Representation

We use the widely spread .BVH motion capture format [24] in which a human skeleton is formed of skeletal limbs linked by rotational joints. It uses Euler angles to quantify rotations of joints having three Degrees of Freedom. This system is not perfect (Gimbal Lock is a possible issue), but allows to gather data easily when using motion capture while keeping track of subcomponents such as the rotations of individual joints. The following choices and assumptions are made:

- Knowing that motion capture data cannot give absolutely exact skeletal displacements of the joints [8] due to soft tissues movements, this work simply seeks to use it to obtain an approximation which would be good enough to characterize the motion.
- The body is simplified to nineteen main joints and it is assumed that this number is sufficient to characterize and understand the general motions of a human skeleton performing boxing combinations.
- Each joint is seen as having three degrees of freedom. The rotations of such joints are represented by Euler ZXY angles. A joint rotation is therefore characterized by three rotation angles Z, X and Y given in degrees by the .BVH motion capture format sampled at the speed of 120 frames per second.

In practice, for every frame, our observed data takes the shape of a nineteen-by-three matrix describing ZXY Euler Angles for all nineteen joints in a simplified human skeletal representation. In other words, 57 continuous variables (each between 0 and 360) characterize a stance at any time.

3 The Learning Method: Fuzzy Gaussian Inference

To learn to recognize a stance, a model needs to be extracted (here a fuzzy membership function) for this stance from learning data. This stance is later identified during a motion by evaluating the membership score of the observed data with respect to the learned model. This section will first describe the novel process by which a fuzzy membership function is generated: Fuzzy Gaussian Inference. Finally, it will show how the degree of membership of observed data to a given template is computed.

3.1 Model Generation

The fuzzy linguistic approach introduced by Zadeh [28] allows us to associate a linguistic variable such as a "guard" stance with linguistic terms expressed by a fuzzy membership function. Using a trapezoid fuzzy-four-tupple (a, b, α, β) which defines a function that returns a degree of membership in [0,1] (see equation 1) seems to be more interesting as there is a good compromise between precision and computational efficiency (compared with, for example, the triangular membership function).

$$\mu(x) = \begin{cases} 0 & x < a - \alpha \\ \alpha^{-1}(x - a + \alpha) & x \in [a - \alpha \quad a] \\ 1 & x \in [a \quad b] \\ \beta^{-1}(b + \beta - x) & x \in [b \quad b + \beta] \\ 0 & x > b + \beta \end{cases} \tag{1}$$

Frames identified as "Guard" of membership equal to one are used as learning samples. The identification of these example data is made by a system similar to Reverse Rating [22], which is to say that, in our case, an expert (a human observer) is asked to do the following: identify a group of frames whose motion indicates a stance that possesses the degree 1.0 of membership in the fuzzy set "Guard". Once these learning data are obtained, a fuzzy membership function can be generated. Many kinds of procedures for the automated generation of membership functions can be found in the literature[11][14][15][23][7][19]. So far, one downside of such techniques has been the difficulty to link the notion of fuzzy membership to the notion of probability distribution. One noticeable attempt to link both concepts in the generation of membership functions has been done by Frantti [9]in the context of mobile network engineering. Unfortunately, this approach is relatively limited as the minimum and maximum of the observed data are the absolute limits of the membership function. As a consequence, such a system ignores motions which are over the extremum of the learning range of the examples. This work presents a method that overcomes this problem by introducing a function that maps the probability that values fall within a given cumulative normal distribution to a degree of membership. This relies on the assumption that, for a population of samples representing a given motion, the Z, X and Y Euler angles characterizing the motion tend to be normally distributed. Assuming that the space of known boxing motions is informationally structured by these hidden Gaussian Distributions, there is a need to build fuzzy membership functions that map to these underlying structures. The mapping from probability distribution to membership score is done by examining the range and center of density of the learning data for one specific motion. In our experiment, there is a limited number of motion capture learning samples of a given stance (let us say a defensive posture called "Guard"). Looking at each Euler angle Z, X, Y for every joint j for this type of motion, it can be observed that, in our training sample, each Euler Angle e in each joint has a global minimum and maximum. This range is defined between minimum and maximum of the learning sample as the range $\delta(e, j)$ of degree of membership one

in the fuzzy set "Guard". Knowing the size of our training sample, it is possible to estimate how much *the range* of our learning sample represents compared to *the range* of all possible guards. For example, if the range of our sample represents around 68.2% of the maximum range of all possible guards ($\gamma = 0.682$), then there is a degree of membership 1 for two standard deviations (one on each side) on the population maximum range. This means that the rest of the distribution that will have membership inferior to one will take three remaining standard deviations on each side. To summarize the salient points of our method, considering the range and center of density of the learning sample, the shape of a fuzzy membership function will be defined by the following four factors:

- The maximum number of standard deviations covered by the fuzzy membership function. In this example, the maximum range is approximated by assuming that it is four standard deviations away in both directions from the mid-point of the range of membership one. This will define the length of the base of the trapezoid shape.
- Depending on the cumulative normal distribution evaluation defining the parameter γ, a portion of the four standard deviations representing the total range will be allocated to the membership-one-range and the remaining part will be allocated to the lower membership degrees. This will define the length of the top part of the trapezoid shape(see Figure 1).

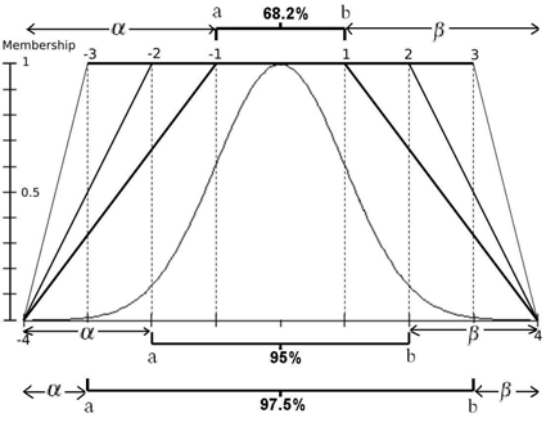


Fig. 1. Influence of the cumulative normal distribution parameter on the shape of the fuzzy membership function

- The average of the means is extracted out of each learning sample. This will correspond to the centroid of the data samples of membership one.
- While the distance $|(b + \beta) - (a - \alpha)|$ will be constant, $a - \alpha$ and $b + \beta$ will be shifted to the side proportionally to the way the centroid is shifted from the midpoint(see Figure 2 and equation 3). This will shift the base of the trapezoid shape to either side.

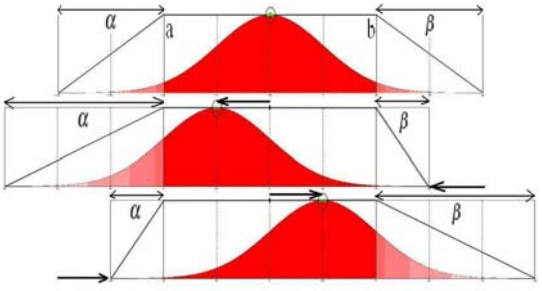


Fig. 2. Moving the centroid shifts the distribution and deforms the fuzzy membership function

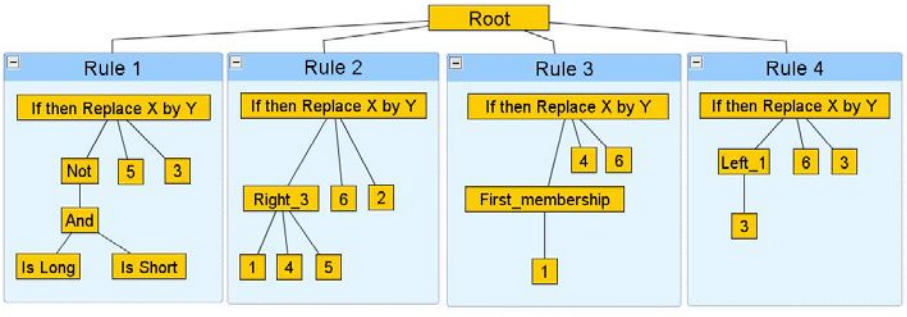


Fig. 3. A Typical set of Rules Generated by GP

For example, if the centroid is at the same position with the middle of the membership-one-range $\delta(e, j)$, and this range is evaluated as representing 95% of the maximum theoretical range, then our fuzzy membership function will be symmetric ($\alpha = \beta = 2$ standard deviations on each side of the membership-one-range). The centroid c and the constant μ are such that :

$$|c - a| = \mu \times |b - a| \tag{2}$$

Let this range be evaluated as representing 95% of the global theoretical range, then the fuzzy membership function would be shifted to the left such that :

$$\begin{cases} \alpha = (1 - \mu) \times (\alpha + \beta) \\ \beta = \mu \times (\alpha + \beta) \end{cases} \tag{3}$$

3.2 Membership Evaluation

Our observed data take the shape of a nineteen-by-three matrix describing ZXY Euler Angles for all nineteen joints. One evaluates how close this matrix is from

a "Guard" stance by calculating the degree of membership of every Euler Angle in every joint (we have previously built a fuzzy-4-tuple corresponding to the "Guard" stance for every one of these Euler angles), and then, an average membership score is computed. This approach could probably be improved in the near future by introducing weighted average for certain joints (for example, the position of the elbow might be more important than the position of the knee when in guard). If a frame has a high membership score for several fuzzy sets, an order of preference of these sets can be established by comparing the Euclidian distance of the observed data to the centroid of each fuzzy set.

4 Mathematical Properties

Fuzzy Gaussian Inference (FGI) does not have the problem linked to dimensionality reduction of methods such as PCA as we keep the initial number of dimensions when building the model. The method decomposes what would normally be a Gaussian Mixture of a number x of m-dimensional Normal distributions into $x \times m$ Fuzzy Membership Functions. In this study nineteen 3-dimensional rotation continuous data are used to produce $19 \times 3 = 57$ fuzzy membership functions. The flexibility of a machine learning method is generally determined by how successfully it can be applied to different application domains. Empirically speaking, making use of supervised machine learning techniques generally involves testing a data sample with different parameter values in order to reach an optimal combination leading to a maximized performance of the given system. Two of the contributing factors to the degree of usability for such methods are the number of parameters in use and the sensitivity the system exhibits to slight variations in parameters values. In other words, if our classifier is parameter dependant like most machine learning techniques, we want to know what is the relationship between the parameters, and how do variations in these parameters influence the overall system performance. Fuzzy Gaussian Inference is based on two parameters which, combined with input data, produce a classification with a certain degree of accuracy. The first parameter is the evaluation of the "relative size" of our sample. Intuitively it could be defined as the ratio of the correct "guard" movements the learning sample represents over the range of all possible correct "guard" movements. This number would be a percentage lying in the interval $]0, \kappa[$ where κ is the percentage of values within 8 standard deviations. This ratio is transformed into a z-score n. To be more precise, this ratio represents the average over $x \times m$ dimensions of the area under the bell curve between $\mu - n\sigma$ and $\mu + n\sigma$ in terms of the cumulative normal distribution function ϕ given by:

$$\phi(n) - \phi(-n) = 2\phi(n) - 1 = erf\left(n/\sqrt{2}\right) \tag{4}$$

where $erf()$ is defined as the error function such that:

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \tag{5}$$

The z-score n can therefore be deduced from the parameter by using the inverse error function. The second parameter is a ratio representing the membership threshold in use with the classifier. A membership threshold of 0.95 means for example that we are interested in identifying all frames which have a membership score $\geq 95\%$ of the fuzzy membership function "Guard". When classifying different types of movements, for a given specific accuracy, there seems to be a mathematical relationship between the parameter ϕ and the membership threshold t. Let g be the function that maps the parameter ϕ (an estimation of the relative-size of the learning sample) to the membership threshold t for a given accuracy such that : $g(\phi) = t$. One can observe that for any parameter ϕ, it seems that: $\dot{g}(\phi) < 0$, that is to say that the slope of the function g is always negative. For a given accuracy, the threshold t seems to vary as a function of ϕ following a general curve with an equation of the form:

$$t = \delta + 1/(\gamma \times \log \phi) \tag{6}$$

where δ and γ are constants linked to the dataset considered. Using the concept of elasticity to evaluate if the threshold t is ϕ-elastic, it becomes noticeable that the elasticity is poor when using a very high ϕ value (superior to 0.95). The maximum elasticity is obtained when ϕ is between 40 and 95%. This means that in our data set, the variations of the ϕ parameter are more likely to influence the threshold t if ϕ is kept between 0.4 and 0.95. Regarding the relationship between accuracy and parameters, the accuracy seems to falter with higher values of ϕ. This makes sense because, our sample being of limited size, over-estimating its relative-size will damage the accuracy of the classifier. The loss in accuracy is determined as a function of ϕ. When classifying a guard, the error is rising with over-estimation of ϕ up to a maximum of 10% which is relatively reasonable.

5 Experiment and Results

5.1 Apparatus and Participants

The motion capture data are obtained from a Vicon Motion Capture Studio with eight infra-red cameras. The motion recognition is implemented in MATLAB 2007 on a single machine: a PC with an Intel core duo 2Ghz with 2 Gigs of RAM. An additional MATLAB toolbox [16] is also used for extracting Euler Angles from .BVH files. Three male subjects, aged between 18 and 21, of light to medium-average size (167cm to 178cm) and weight (59 to 79kgs), all practising boxing in competition at the national level. None of them presented any abnormal gait. Optical Markers were placed in a similar way on each subject to ensure a consistent motion capture.

5.2 Procedure

The motion capture data is obtained from several subjects performing each boxing combination four times. There are twenty-one different boxing combinations,

each separated by a guard stance. These are performed at two different speeds (medium-slow and medium fast). We extract a fuzzy membership function template corresponding to a specific stance (e.g. a Guard) from various samples. First all three participants are used to learn and to test how well the system recognizes some of their Guard stances. Then, an evaluation is done to see how the system cope to learn from only two participants, and test how well it recognize stances from a third different participant. The accuracy of the system is examined when learning to recognize five different boxing stances simultaneously. At first there is an evaluation on how accurately each frame is classified individually. Then, Genetic Programming is used to generate fuzzy rules that consider groups of frames and their relative positions. The inputs for each given time frame are the seven membership scores of each known move. These membership scores s_i are re-scaled when by fine-tuning the thresholds t_i linked to each input i the following way:

$$s_i = (s_i - t_i) \div / (1 - t_i) \tag{7}$$

Fuzzy rules are generated using Strongly-Typed Genetic Programming (a specific Python based open source package has been built for this purpose [12]). The GP system evolves rules of the type *If Then Replace X by Y* that are applied to qualitative output of each frame.

The rules are defined by the GP terminal and function sets, and specific constraints that dictate the structure of generated trees. One individual consists of four interconnected *If Then Replace X by Y* rules(see Figure 3). They produce as output groups of frames with modified first membership scores (see Figure 4). Each *If Then Replace X by Y* statement uses specific operators to identify groups of frames and replace their best motion membership score by a different one. e.g. *If Then Replace X by Y(membership_2(guard),jab,cross)* replace the "jab" first membership score with "cross" in groups of frames defined by a "guard" second best membership score. The first branch of a *If Then Replace X by Y* node will contain the logic rules of the if statement. The last two branches will contain terminal nodes referring to the motions being replaced. The children nodes of the if statement contain logical operators(e.g. and, or, not) and operators capturing movement duration and therefore velocity. e.g. *is_short* expresses a duration of less than 5 frames. There are also operators returning the first, second and third best membership scores of a motion (e.g. *membership_2(left_hook)* returns groups of frames with the second best membership score for a motion as a left hook) and operators allowing the recognition of combinations of moves (e.g. *left_2(guard,jab)* returns groups of frames preceded in order by a guard and then a jab motion).

There can be seven different types of moves, therefore seven possible qualitative outputs for a group of frames. Each rule f_j can be seen as a function of the form:

$$\{f_j : X \to X | X = 1, 2, ..., N\} N = 7, \text{ and } j = 1, 2, .., 4 \tag{8}$$

Each rule produces an output which is used in turn as an input for the next rule. This means that one set of 4 rules is in fact a function composition of the type:

$$\{f_1 \circ f_2 \circ f_3 \circ f_4(X) : X \to X | X = 1, 2, ..., N\} N = 7 \tag{9}$$

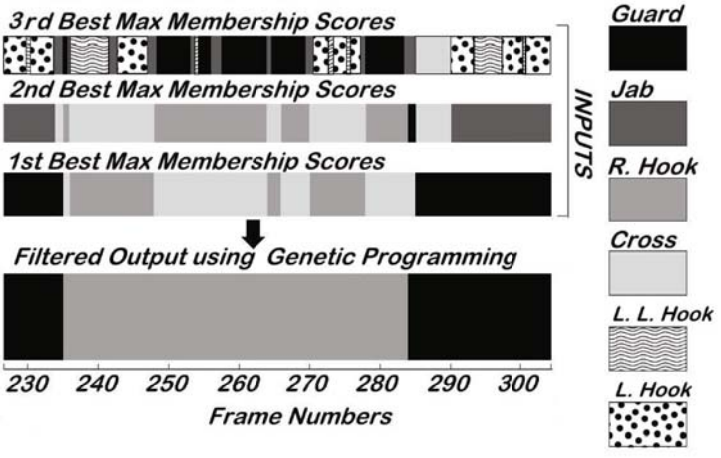


Fig. 4. Fuzzy Rule-Based Improved Classification of a Right Hook Movement

The fitness of set of rules is evaluated by looking at the accuracy the classification of groups of frames depending on their relative positions. This "context-aware" accuracy (as opposed to the "short-sighted" accuracy of an individual frame) is evaluated by summing the overall number of frames that differ from the classification made by a human observer. The Genetic Programming system use a population size of 1000. Its crossover, mutation and reproduction rates are respectively 50%, 49% and 1%.

5.3 Results and Discussion

Figure 5[13] shows a comparison between the accuracy of Fuzzy Gaussian Inference(FGI) and a standard Gaussian Mixture Models(GMM) algorithm when classifying seven different stances (Guard, Jab, Cross, Lower Cross, Right Hook, Left Hook, and Lower Left Hook). The accuracy of FGI for individual frames has been measured and has shown satisfying results in previous work [13]. This

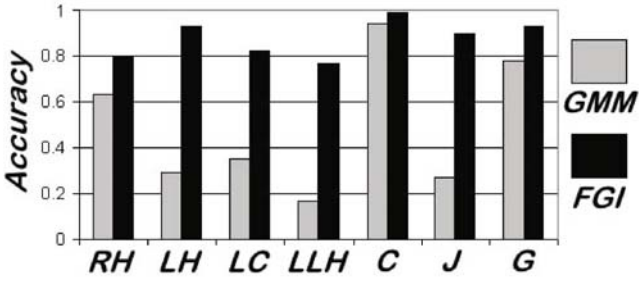


Fig. 5. Comparing accuracy on seven stances: GMM versus FGI

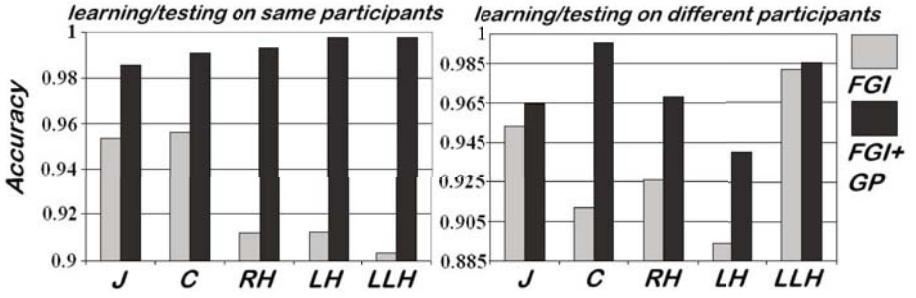


Fig. 6. Comparing "context-aware" accuracy: FGI versus FGI+GP

paper focuses on the accuracy of the classifier regarding frames depending on their relative positions. We compare this "context-aware" accuracy (as opposed to the "short-sighted" accuracy of an individual frame) of FGI to a mixture of FGI and Genetic Programming(see Figure 6) over five different moves (Jab, Cross, Right Hook, Left Hook, and Lower Left Hook) performed by each one of all three individuals (data are 3-fold validated). Results show the accuracy when the same individuals are used for learning and testing and when the individuals used for learning are different from the individuals used for testing. Generally speaking, no assumptions are made about the prevalence of the positions of certain joints. We only consider the general average of membership scores of all joints. If the system can cope with this minimal information, it is reasonable to assume it would also give satisfaction with extra information about joint positions and coupling. There is potential for further improvement at this level. A t-test shows with 95% confidence that FGI seems to be significantly more accurate than the GMM-based one (besides, it is worth noting that it has a general average individual frame accuracy of 87.71% while the GMM algorithm is 49% accurate). Another t-test confirms with 95% confidence that the mixture of FGI and Genetic Programming performs significantly better than FGI alone, even with as little as four rules in total. Although the models performed well when the individual concerned formed part of the training group, the classifier performance worsened significantly when they were removed. Despite this phenomenon in line with previous findings [27] [6], it is worth noticing that the association of FGI and GP still shows consistently better results than FGI on its own, while keeping the ability of FGI to learn from small data sets without pre-processing.

6 Conclusion

The proposed combination of Fuzzy Gaussian Inference and Genetic Programming can learn and classify successfully complex real 3d human motion capture data while outperforming a GMM based classifier. There are several distinct advantages to this technique. First, there is no need for pre-processing the data

(which can be a problem when using techniques such as GMM). Also, static models can be obtained from very few examples, and parameters can be set to tailor the precision of every model to the quantity of available data. The addition of a GP filter consistently improves the "context-aware" accuracy of the classifier. Having validated our method on a real-life data set, the next step is to prioritise the recognition of new moves from partial information. Future work might include a robot kinematics[17][18][5] representation system to deal with the occluded data.

Acknowledgements

The project is funded by EPSRC Industrial CASE studentship and MM2G Ltd under grant No 07002034. Many thanks to Portsmouth University Boxing Club and to the Motion capture Team: A. Counsell, G. Samuel, O. Seymour, I. Sedgebeer, D. McNab, D. Shipway and M. Mitrofanov.

References

1. Aggarwal, J.K., Cai, Q., Liao, W., Sabata, B.: Articulated and elastic non-rigid motion: A review. In: Proc. IEEE Workshop on Motion of Non-Rigid and Articulated Objects, pp. 2–14 (1994)
2. Anderson, D., Luke, R.H., Keller, J.M., Skubic, M.: Modeling human activity from voxel person using fuzzy logic. IEEE Transactions on Fuzzy Systems (accepted for publication, 2009)
3. Bobick, A.F., Wilson, A.D.: A state based technique for the summarization and recognition of gesture. In: International Conference on Computer Vision, pp. 382–388 (1995)
4. Chan, C.S., Liu, H., Brown, D.J.: Recognition of human motion from qualitative normalised templates. Journal of Intelligent and Robotic Systems 48(1), 79–95 (2007)
5. Chang, C.S., Liu, H.: Fuzzy qualitative human motion analysis. IEEE Transactions on Fuzzy Systems 17(4), 851–862 (2009)
6. Darby, J., Li, B., Costen, N.: Human activity recognition: Enhancement for gesture based game interfaces. In: 3rd Int. Conf. on Games Research and Development, CyberGames (2007)
7. Devi, B.B., Sarma, V.V.S.: Estimation of fuzzy memberships from histograms. Inf. Sci. 35(1), 43–59 (1985)
8. Favre, J., Aissaoui, R., Jolles, B.M., Siegrist, O., de Guise, J.A., Aminian, K.: 3d joint rotation measurement using mems inertial sensors: Application to the knee joint. In: ISB-3D: 3-D Analysis of Human Movement, Valenciennes, France, June 28-30 (2006)
9. Frantti, T.: Timing of fuzzy membership functions from data. Academic Dissertation - University of Oulu - Finland (July 2001)
10. Guo, Y., Xu, G., Tsuji, S.: Understanding human motion patterns. In: International Conference on Pattern Recognition, pp. B:325–329 (1994)
11. Iokibe, T.: A method for automatic rule and membership function generation by discretionary fuzzy performance function and its application to a practical system. In: Proceedings of the First International Joint Conference of the North American Fuzzy Information Processing Society Biannual Conference, pp. 363–364 (1994)

12. Khoury, M.: pystep or python strongly typed genetic programming, http://pystep.sourceforge.net/
13. Khoury, M., Liu, H.: Fuzzy qualitative gaussian inference: Finding hidden probability distributions using fuzzy membership functions. In: IEEE Workshop on Robotic Intelligence in Informationally Structured Space, RiiSS 2009 (2009)
14. Kim, C., Russell, B.: Automatic generation of membership function and fuzzy rule using inductive reasoning. In: Third International Conference on Industrial Fuzzy Control and Intelligent Systems, 1993. IFIS 1993, pp. 93–96 (1993)
15. Kim, J., Seo, J., Kim, G.: Estimating membership functions in a fuzzy network model for part-of-speech tagging. Journal of Intelligent and Fuzzy Systems 4, 309–320 (1996)
16. Lawrence, N.D.: Mocap toolbox for matlab/, http://www.cs.man.ac.uk/~neill/mocap
17. Liu, H.: A fuzzy qualitative framework for connecting robot qualitative and quantitative representations. IEEE Transactions on Fuzzy Systems 16(8), 1522–1530 (2008)
18. Liu, H., Brown, D.J., Coghill, G.M.: Fuzzy qualitative robot kinematics. IEEE T. Fuzzy Systems 16(3), 808–822 (2008)
19. Nieradka, G., Butkiewicz, B.S.: A method for automatic membership function estimation based on fuzzy measures,'. In: Melin, P., Castillo, O., Aguilar, L.T., Kacprzyk, J., Pedrycz, W. (eds.) IFSA 2007. LNCS (LNAI), vol. 4529, pp. 451–460. Springer, Heidelberg (2007)
20. Pentland, A.P., Oliver, N., Brand, M.: Coupled hidden markov models for complex action recognition. In: Massachusetts Institute of Technology, Media Lab (1996)
21. Remagnino, P., Tan, T.N., Baker, K.D.: Agent orientated annotation in model based visual surveillance. In: International Conference on Computer Vision, pp. 857–862 (1998)
22. Sanghi, S.: Determining membership function values to optimize retrieval in a fuzzy relational database. In: Proceedings of the 2006 ACM SE Conference, vol. 1, pp. 537–542 (2006)
23. Simon, D.: H infinity estimation for fuzzy membership function optimization. Int. J. Approx. Reasoning 40(3), 224–242 (2005)
24. Thingvold, J.: Biovision bvh format (1999), http://www.cs.wisc.edu/graphics/Courses/cs-838-1999/Jeff
25. Wang, L., Hu, W., Tan, T.: Recent developments in human motion analysis. Pattern Recognition 36(3), 585–601 (2003)
26. Yacoob, Y., Black, M.: Parameterized modeling and recognition of activities. In: Sixth International Conference on Computer Vision, 1998, pp. 120–127 (1998)
27. Yamato, J., Ohya, J., Ishii, K.: Recognizing human action in time-sequential images using hidden markov model. In: IEEE Computer Vision and Pattern Recognition, pp. 379–385 (1992)
28. Zadeh, L.: Fuzzy sets. Information and Control 8, 338–353 (1986)

Pointing Gestures for a Robot Mediated Communication Interface

John-John Cabibihan[1,2], Wing Chee So[3], Medi Nazar[2], and Shuzhi Sam Ge[1,2]

[1] Social Robotics Laboratory, Interactive and Digital Media Institute
[2] Department of Electrical and Computer Engineering
[3] Department of Psychology,
National University of Singapore, Singapore
{elecjj,psyswc,medinazar,elegesz}@nus.edu.sg

Abstract. This paper asked whether the pointing gesture accompanying with speech would facilitate comprehension of spatial information in the videoconference communication. Ten adults participated in our study and communicated with the experimenter over Skype (Skype Technologies, Luxembourg). The experimenter described the spatial layout of items in a room to the participants in two conditions – dynamic and static. In the static condition, the notebook was not moving; in the dynamic condition, the notebook moved around with the arms pointing to abstract spatial locations that represented the locations of items in the room. The movement was done by putting the notebook on the three-wheeled Wi-Fi enabled device that was equipped with two artificial arms and was controlled by the experimenter over the Internet. At the end of each description, the participants were asked to lay out the items properly. Reaction times and accuracy rate were recorded. The findings showed that the accuracy rate was higher in the dynamic condition than in the static condition. In addition, the response time was faster in the dynamic condition than in the static condition. It turned out that pointing gestures facilitated the speech comprehension of spatial information.

Keywords: Videoconferencing, pointing gesture, communication modality, robot-mediated interaction.

1 Introduction

Communication through Internet is the most common way of keeping in touch with our friends and family members from a distance. Email, chat, videoconference, and social networking sites (SNS) are becoming part of our daily lives. In face to face communication, it is very natural for us to combine our different senses to fully express ourselves and to fully understand what the other person is saying. We often use our audio-visual sensory system and incorporate non-verbal cues such as gestures to convey information using these multiple channels unconsciously to different degrees. In mediated communication such as videoconferencing, nonverbal cues are either limited or filtered thus the richness of information sent/received decreases.

There are various existing works focusing on different aspect of mediated communication. In [1], Hwang and Park mentioned that computer mediated communication

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 67–77, 2009.

(CMC) is a medium with low social presence especially text-based CMC environments since non-verbal cues like gestures, direction of gaze and posture are missing, where social presence here is defined as the ability of communicating parties to project themselves, socially and emotionally, as real people through a medium of communication [2]. In [3], the difference among communication modes (text chat, audio, audio–video, and avatar) in an Internet-based collaboration were investigated and significant differences were found between text-based chat and all other communication modalities. Although some researchers [4] argued that the quality of communication is based on the medium, Fish and colleagues [5] showed that video does not significantly improve telecommunication as compared to audio with respect to efficiency or user satisfaction. Experiments were conducted between human-robot interaction [6] to test the effect of social presence. These current works motivate us to find other means of increasing the level of social presence in existing communication through the Internet.

In this paper, we added pointing gesture to a typical videoconference over Skype. To produce such gesture, we thought of an embodiment of a person (a robot) on the other side of communication channel that can be controlled through the Internet. Furthermore, we used memory test in our experiment to investigate the effect of such communication compare to a typical videoconference. In our experiment, ten adults participated in our study where the spatial layout of items in a room is described by the experimenter with and without pointing gesture to each participant. Then they were asked to layout the items properly. Reaction times and accuracy rate were recorded.

In [7], Kita argued that pointing gesture lessens the confusion between the *left* and *right* concepts. Thus, we hypothesized that adding pointing gesture will increase the accuracy rate of the participants and also, their response will be faster.

The succeeding sections are organized as follows: Section 2 presents the methodology that we used in our experiment, Section 3 shows the results, Section 4 contains the discussion and analysis, and finally, the conclusion in presented in Section 5.

2 Methodology

2.1 The Robot Interface

The aim of the experiment is to find out if there will be a difference when we add pointing gesture in describing the spatial location of the objects seen by the person in a room over a Skype-based communication. In order to do this, we built a robot interface that consists of three components: (1) Rovio (WowWee Group Limited, Hong Kong), a three-wheeled Wi-Fi enabled robot; (2) mini notebook put on top of Rovio; and (3) a static pointing arm attached to the combined Rovio and mini notebook, see Figure 1a. The movement of the robot interface is controlled over the Internet through a software joystick installed in another Wi-Fi enabled computer (Computer A), see Figure 1b. During the experiment, the face of the experimenter who is communicating with the participant appears on the monitor of the notebook through a Skype videoconference.

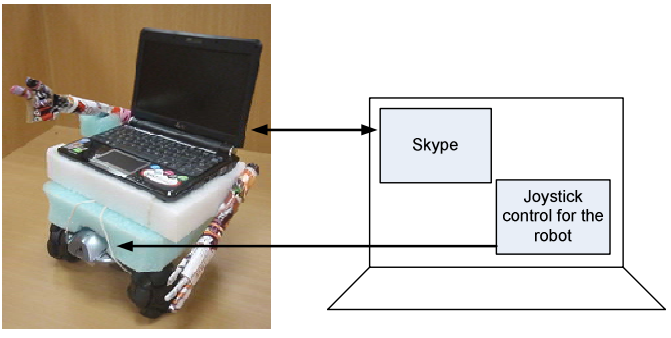


Fig. 1. The Robot-Mediated Communications Interface

2.2 Communication Set-Up

We used two communication set-ups in our experiment, first is the typical videoconference over the Internet which we called *static mode* (Figure 2a) and the second is the one with additional robot movement called *dynamic mode* (Figure 2b). In static mode, the experimenter and the research subject will communicate through a typical videoconference (without movement) using Computer A and the mini notebook while in dynamic mode the robot interface will be used to create the pointing gesture, the unidirectional arrow in Figure 2b illustrates this additional communication modality towards the research subject. In our experiment, we studied the difference of these two set-ups: static and dynamic modes.

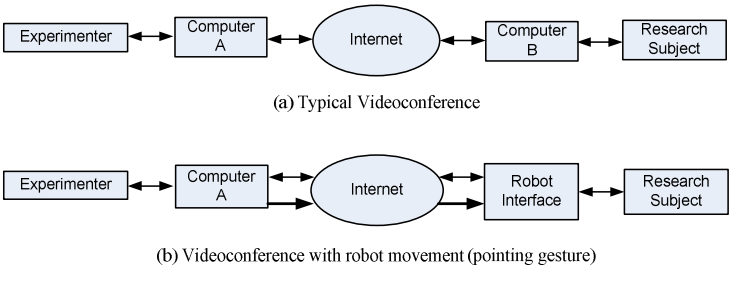


Fig. 2. Communication set-up for *static* and *dynamic* modes

2.3 Experimental Set-Up

Figure 3 shows the experimental set up in the laboratory room. The robot interface is placed on top of a wide table to give space for its movement in a dynamic condition facing the subject who sits on the chair. The dashed area that surrounds the subject's chair is the empty part of the room where the items described in the scripts are located. Ten layout sheets and ten sets of small cut pieces of papers needed for the memory test are provided on a separate table (not included in the figure).

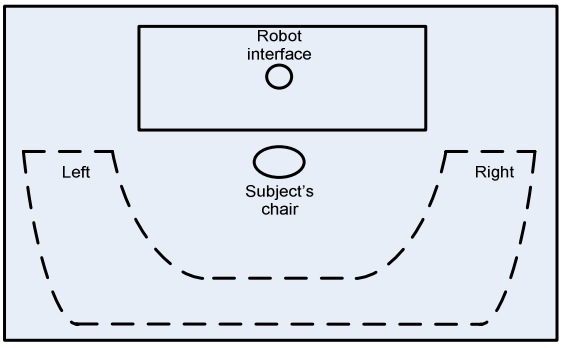


Fig. 3. Experimental Set-up

The layout sheet contains the layout of the room shown in Figure 3 which is basically empty. The items to be laid out on the sheet were written on the small cut pieces of papers. Each set of cut papers has its own color and has a label in a room-mode format (e.g. kitchen-dynamic). These ten sets were arranged in a right to left manner based on the random sequence of the script for each subject.

2.4 The Scripts

For the scripts, we considered five venues that can be found at home such as kitchen, living room, bedroom, study room and kid's playroom. We thought of five items that can be found in each room, then, prepared a description/script of the room with the five items and their specific location in a given room. There are ten scripts, five for the static condition and another five for the dynamic condition, see Appendix. One room is described in each script. Scripts 1 to 5 are the static mode description of the five given rooms whereas scripts 6 to 10 are the description for the dynamic mode, same rooms were used but with different list of items. The first statement in the script varies according to the sequence of delivery which is random for each subject.

In the *static* mode, the locations of the imaginary items in a given room were described using the following keywords: "on your left hand side", "on the right corner of the room". In the dynamic mode, since the robot is moving and pointing towards the direction of its movement, the scripts were shortened and simplified to "on that side" and "on that corner". The word "that" is associated with the location of the item where the robot is facing and pointing at, thus, no need for long description. Cues such as "behind you" and "at your back" remain to avoid ambiguity while the robot is facing to the subject.

2.5 Experimental Procedure

Participants

We invited ten participants in our pilot test, eight were female and two were male whose age is from 22 to 39 with different nationality and field of studies.

Procedure

Upon arrival of the participant, the communication set-up was explained and instructions were given by the experimenter. The experimenter set up the Skype connection between the mini notebook and Computer A which is in the adjacent room. Then, the experimenter entered to the adjacent room leaving the participant and the robot interface facing each other. The experimenter started the communication by giving introductory statements; consequently, the first script was recited. Each script contains the spatial description of five imaginary items memorized by the participant.

After the first script, the participant stood, picked the first set of small cut pieces of paper where the mentioned items were written, and laid out the items in their proper location on the provided layout sheet. After arranging the items on the layout sheet, he/she went back to his/her sit for the next script. The procedure was repeated for ten scripts. Different sequence of the scripts was given to each participant which are all randomized.

We used a timer written in C programming language (ran in Dev-C++ [8]) to measure the *response time* of the participant in each script. The timer starts when the participant gives a signal that he/she will start to lay out the items and it ends when a finish signal is sent by the participant.

After the ten scripts, the experimenter went out from the adjacent room to meet the subject in the experimental room. Then, checked the *accuracy* of the laid out items while having a conversation with the participant about what he/she's thinking regarding the location and the items and why they have mistakes on the item's location. Each script has five items and we have five scripts for each mode (see Appendix), thus, the perfect score per script is 5. The accuracy is calculated separately for static and dynamic mode.

3 Results

Figure 4 shows the response time of each subject in static and dynamic modes. It is evident that subject number 4 and 7 were faster in answering the memory test in dynamic compare to static mode. All subjects were relatively faster in dynamic mode except subject number 2.

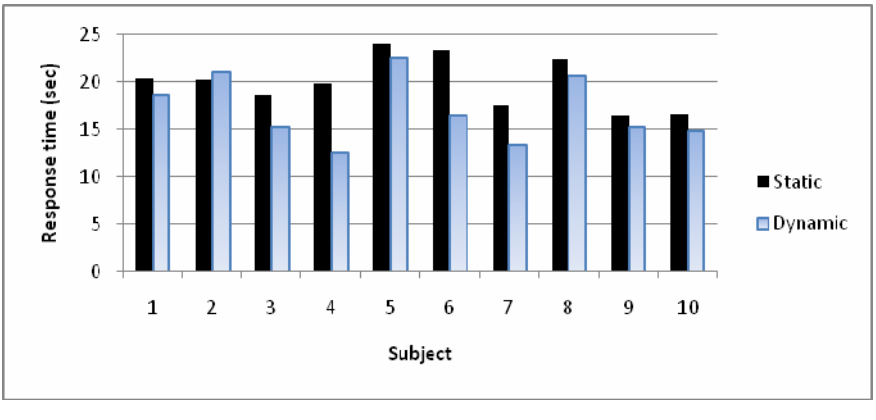


Fig. 4. Subjects' average response time

Note that the average response time varies per subject, which may be due to various factors such as alertness, current mental state of the subject and how they process data in their brain. These factors are out of the scope of this study.

Figure 5 shows the result of the memory test, the accuracy on the vertical axis shows the average correct laid out items of each subject with maximum accuracy level of five. The accuracy level of subject 1, 4 and 7 are much higher in dynamic compare to static mode. All the subjects got higher accuracy in dynamic mode except subject number 6 who got the same accuracy in both modes.

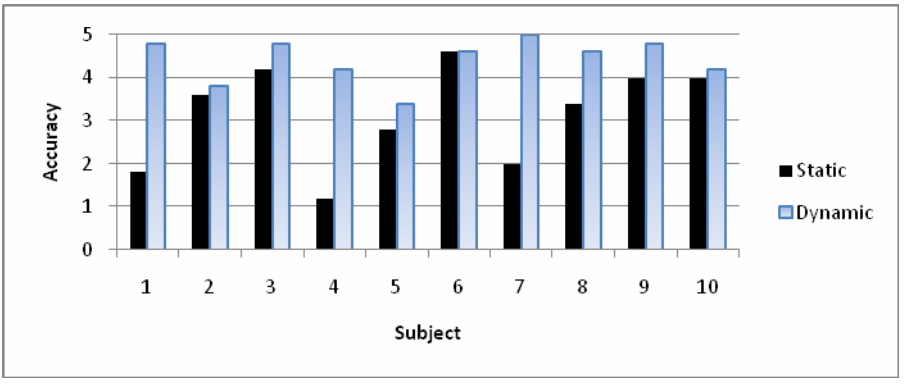


Fig. 5. Subjects' Average Accuracy

The overall response time and accuracy rate for all the subjects are shown in Figure 6. The accuracy rates in static and dynamic modes are 63.2% and 88.4% respectively. It clearly shows that the response time in dynamic mode is faster than in static mode, moreover, the accuracy rate is higher in dynamic mode.

A paired-samples t-test was conducted to compare the response time of the research subject in static and dynamic conditions. The static condition makes use of the typical conversations while the dynamic condition has a robot moving and pointing directions. There was a significant difference at the 5% level in the scores for the static (mean = 19.94, sd = 2.70) and dynamic (mean = 17.04, sd = 3.45) conditions; t (3.56, p = 0.006).

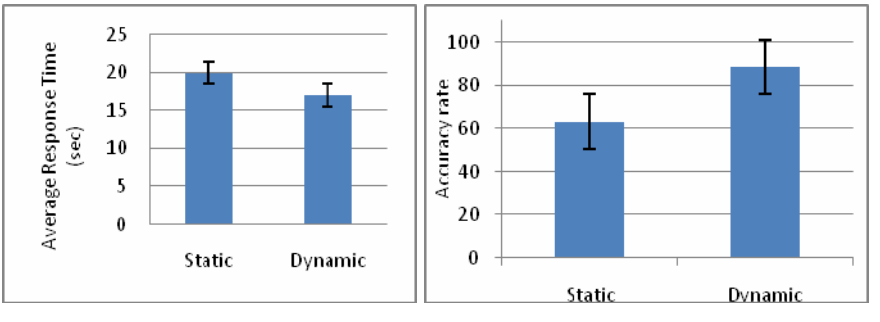


Fig. 6. Overall average response time and accuracy rate

Paired-samples t-test was also conducted to compare the accuracy of the research subject in the memory test for static and dynamic conditions. Likewise, there was a significant difference at the 5% level in the scores for the static (mean = 3.16, sd = 1.16) and dynamic (mean = 4.42, sd = 0.51) conditions; t (-3.194, p = 0.011).

4 Discussion

According to most of the participants, the movement/pointing gesture helped them in recalling the location of the item and the only thing they were memorizing during the *dynamic* condition are the items being mentioned whereas in static condition they have to remember the items at the same time. Subjects no.2 and no.9 mentioned that the movement distracted them. If we look on to the average response time they got, it is indeed confirmed that the movement/pointing gesture did not help subject no.2 in memorizing the items but we cannot say that it has the same effect with subject no. 9 because she had a higher response time during static than dynamic mode (see Figure 4). Most of the remaining subjects mentioned that they were having a hard time memorizing the items in *static* mode and it is confirmed by the results we just have presented in Section 3.

Figure 7 shows the frames from the video during the experiment. Figure 7a shows five frames while the scene is in the kitchen *static* mode description. Frames 1 to 5 were taken after each item was mentioned in the script. From the frames shown, it is apparent that the participant is trying to remember the items and the location being described by merely listening to the experimenter, notice that the position of the robot interface is not changing. In Figure 7b, during the *dynamic* mode description of the kitchen, the participant is following the movement of the robot interface, looking towards the direction where it is pointing while listening to the item being mentioned by the experimenter.

Frame 1 Frame 2 Frame 3 Frame 4 Frame 5

Frames taken from a *static* mode video

Frame 6 Frame 7 Frame 8 Frame 9 Frame 10

(b) Frames taken from a *dynamic* mode video

Fig. 7. Frames taken from the video during the experiment

In terms of accuracy, most of the wrong answers were due to swapped items. One participant was confused about the corner of the room though the layout of the room in Figure 3 was presented to them before the experiment started.

Unfamiliar words were compensated because they were written on the cut papers, so even if the item name is not that familiar to some of the subjects they were able to figure out what it is upon reading the words during the test.

During the experiment, the experimenter used to make a forward movement towards the subject in between scripts to make the interaction interactive and to take away the boredom from the subjects in case there is and to thank them at the end of the tenth script. Most of them smiled and responded positively and some of them felt anxious at the same time. It indicates that the presence of the embodiment strengthens the level of social presence of the person communicating on the other side of communication channel.

5 Conclusion

One of the most commonly used way of communication over the Internet today is videoconferencing, a combination of audio and video. In our experiment, we added another modality to this current set-up. We used an embodiment to add pointing gesture to the typical Skype-based communication. We used memory test to measure the effect of this embodiment.

From the result of our experiment, it appears that the dynamic mode significantly decreases the response time of the subjects and increases their accuracy rate in answering the memory test.

In the memory test, the sense of hearing was used by the subjects during the static mode while in dynamic mode two senses were used, sense of hearing and sight that is hearing the name of the items from the script and seeing the robot interface movement to the direction of the items. The results indicate that using these two senses at the same time to recall object's specific location is better than just using the sense of hearing alone. Thus, the addition of modality which is the pointing gesture movement in our experiment made a significant contribution in current Skype-based communication.

6 Limitations and Future Directions

There are several limitations to our current work. First, we utilized static arms and used the moving robot to point one arm to the direction of the items being described in the scripts. In our future work, we plan to develop moving arms that would point without the robot body's movement. Second, we used the available three-wheeled WiFi enabled device (Rovio) as our robot mediated interface, mounted a mini notebook above it and attached static arms. For the arm to move independently, this must be part of the interface that can be manipulated through the network, thus, the need to develop a robot interface with moving arms.

For our future studies, extending the number of non-verbal cues such as waving of the robot's arm and handshaking will be investigated. This requires us to modify our current experimental procedure or design a new one since our current memory test

might not be applicable to these new gestures. It is also interesting to explore the effect of proximity.

Acknowledgments. The authors would like to thank Dr. Marcelo Ang, Department of Mechanical Engineering-NUS, for lending us the robot Rovio from the Control & Mechatronics Lab. This research was funded by the Social Robots: Breathing Life Into Machine Project (NRF-R705-000-017-279), Media Development Authority of Singapore.

References

1. Short, J.A., Williams, E., Christie, B.: The social psychology of telecommunications. John Wiley & Sons, Ltd., London (1976)
2. Garrison, D.R., Anderson, T.D.: Learning in the 21st Century, p. 49. RoutledgeFalmer, London (2003)
3. Bente, G., Rüggenberg, S., Krämer, N.C., Eschenburg, F.: Avatar-mediated networking: Increasing social presence and interpersonal trust in net-based collaborations. Human Communication Research 34(2), 287–318 (2008)
4. Hwang, H.S., Park, S.: Being together: User's subjective experience of social presence in CMC environments, Beijing. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 844–853 (2007)
5. Fish, R.S., Kraut, R.E., Chalfonte, B.: The VideoWindow system in informal communications. Paper presented at the Proceedings of the Third Conference on Computer-Supported Cooperative Work Los Angeles, CA (1990)
6. Bainbridge, W.A., Hart, J., Kim, E.S., Scassellati, B.: The Effect of Presence on Human-Robot Interaction. In: Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication, Munich, Germany (2008)
7. Kita, S.: Pointing: Where Language, Culture, and Cognition Meet. Lawrence Erlbaum Associates, Inc., New Jersey (2008)
8. Bloodshed Software, `http://www.bloodshed.net/devcpp.html`

Appendix: Scripts

Five scenes were used in our scripts namely the kitchen, living room, bed room, study room and kid's playroom. One room is described in each script. Scripts 1 to 5 are the static mode description of the five given rooms whereas scripts 6 to 10 are the description for the dynamic mode.

Script 1: Scene - Kitchen

Hi, ____(<subject's name>), I would like to introduce to you my kitchen.
There is a stove on your right side.
The sink is adjacent to the stove.
There are piles of dishes beside the sink.
The refrigerator is just right behind you and there are notes posted on its door.
And on your left side, there is the dining table surrounded by chairs.

Script 2: Scene - Living Room Now, let me describe my living room to you. The *door* is on your left hand side. Next to the door, I can see an *aquarium* on top of wooden shoe rack. At your back, there is a *small table with a dvd player* put on its top. There are *two couches* adjacent to each other on the right side wall. There is a *small table* right in front of the couches.
Script 3: Scene - Bed Room Let's go to the bedroom There is an *exit door* at your left. At the left corner there's a *computer desktop* on top of the computer table. There are *five paintings* hanging on the wall behind you. Next to the paintings, there's a *window* with open horizontal blinds near the right corner of the room. And there's a *bed* opposite to the window.
Script 4: Scene - Study Room Let's go to my study room. The *entrance door* is on your right. There's a *computer and a printer* on the right corner of the room. I can see an *office table* with two chairs in front of it behind you. On the left corner, there is a *wooden shelf* full of books and documents. Next to the shelf, there is a *glass wall* with vertical blinds.
Script 5: Scene - Kid's Playroom Finally, let's go to the kid's playroom. There is a *small table* surrounded by small chairs on your left. At the left corner there is a *shelf* with colourful books in different sizes. Adjacent to the shelf, behind you, there is a *stackable bin* full of toys. A *painting on a canvas* stands on the right corner. There is an *inflatable couch* on your right.
Script 6: Scene - Kitchen Hi, ____(<subject's name>), I would like to introduce to you my kitchen. There is a *trashcan* on that side. Then, there is the *barbecue stand* beside. Next to that is a *water dispenser*. The *cabinet* is just right behind you On this side, there is a *breakfast nook* with chairs beside it.
Script 7: Scene - Living Room Now, let me describe my living room to you. The *fireplace* is on this side. I can see a *piano* on that corner. At your back, there is a *window* with colourful curtain. On that corner, I can see *two couches* adjacent to each other. And there is a *coffee table* right in front of the couches.
Script 8: Scene - Bed Room Let's go to the bedroom. The *bed* is on this side. Then, on that corner I can see *clothes* hanging. There is a *lampshade* behind you. There are *candles* on top of a rectangular table on that corner. And there's a *television* on this side.

Script 9: Scene - Study Room
Let's go to my study room.
There is the *study table*!
There's a *bookshelf* on that corner.
At your back, I can see *medals & certificates* hanging on the wall.
On that corner, there is an *opened dictionary* on its stand.
I can see the *exit door* on this side.

Script 10: Scene - Kid's Playroom
Finally, let's go to the kid's playroom.
Oh, I can see different sizes of *dolls* on the floor.
And at that corner, there are *coloring books* on top of a small table.
Behind you, there are *artworks* posted on the wall.
A *drawing board* stands on that corner.
On this side I can see a *toy house*.

View-Invariant Human Action Recognition Using Exemplar-Based Hidden Markov Models

Xiaofei Ji[1,2] and Honghai Liu[1]

[1] The Institute of Industrial Research
The University of Portsmouth, UK
[2] The College of Automation Engineering
Nanjing University of Aeronautics and Astronautics, China

Abstract. An exemplar-based Hidden Markov Model is proposed for human action recognition from any arbitrary viewpoint image sequence. In this framework, human action is modelled as a sequence of body poses (i.e., exemplars) which are represented by a collection of silhouette images. The human actions are recognized by matching the observation image sequence to predefined exemplars, in which the temporal constraints were imposed in the exemplar-based Hidden Markov Model. The proposed method is evaluated in a public dataset and the result shows that it not only reduces computational complexity, but it also is able to accurately recognize human actions using single cameras.

1 Introduction

Human action recognition from video is an important and challenging research topic in human-robot interaction with many potential applications involving human motion understanding such as visual surveillance, content based video retrieval, athletic performance analysis *etc.*, in which the actions are often observed from arbitrary camera viewpoints, for instance as shown in Fig. 1 [1]. Hence it is desired that the recognition algorithms exhibit some view invariance, that is to say, an action should remain recognizable while the viewpoint of the camera is changing [2, 3]. The viewpoint issue in visual-based human action recognition has attracted more and more research attentions [4, 5, 6, 7, 8, 9, 10], and those methods can be classified into two categories: template-based methods and state-space methods. The advantages of template-based methods are the low computational cost and the simple implementation, however they are usually more sensitive to noise and variance of movement duration. Rao *et al.* [11] presented a view-invariant computational representation of human action to capture dramatic changes in the speed and direction of a motion trajectory, which was presented by spatio-temporal curvature of a 2D trajectory. Parameswaran and Chellappa [12, 13] handled the problem of view-invariant action recognition based on point-light displays by investigating 2D and 3D invariant theory. Furthermore a representative spatio-temporal action volumes (STV) was proposed by Yilmaz and Shah [14, 15] to achieve view-invariant action recognition. The above methods are all based on the assumption that point correspondences are

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 78–89, 2009.

Fig. 1. A surveillance scene in CMU dataset [1]

available in parts of images. So their applications are limited to some special occasions. Another approach was proposed by Blank and Gorelick [16] that represented human actions as three-dimensional shapes induced by the silhouettes in the space-time volume. This method extracts space-time features that do not require computing point correspondence. This method is not fully view-invariant, it is however robust to large changes in viewpoint(up to 54 degrees). The key to template-matching approaches is finding the vital and robust feature sets, and then an action recognition may be simply considered as a classification problem of those feature sets.

On the other hand the methods based on state-space models, *e.g.* Hidden Markov Models (HMMs), have been widely applied to express the temporal relationships inherent in human actions [17, 18, 19, 20, 21]. These methods usually define each static posture as a state. These states are connected by certain probabilities, any motion sequence is considered as a tour going through various states of these static poses. Lv and Nevatia [22] presented an example-based view-invariant action recognition system that explored the use of contextual constraints. Those constraints were inherently modelled by a novel action graph model representation called *Action Net*. Each link in the action net specified the possible transition of the key poses within an action class or across different action classes. This approach was demonstrated on challenging video sets consisting of 15 complex action classes. Owing to the complexity of the action net, modelling transitional probability for each link is not applicable in practice. So, this action net representation neglects the transitional probability. A similar work on exemplar-based HMMs was proposed for view-invariant human motion analysis [23]. This model can account for dependencies between three dimensional exemplars, *i.e.* representative pose instances and image cues. Inference is then used to identify the action sequence that best explains the image observations. This work uses a probabilistic formulation instead of the deterministic linked action graph introduced in [22], it can handle uncertainties inherent to actions performed by different people and different styles. However the learning process is relatively complex.

Inspired by the works of [22,23], we propose a simplified view-invariant human action recognition framework using exemplar-based HMMs. In our framework, each human action is modelled by a sequence of key poses, which are represented by a set of 2D silhouette images captured from multiple camera viewpoints. Action recognition is achieved using Viterbi search on the exemplar-based HMMs. Furthermore, we use the silhouette distance signal as shape feature, which is efficiently obtained. Hence the reduction of the computational complexity is achieved in model training and testing.

The remainder of this paper is organized as follows. Section 2 overviews the proposed framework. Exemplar selection and representation are introduced in Section 3. Action modelling and recognition are proposed in Section 4. The results are presented and discussed in Section 5. The paper is concluded in Section 6 with analysis on future research challenges and directions.

2 The Framework of the Proposed Approach

The framework is shown in Fig. 2, in which each human action is modelled by an exemplar-based HMM. Each state in the HMM accords with a key pose exemplar which is represented by a collection of silhouette images observed in 36 different viewpoints under the assumption that only the orientation of an actor around the vertical axis is variable.

In the process of model training, the key poses of each action class are extracted from a given small set of action sequences from Inria Xmas Motion Acquisition Sequences(IXMAS) dataset by clustering the 3D visual hull representation. This dataset is a multiple-actor and multiple view dataset. The 3D key poses are projected into multiple view 2D images using the camera projection principle. Finally the dynamics, *i.e*, transfer matrices are learned by utilizing

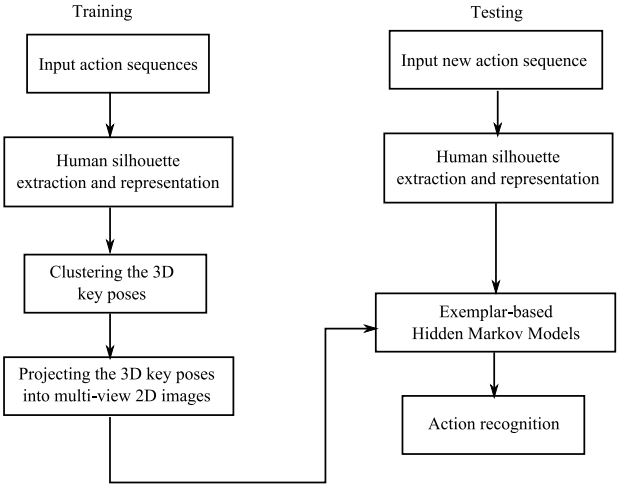


Fig. 2. The framework of proposed approach

the expectation maximization (E-M) algorithm from the given action sequences. On the other hand, the exemplar-based HMMs are used to find the most likely sequence of actions seen in a single viewpoint video in recognition phrase. At each frame the observation probability is computed based on shape similarity between the observation and the exemplar. In this process, we force the viewpoint to remain constant or change smoothly from one key pose to the next consecutive pose. The steps in details are provided in the following sections.

3 Exemplar Selection and Representation

In the proposed framework, an action is modelled as a sequence over a set of key poses, the exemplars, which are described by multiple view silhouette images. It is evident that collecting multiple view pose images from real experimental conditions is a difficult task. So we directly obtain multiple view pose dataset by projecting 3D key poses into multiple view 2D silhouette images where the camera parameters are known [23]; 3D key poses are extracted by k-means clustering.

3.1 Key Pose Extraction

There are some popularly used methods of extracting key poses including motion energy minima and maxima [22], optical flow magnitude of foreground pixels extremum [4], k-means clustering and wrapper method [23]. In the framework, four key poses are extracted for each action by using k-means method from a small set of action sequence of IXMAX dataset, in which human poses in every frame are represented in 3D visual hulls that have been computed using a system of 5 calibrated cameras as shown in Fig. 3.

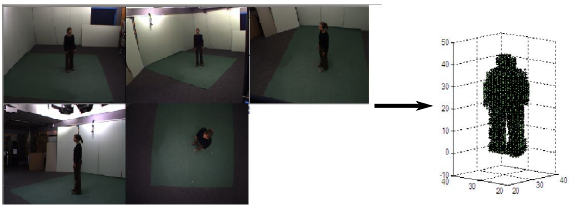


Fig. 3. 3D visual hull presentation

3.2 2D Silhouette Image Generation

Due to the fact that a camera is a mapping device between a 3D world (object space) and a 2D image, we project 3D visual hull presentation into multiple view 2D silhouette images using the camera projection principle. It is only considered in the paper that the orientation of a person around the vertical axis is variable. The orientation angle is discretized into 36 equally spaced angles within [0,2π],

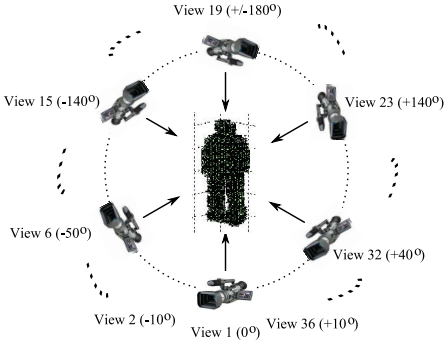


Fig. 4. Multi-views for a given key pose

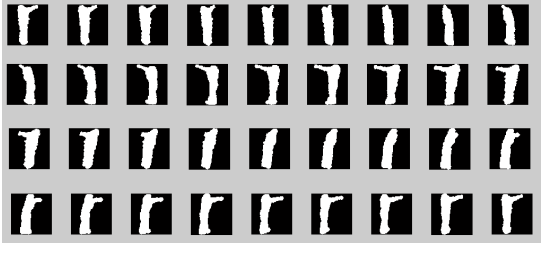


Fig. 5. Multi-view silhouette images of a given key pose

the multi-views for a given key pose are shown in Fig. 4; multiple view silhouette images of a given key pose are provided in Fig. 5.

The silhouette images are centred and normalized in order to contain as much foreground as possible, it leads to the fact that the motion shape is not distorted and all input frames are equal dimensions.

3.3 Contour Shape Feature

There are some representative shape features of the silhouette image in the previous papers, such as shape context descriptor [24], width feature [25]. We describe the silhouette images using the silhouette distance metric, in that it not only can capture both structural and dynamic information for an action, but it also can be efficiently obtained [26]. An example of the distance signal is shown in Fig. 6, which is generated by calculating the Euclidean distance between the centre of the mass points and each edge point of the silhouette images in clockwise direction.

In order to obtain image scale and rotation invariance, firstly the principle axis of the silhouette is computed before computing shape feature, the rotation angle of the principle axis is compensated so that the principal axis is vertical.

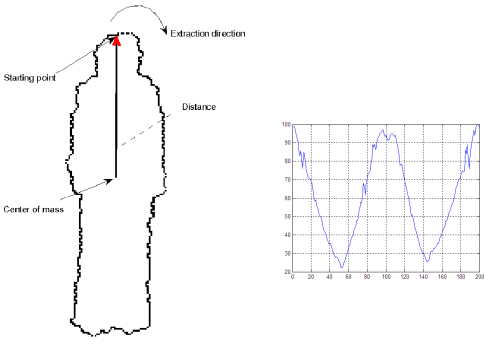


Fig. 6. Contour shaper feature of the silhouette image

Then the contour of the silhouette image is uniformly re-sampled to 200 edge points, and the distance is normalized into [0, 100].

4 Action Modelling and Recognition

Each action class is modelled from multiple-person and multiple-view datasets by learning an exemplar-based HMM in the paper. Each state in this graph model corresponds to one key pose, which is represented as a set of contour shape features of multiple view 2D silhouette images. Action recognition is achieved by using the standard HMM algorithm, *i.e.*, the maximum a posteriori estimate, to find the most likely sequence of actions seen in a single viewpoint video.

4.1 Exemplar-Based Hidden Markov Model

The exemplar-based Hidden Markov Model has been used in action recognition to solve the problem that the space of observations is not Euclidean [17, 23]. The novelty of the exemplar-based Hidden Markov Model is that mixture density functions are not entered on arbitrary means values, but centred on prototypical data instances, the exemplar. A representative graphical model is shown in Fig. 7.

An action class is modelled as a hidden state sequence Q, *e.g.* a motion sequence in a pose space. At each time, exemplar x_t is drawn from the motion sequence Q. Observation y_t derived from a geometric transformation of exemplars. α_t is transformation parameters. Learning this probability model involves learning the exemplars from the training set, learning the dynamics, *i.e*, transfer matrix in the form of $P(x_t|x_{t-1})$, learning the exemplar probability given the state $P(x_t|q_t)$.

4.2 Action Modelling

Exemplars estimation is no longer coupled with the HMMs estimation in the proposed work. Exemplars firstly are extracted using the method introduced in

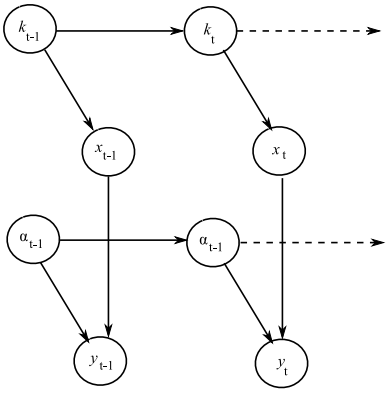


Fig. 7. Graphical model for exemplar-based tracking [17]

Section 3. Then the exemplars are represented by a set of contour shape feature of multiple view 2D silhouette images, which don't deterministically linked to motion states q. So there is no coupling between the states and the exemplars. Under this condition, exemplar probability given the state $P(x_t|q_t)$ need not estimate. Only the dynamic $P(x_t|x_{t-1})$ is learned using the traditional expectation maximization(E-M) approach, in which each exemplar can be treated as a discrete symbol.

4.3 Action Recognition

Separate action models H_c are learned for each action class $c \in 1, ..., C.$, its observation probability $p(y|x)$ is defined as

$$p(y|x = i) = \frac{1}{Z_i} \exp(-d(y, x_i)/\sigma_i^2) \tag{1}$$

Where d is a distance function between the observation and the exemplar, where the Euclidean distance is employed. The variance σ_i and the normalization constant Z_i are selected as the method proposed in [17], *i.e.*, $\sigma_i = \sigma, i = 1, ..., M$ (M is the number of exemplars).

A sequence of observation Y is recognized with respect to the maximum a posteriori estimate:

$$p(Y) = \arg\max_c p(Y|H_c)p(H_c) \tag{2}$$

The joint probability of observation sequence $p(Y|H_c)$ can be obtained by the Viterbi algorithm. The prior probabilities $p(H_c)$ is used as a uniformly distributed prior in the framework.

5 Experiments

We demonstrate the proposed framework on a public dataset, IXMAS dataset. It contains 12 actions, each of which was performed 3 times by 12 actors(taken

Fig. 8. The IXMAS database [9]

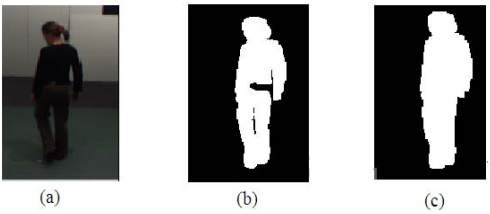


Fig. 9. An example of defect in the provided silhouette image, (a)observe image (b) the silhouette image (c) the image after repairing

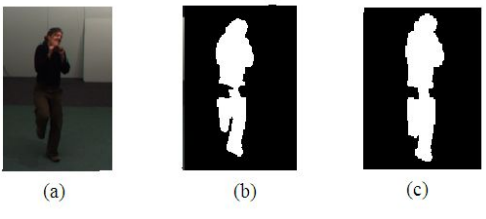


Fig. 10. Other example of defect in the provided silhouette image, (a)observe image (b) the silhouette image (c) the image after repairing

from 5 cameras). In this dataset, actor orientations are arbitrary since no specific instructions were given during the acquisition, as shown in Fig. 8.

Human silhouette images of the observation sequences are provided with the dataset. The quality of the silhouette image is general good but many defects are also present. So morphological close and open operations are applied to the image in order to deal with noise due to background subtraction, as shown in Fig. 9. Not all the defects can be repaired, as shown in Fig. 10. It is a challenging task to recognize the actions from this dataset.

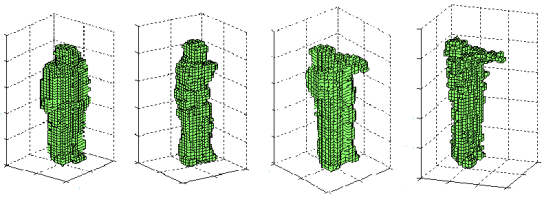


Fig. 11. The exemplars for an action punch

Table 1. The Recognition rate

action	recognition rate(%)
check watch	78.0
cross arms	80.0
scratch head	75.0
sit down	86.7
get up	85.0
turn around	71.7
walk in a circle	70.0
wave a hand	76.7
punch	83.3
kick	81.7
overall	78.8

Since male and female actors' execution styles in action are significantly different, we chose 10 actions (i.e., check watch, cross arms, scratch head, sit down, get up, turn around, walk in a circle, wave a hand, punch, kick), performed by 5 female actors, each 3 times, and viewed by 4 cameras (except top camera) as training and testing objects in our experiment. The action sequences were all manually segmented in advance, so no action segmentation was considered. 4 actors were used for exemplar extraction and model learning each time, another one was used to test the models. Finally the average recognition rate was calculated.

We extracted 4 exemplars for each action, which is sufficient to accurately recognize those actions. The exemplars of the action punch are shown in Fig. 11.

One HMM is used to model each action class, then a sequence of observation is recognized by using the maximum a posteriori estimate. The recognition rates for each action class are listed in Table 1. The confusion matrix is provided in Fig 12 to show the effectiveness of the method.

The results show that, among those 10 actions, "sit down" and "get up" are the easiest actions to recognize because they are more remarkable than the other actions. Some arm related actions such as "check watch", "cross arm", "scratch head" and "wave hand" got relatively low recognition rate in that some silhouette images of the key poses are very similar from single cameras. Our system achieves a satisfied average recognition rate of 78.8% using a single camera, which is lower than the work (80.6%) in [22]. However our system runs at 15 frames/sec, which

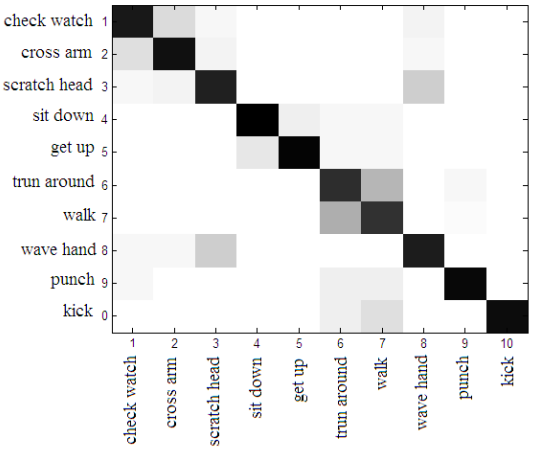


Fig. 12. Confusion matrix for recognition

is faster than 5 frames/sec in [22]. It has potential to implement the proposed work into the real time human motion recognition.

6 Conclusion

An exemplar-based view-invariant action recognition framework has been proposed in this paper. Experimental results have demonstrated that the proposed framework can effectively recognize human actions performed by different people and different actions types in real time without the camera viewpoint constraints. Future work are targeted as follows: 1) Fuse multiple features of actions to achieve higher recognition rate in order to meeting the requirement of the scenarios provided in [25]; 2) Replace the HMMs with Conditional Random Fields to improve the recognition performance and to decrease the computational cost [27,28]; 3) Extend the proposed framework to complex unconstrained scenes.

References

1. Gross, R., Shi, J.: The CMU motion of mody (MoBo) database (CMU-RI-TR-01-18) (June 2001)
2. Moeslund, T., Hilton, A., Krüger, V.: A survey of advances in vision-based human motion capture and analysis. Computer Vision and Image Understanding 104(2-3), 90–126 (2006)
3. Ji, X., Liu, H., Li, Y., Brown, D.: Visual-based view-invariant human motion analysis: A Review. In: Lovrek, I., Howlett, R.J., Jain, L.C. (eds.) KES 2008, Part I. LNCS (LNAI), vol. 5177, pp. 741–748. Springer, Heidelberg (2008)
4. Ogale, A., Karapurkar, A., Aloimonos, Y.: View-invariant modeling and recognition of human actions using grammars. In: Proc. IEEE Conf. Computer Vision, vol. 5, pp. 115–126 (2005)

5. Rogez, G., Guerrero, J., Martınez, J., Orrite, C.: Viewpoint independent human motion analysis in man-made environments. In: Proc. British Machine Vision Conference, pp. 659–668 (2006)
6. Elgammal, A., Lee, C.: Inferring 3D body pose from silhouettes using activity manifold learning. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 681–688 (2004)
7. Lee, C., Elgammal, A.: Simultaneous inference of view and body pose using torus manifolds. In: Proc. Int. Con. Pattern Recognition, vol. 3, pp. 489–494 (2006)
8. Yu, S., Tan, D., Tan, T.: Modelling the effect of view angle variation on appearance-based gait recognition. In: Proc. Asian Conf. Computer Vision, vol. 1, pp. 807–816 (2006)
9. Weinland, D., Ronfard, R., Boyer, E.: Free viewpoint action recognition using motion history volumes. Computer Vision and Image Understanding 104, 249–257 (2006)
10. Natarajan, P., Nevatia, R.: View and scale invariant action recognition using multiview shape-flow models. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 1–8 (2008)
11. Rao, C., Yilmaz, A., Shah, M.: View-Invariant representation and recognition of actions. International Journal of Computer Vision 50(2), 203–226 (2002)
12. Parameswaran, V., Chellappa, R.: View invariants for human action recognition. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, vol. 2, pp. 83–101 (2003)
13. Parameswaran, V., Chellappa, R.: View independent human body pose estimation from a single perspective image. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, vol. 2, pp. 16–22 (2004)
14. Yilmaz, A., Shah, M.: Actions as objects: a novel action representation. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 984–989 (2005)
15. Yilmaz, A., Shah, M.: Actions sketch: a novel action representation. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, vol. 1, pp. 984–989 (2005)
16. Blank, M., Gorelick, L., Shechtman, E., Irani, M., Basri, R.: Actions as space-time shapes. In: Proc. IEEE Conf. Computer Vision, vol. 2, pp. 1395–1402 (2005)
17. Elgammal, A., Shet, V., Yacoob, Y., Davis, L.S.: Learning dynamics for exemplar-based gesture recognition. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, vol. 1, pp. 571–578 (2003)
18. Mori, T., Segawa, Y., Shimosaka, M., Sato, T.: Hierarchical recognition of daily human actions based on Continuous Hidden Markov Models. In: Proc. IEEE Conf. Automatic Face and Gesture Recognition, pp. 779–784 (2004)
19. Shi, Y., Bobick, A., Essa, I.: Learning temporal sequence model from partially labeled data. In: Computer Vision and Pattern Recognition, pp. 1631–1638 (2006)
20. Ahmad, M., Lee, S.: Hmm-based human action recognition using multiview image sequences. In: Proc. Int. Conf. Pattern Recognition, vol. 1 (2006)
21. Patrick, P., Svetha, V., Geoff, W.: Tracking as recognition for articulated full body human motion anlysis. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 1–8 (2007)
22. Lv, F., Nevatia, R.: Single view human action recognition using key pose matching and viterbi path searching. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 1–8 (2007)
23. Weinland, D., Grenoble, F., Boyer, E., Ronfard, R., Inc, A.: Action recognition from arbitrary views using 3D exemplars. In: Proc. IEEE Conf. Computer Vision, pp. 1–7 (2007)

24. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. IEEE Trans. Pattern Analysis and Machine Intelligence 24(4), 509–522 (2002)
25. Cherla, S., Kulkarni, K., Kale, A., Ramasubramanian, V.: Towards fast, view-invariant human action recognition. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 1–8 (2008)
26. Dedeoğlu, Y., Töreyin, B., Güdükbay, U., Çetin, A.: Silhouette-based method for object classification and human action recognition in video. In: Proc. European Conf. Computer Vision, pp. 62–77 (2006)
27. Sminchisescu, C., Kanaujia, A., Metaxas, D.: Conditional models for contextual human motion recognition. Computer Vision and Image Understanding 104(2-3), 210–220 (2006)
28. Wang, L., Suter, D.: Recognizing human activities from silhouettes: motion subspace and factorial discriminative graphical model. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 1–8 (2007)

A Local Interaction Based Multi-robot Hunting Approach with Sensing and Modest Communication

Wenwen Zhang[1], Jing Wang[2], Zhiqiang Cao[1], Yuan Yuan[1], and Chao Zhou[1]

[1] Laboratory of Complex Systems and Intelligence Science,
Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
[2] Yantai Engineering and Technology College, Yantai 264006, China
zhwenwen@126.com, wj814@yahoo.com.cn,
{zhiqiang.cao,yuan.yuan,chao.zhou}@ia.ac.cn

Abstract. A local interaction based hunting approach for multi-robot system in unstructured environments is proposed in this paper. The hunting task is modeled as three modes: initial leader-fixed following&search mode, leader-changeable following&search mode and hunting mode. The conditions for modes switching are given. In order to reduce the dependence on communication, an event-trigger communication scheme based on the evader's observation state is designed. For individual robot, it integrates local information from vision system, sonar sensors and encoders in its local coordinate frame as well as modest communication data to acquire situation-suited task mode, and then makes decisions based on behaviors with appropriate local coordination rules. The experiments with physical mobile robots verify the effectiveness of the proposed approach.

Keywords: Multi-robot system, hunting, sensing, modest communication, local coordination rule.

1 Introduction

To satisfy the requirements of robotics research and applications in industry, military and aerospace etc., multi-robot system with function distribution and resource sharing characteristics has become an important research direction[1][2]. Compared with centralized approaches, those with each robot making decisions autonomously are more attractive. Explicit communication plays an important role for group performance. However, as the amount of robots increases, the communication burden among them might become much heavier. In order to reduce the communication amount, one better solution is to take full advantage of sensory information with elaborated event-trigger information exchange schema. This paper addresses the problem in the context of multi-robot hunting test-bed.

The hunting with multiple robots is a challenging task with inherent dynamic characteristics, and it has various potential applications, such as security, urban search and rescue, et al. Many researchers have studied the problem. In BErkeley AeRobot (BEAR) project, the pursuit-evasion game is studied and it integrates heterogeneous autonomous agents both unmanned aerial vehicles (UAV) and unmanned ground

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 90–99, 2009.

vehicles (UGV). [3][4] describe the implementation in a probabilistic framework by which the game and map-building problems are combined, and a hierarchical hybrid system architecture to coordinate the team is given and computationally feasible pursuit policies are considered. G. Hollinger et al. discretize the environment, and define cost heuristics in planning based on probabilistic formulation[11]. The proposed algorithm is testified by simulations in a complex environment. Yamaguchi proposes a feedback control law for coordinating motions of multiple nonholonomic mobile robots[5][6], and cooperative behaviors to capture/enclose a target are generated by making troop formations, which are controllable by formation vector. Cao considers the hunting problem with an evader under certain simulation assumptions[7][8] and a line searching formation is adopted for the group to find the evader cooperatively[8]. In addition, [14] studies the visibility-based pursuit-evasion problem with limited field of view.

Influenced by social behaviors of insects or animals and sensor network with an unprecedented information quantity, biologically inspired hunting model[9][10] and sensor network assisted pursuit-evasion[12][13] are also studied. [9] presents a wolf pack-inspired hunting model, and the experiments with Sony AIBO robots equipped with local cameras are conducted with image processing in a remote computer. [10] considers the pursuit-evasion game inspired by the phenomenon of stigmergy in social insect colonies such as ants to local food. [12] simulates pursuit-evasion with sensor network system in both centralized and distributed manners. [13] investigates the problem with a sensor network, where a seven-layer control system is designed to implement multi-sensor fusion, multi-target tracking and multi-agent coordination.

This paper considers the hunting problem with multiple sensor-enabled wheeled mobile robots in a distributed manner under unstructured model-free environments, and more emphases self-sensing by on-board CCD cameras, encoders and sonar sensors. There's no global coordinate and the individual robot describes the environment in its local frame. One important point of the proposed hunting approach is that it is helpful for the robot to find the evader by maintaining a following relationship with a teammate that has detected the evader. In addition, in order to reduce the dependence on communication, a modest communication scheme is given. The local coordination rules prevent inter-robot confliction and guide the robots' motion to emerge group-level cooperation. The task is considered to be completed when the relative distance between the robot and the detected evader is less than a predefined capture distance.

The rest of the paper is organized as follows. In Section 2, the proposed hunting approach is described in detail. Individual control architecture, task mode library and basic behaviors as well as coordination rules are presented. Experiment results are demonstrated in Section 3 and Section 4 concludes the paper.

2 Multi-robot Hunting Approach

This paper considers a hunting scenario of multiple robots and a single evader. Each robot is equipped with vision system, encoder, sonar sensors and radio device, and can be distinguished by a unique on-board two-color coded cylinder. Multiple robots, the evader and a monitor console constitute a wireless local area network. The monitor console is only responsible for receiving and transmitting very limited communication

data. The individual robot establishes a local polar coordinate frame whose pole is the robot center with the polar axis direction of its heading.

The robot distinguishes its interested objects both the teammates and the evader with its on-board vision system and estimates their relative positions. Sonar sensors detect the surrounding obstacles. The individual robot integrates all acquired data in its local frame and makes decisions based on the local information as well as limited communication data. For individual robot, the interested objects may run out of the camera's sight or the vision system might fail to recognize the interested objects due to the limited field of view, environmental disturbances and et al. In this case, the estimation of relative positions is necessary by integrating historical visual data with encoder information. During the hunting process, the robot may follow a contributive teammate or hunt the evader, and we call the leader teammate or the evader as target collectively.

For robot R, the position of the evader is denoted as (d_e, θ_e), where d_e, θ_e are the relative distance and bearing in its local frame, respectively. Similarly, (d_{R_i}, θ_{R_i}) is the position of teammate R_i. Sonar sensors $S_k (k = 0,1,...,k_s - 1)$ are distributed evenly around the robot and the perception information is described by (ρ_s^k, ϕ_s^k), where k_s is the number of sonar sensors, ρ_s^k is the distance of obstacle detected by sensor S_k in the local frame, and $\rho_s^k = 0$ if it doesn't detect obstacles. $\phi_s^k = -\pi + \frac{2k\pi}{k_s}$ is the angle corresponding to sensor S_k.

2.1 Control Architecture and Communication Scheme for Individual Robot

The control architecture for individual robot is shown in Fig. 1.

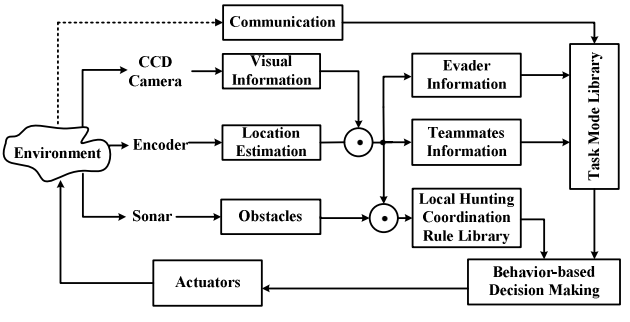


Fig. 1. Individual robot's control architecture

Based on the current local sensing information and communication data, the individual robot firstly chooses proper mode from the predefined task mode library, and then makes behavior-based motion decisions according to the appropriate coordination rule from local hunting coordination rule library. The motion control commands are then sent to the actuators.

To reduce the dependence on communication, an event-trigger modest communication scheme based on the evader's observation state is designed. The observation state indicates whether the evader is visible or not and we call the state changes as events. There are two events will be considered:

Event 1: Once the robot detects the evader by its vision system, the event of evader being detected occurs.

Event 2: When the robot fails to detect the evader and the positions estimation based on encoders is also invalid, the evader lost event occurs.

The communication arises only when the above events occur, thus the designed communication scheme will reduce communication amount. Furthermore, the communication data are only about the evader's observation state and no any position information are transmitted. Each robot stores a teammate-evader state table T_R, which indicates the evader's observation state of each teammate. Once the robot receives communication data, it will update the table immediately.

2.2 Task Mode Library

The proposed hunting approach is summarized as follows. At the beginning, multiple robots cooperative explore the environment to find the evader by predefined dynamic following&search. The only specified navigator guides the direction of the search, while other robots follow their respective predefined leaders. This may improve the possibility of detecting the evader and be beneficial to group hunting. The predefined leader-follower relationship will be kept until the following two cases occur:

1) Once the robot detects the evader, it begins to pursue the evader and informs all teammates about its evader's observation state;
2) For the robot that hasn't found the evader since the beginning, if there are detected teammates that have detected the evader according to the table T_R, it will select the teammate with the smallest ID as its leader.

When the robot loses tracking of the previous selected leader or the leader has lost the evader according to the table T_R, it will consider changing its selected leader. In addition, the evader may be lost during the pursuit process, and the robot will consider reselecting an appropriate leader by inquiring the table T_R.

Based on above description, the hunting task is modeled as three modes: initial leader-fixed following&search mode, leader-changeable following&search mode and hunting mode, and they constitute the task mode library.

The modes switching is shown in Fig. 2. The meanings of symbols C_i (i=1,…,4) are shown as follows:

C_1: The robot detects the evader by its vision system;
C_2: The robot has lost the evader;
C_3: There are teammates that have detected the evader according to table T_R;
C_4: The robot detects at least one teammate mentioned in C_3.

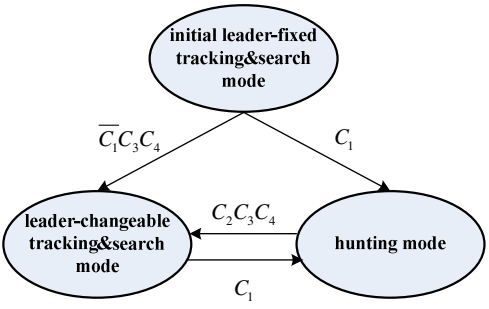


Fig. 2. The hunting task modes

2.3 Behaviors for Individual Robot

Individual robot has different targets in different task modes. A behavior library is presented and it includes: random search, round-obstacle, following and hunting behaviors. Different task modes have different behaviors combination. If the robot is in hunting mode, the behaviors set is $\{random\ search, round\text{-}obstacle, hunting\}$, or else, the set is given by $\{random\ search, round\text{-}obstacle, following\}$.

1) *Random search behavior*. This behavior is activated when there's no target to be followed or hunted. Whether the robot is in initial leader-fixed tracking&search mode or leader-changeable tracking&search mode, if it loses tracking of the pre-assigned or the selected leader, and there's no teammate can be selected to sever as its leader, it has to execute random search behavior. When the robot performs hunting mode, once the evader is lost, it needs to search the environment to find the evader or a helpful leader.

2) *Round-obstacle behavior*. No matter the target is teammate or evader, the obstacles are always considered. The collision avoidance is the most important if the obstacles block the robot's path. The robot has to firstly determine the prior rounding direction, namely which side has the higher priority to move around based on the distribution of obstacles with respect to the target in its local frame. If the target is the evader, the prior rounding direction is selected with consideration of certain coordination rule.

3) *Following behavior*. The robot follows its leader with desired distance and observation angle. In order to cope with imprecise sensory information, once the relative distance between the robot and its leader is smaller than the desired distance, the robot will maintain detection of the target with the constraint of a limited observation angle by rotation motion and resumes following to its leader only when the relative distance is larger than a bigger value compared with the desired distance.

4) *Hunting behavior*. In this behavior, the robot pursues the evader. In order to avoid the interference amongst the robots, coordination rules will be used.

2.4 Coordination Rule Library

In our multi-robot hunting system, multiple mobile robots try to capture the common evader in a limited workspace, it's necessary to coordinate the robots' motions to avoid conflicts and improve the group performance. It's a better solution to keep the

separation angles among robots with respect to the evader being big enough. On the basis of above, we give a coordination rule library based on geometric restricts.

When the robot performs hunting mode and tries to capture the evader, it is necessary to decide whether or not to coordinate with teammates based on the analysis of current situations. Then the robot chooses an appropriate rule from coordination rule library to better complete the task based on local interaction with teammates.

Take the robot R as an example, we denote with R_l and R_r the teammates with the smallest angle between the teammate and robot R on each side with respect to the evader (see Fig. 3). The smallest angles are denoted with θ_l^r and θ_r^r respectively. Let O_{lr} be the midpoint of segment connecting R_l with R_r, O_{rlr} is the midpoint of line between R_l and R, and $\theta_{th}^{rp_{co}}$ is the given coordination angle.

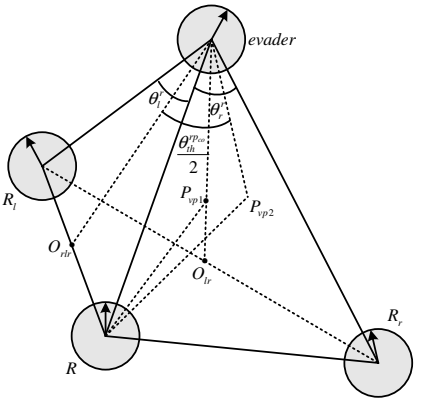


Fig. 3. Hunting Coordination among the robots

1) *Coordination Rule 1*($\theta_l^r < \theta_{th}^{rp_{co}} \bigcap \theta_r^r < \theta_{th}^{rp_{co}}$): The robot needs to coordinate with teammates R_l and R_r from both sides as shown in Fig. 3. It will move along the dash line between point O_{lr} and the evader to perform coordination, and a point P_{vp1} is selected to guide its motion. The point P_{vp1} is on the dash line connecting the evader and O_{lr} with a distance $\frac{d_e}{k_{co}}$ to the evader, where $k_{co} = 2$ and d_e is the evader's relative distance.

2) *Coordination Rule 2*($\theta_l^r < \theta_{th}^{rp_{co}} \bigcup \theta_r^r < \theta_{th}^{rp_{co}}$): The robot will coordinate with the teammate at only one side. Take left teammate R_l as an example. The better path is the ray generated by rotating the line between the evader and point O_{rlr} away from R_l with an angle of $\frac{\theta_{th}^{rp_{co}}}{2}$. The point P_{vp2} (see Fig. 3) lying on the ray with a distance $\frac{d_e}{k_{co}}$ to the evader will guide the robot's motion.

Besides the above coordination rules for hunting behavior, another rule for round-obstacle behavior is designed to prevent the robots from moving around obstacles along only one side.

3) *Coordination Rule 3*: For round-obstacle behavior of hunting mode, one important factor to determine the prior rounding direction is the distribution of teammates with respect to the evader. The side with fewer teammates has higher priority.

3 Experiments and Results

We adopt a team of wheeled mobile robots developed by CASIA to testify the proposed approach. Each robot is equipped with 3 CCD cameras with field of view 60°, and a 16-sonar ring is evenly arranged around the robot. We represent the experimental results with task modes, the trajectories of robots/evader based on encoders, and snapshots of video sequences. We denote with m_mode task modes and m_mode is 1, 2, 3 when the robot is in initial leader-fixed following&search mode, leader-changeable following&search mode and hunting mode, respectively.

Experiment 1 adopts two robots R_1 and R_2 to pursue a static evader E. Their initial positions and obstacles are shown in Fig. 4. R_1 is appointed as the navigator as well as the leader of R_2. Several selected images and task mode variations are shown in Fig. 4 and Fig. 5, respectively. At the beginning, R_1 detects the evader. It sends communication information that the evader observation state is detected to R_2 and hunts the evader, while R_2 only sees R_1 and it has to follow this predefined leader. Once R_2 receives the communication information from R_1, table T_R is updated and m_mode is switched from 1 to 2. During the following process, R_2 finds the evader (see Fig. 4(b)) and starts to pursue the evader. After they move around the obstacle from two sides respectively based on the local hunting coordination rules, the task is completed smoothly.

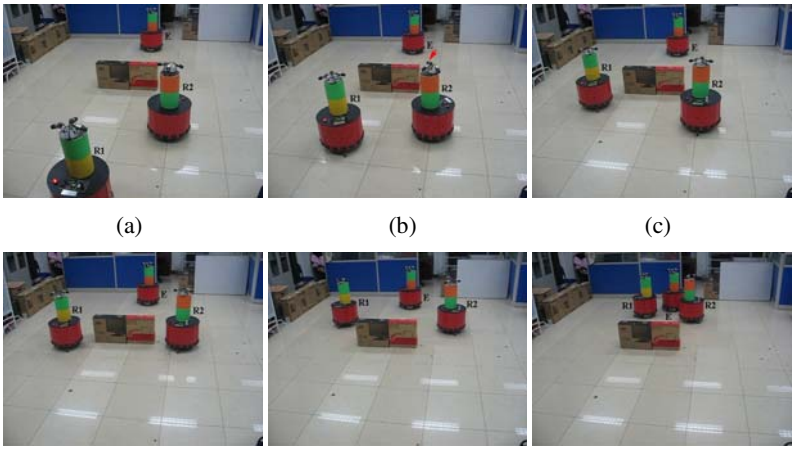


Fig. 4. Selected images of experiment 1

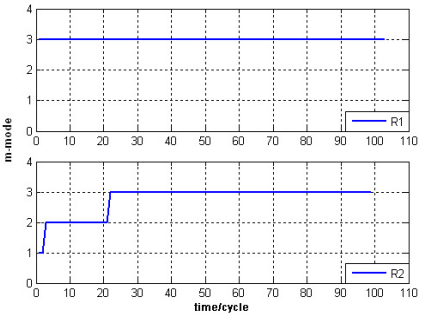
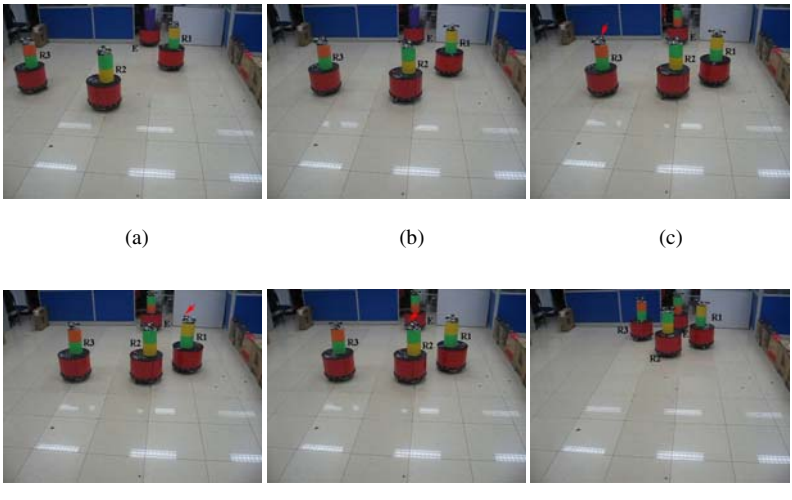


Fig. 5. Task modes of robots in experiment 1

In experiment 2, three robots R_1, R_2, R_3 are required to hunt a static evader E co-operatively. Fig. 6 gives several selected images and task mode variations of each robot are shown in Fig. 7. R_1 is the navigator as well as the leader of R_2, and R_2 is the leader of R_3. Initially, the evader is concealed and the robotic system searches the environment. Later, the evader is unshielded as shown in Fig. 6(c). R_3 will firstly detect the evader, and it sends the information to teammates R_1 and R_2. R_1 switches its task mode from 1 to 2 after receiving the communication message as it finds R_3. During the following process with R_3, R_1 sees the evader by itself (see Fig. 6(d)). After R_2 have detected the evader (see Fig. 6(e)), the task is fulfilled by all robots.



(a) (b) (c)

(d) (e) (f)

Fig. 6. Selected images of experiment 2

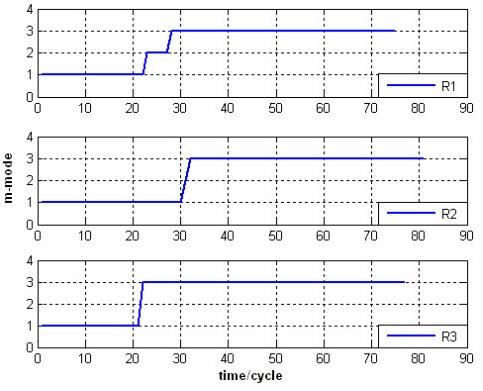


Fig. 7. Task modes of robots in experiment 2

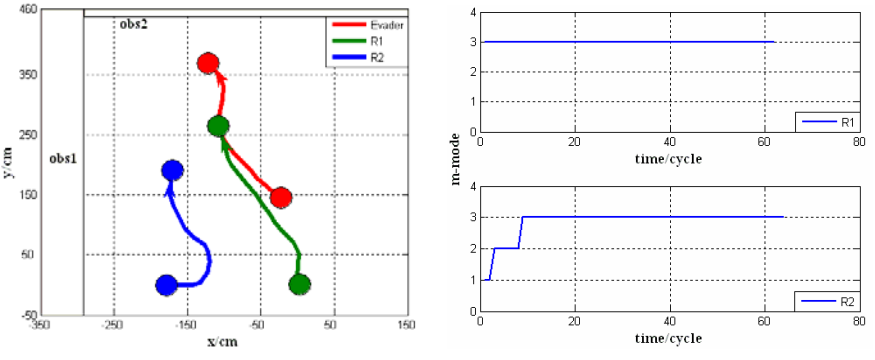


Fig. 8. (a) The trajectories of robots and evader. (b) Task modes of robots in experiment 3.

Two robots R_1, R_2 are adopted to hunt an intelligent evader in experiment 3. The trajectories of all robots and the evader are shown in Fig. 8(a), and Fig. 8(b) gives the task mode variations of all robots involved. It suggests that with the guidance of R_1, R_2 changes from following R_1 to cooperative hunting with R_1 until the completion of the task.

4 Conclusions

Considering the problem of multiple sensors enabled mobile robots hunting a dynamic evader, a distributed hunting approach is proposed in this paper. The designed task modes, behaviors and local coordination rules enable robots to complete the task collaboratively, and the event-trigger modest communication scheme may reduce the network burden. The proposed approach is practicable, and its effectiveness is supported by several typical experiments. Due to the fact that the practical physical multi-robot system is complex and unpredicted, some other issues such as leader failure, package loss, et al. are still required to be further researched.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China under Grants 60805038, 60725309, and in part by the National High Technology Research and Development Program of China (863 Program) under Grant 2006AA04Z258.

References

1. Cao, Y.U., Fukunaga, A.S., Kahng, A.B.: Cooperative mobile robotics: antecedents and directions. Autonomous Robots 4, 7–27 (1997)
2. Tan, M., Wang, S., Cao, Z.: Multi-robot Systems. Tsinghua University Press, Beijing (2005) (in Chinese)
3. Vidal, R., Rashid, S., Sharp, C., Shakernia, O., Kim, J., Sastry, S.: Pursuit-evasion games with unmanned ground and aerial vehicles. In: IEEE International Conference on Robotics and Automation, pp. 2948–2955 (2001)
4. Vidal, R., Shakernia, O., Kim, J., Shim, H., Sastry, S.: Probabilistic Pursuit-Evasion Games: Theory, Implementation, and Experimental Evaluation. IEEE Transactions on Robotics and Automation 18(5), 662–669 (2002)
5. Yamaguchi, H.: A cooperative hunting behavior by mobile-robot troops. International Journal of Robotics Research 18(8), 931–940 (1999)
6. Yamaguchi, H.: A distributed motion coordination strategy for multiple nonholonomic mobile robots in cooperative hunting operations. Robotics and Autonomous Systems 43(4), 257–282 (2003)
7. Cao, Z., Gu, N., Tan, M., Nahavandi, S., Mao, X., Guan, Z.: Multi-robot Hunting in Dynamic Environments. International Journal of Intelligent Automation and Soft Computing 14(1), 61–72 (2008)
8. Cao, Z., Zhang, B., Wang, S., Tan, M.: Cooperative Hunting of Multiple Mobile Robots in an Unknown Environment. Acta Automatica Sinica 29(4), 536–543 (2003)
9. Weitzenfeld, A., Vallesa, A., Flores, H.: A Biologically-Inspired Wolf Pack Multiple Robot Hunting Model. In: Proc. of Latin American Robotics Symposium
10. Li, F., Prithviraj, D.: A Stigmergy-Based Model for Solving Cooperative Pursuit-Evasion Games in Unknown Environments. In: IEEE International Conference on Self-Adaptive and Self-Organizing Systems, pp. 467–468 (2008)
11. Hollinger, G., Kehagias, A., Singh, S.: Probabilistic Strategies for Pursuit in Cluttered Environments with Multiple Robots. In: IEEE International Conference on Robotics and Automation, pp. 3870–3876 (2007)
12. Zhuang, P., Shang, Y., Shi, H.: A New Method of Using Sensor Network for Solving Pursuit-Evasion Problem. Journal of networks 2(1), 9–16 (2007)
13. Oh, S., Schenato, L., Chen, P., Sastry, S.: Tracking and Coordination of Multiple Agents Using Sensor Networks: System Design, Algorithms and Experiments. Proceedings of the IEEE 95(1), 234–254 (2007)
14. Gerkey, B.P., Thrun, S., Gordon, G.: Visibility-based pursuit-evasion with limited field of view. International Journal of Robotics Research 25(4), 299–316 (2006)

Robot Formations for Area Coverage

Jürgen Leitner[1,2,3]

[1] Helsinki University of Technology, Finland
[2] Luleå University of Technology, Sweden
[3] The University of Tokyo, Japan
juxi.leitner@gmail.com

Abstract. Two algorithms for area coverage (for use in space applications) were evaluated using a simulator and then tested on a multi-robot society consisting of LEGO Mindstorms robots. The two algorithms are (i) a vector force based implementation and (ii) a machine learning approach. The second is based on an organizational-learning oriented classifier system (OCS) introduced by Takadama in 1998.

1 Introduction

This paper aims to describe the usage of a machine learning (ML) and evolutionary computing approach to use robot formations for area coverage, e.g. in wireless coverage and astronaut support. ML was chosen because of its capabilities with dynamic and unknown environments, the results will be compared with results based on a simple vector approach.

An interesting scenario for multi-robot cooperation in space is the exploration of the Moon and Mars, where rovers are currently on the forefront of space exploration. For permanent human settlement will depend heavily on robotic reconnaissance, construction and operational support. Tasks for the robots will include mapping landing sites, constructing habitats and power plants, communicating with and acting as a communication relay to Earth, and so forth. Scenarios for this assume many different robots, with different, but overlapping, capabilities working together. One scenario for the use of multiple robots teaming up and working together can be taken from the recent NASA plans for building *human outposts* on the Moon and Mars. These plans outline also a need for robotic support for the astronauts and explorers. In this scenario robots will, for example, search cooperatively for a location suitable in size and other properties to harbor a permanent human settlement. These tasks will include soil preparation and movement as well as, for example, carrying solar panels in tight-cooperation between two robots. The heterogeneity of the rovers is exploited throughout the whole mission to allow for better performance. Meanwhile, other rovers will begin with the exploration and surveying of the region around the construction site. Rovers with specialized sensing instruments are sent to investigate and cover as much of the area as possible, possibly transported in a larger mother-ship type robot at first. Formations of rovers will generate a wireless communication and emergency network for the robots as well as future human explorers.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 100–111, 2009.

Coverage Problem. The *coverage problem* (sometimes also referred to as *covering problem*) is widely defined as to *"cover a search space consistently and uniformly"* (1), or more informally (in robotic exploration) to "see" every point of a given area. The coverage problem for a team of robots was categorized into three types (2): blanket coverage, barrier coverage, and sweeping coverage. Blanket coverage, which is the closest problem to the area coverage presented here, has the objective of maximizing the total covered area by a static arrangement.

Machine Learning. Machine learning (ML) is using artificial intelligence algorithms to allow a system to improve over time using input data. ML is (i) useful when human expertise is not available, (ii) needed when solutions change over time or need to be adapted and (iii) helping to understand how humans learn.

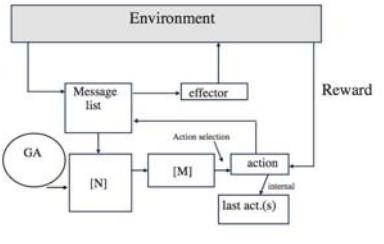
Evolutionary computing describes search algorithms based on natural selection and genetics. The idea is to search a problem space by evolving from a random starting solutions to better/fitter solutions, generated by selection, mutation and recombination.

This paper describes two implemented algorithms: a lightweight and simple force vector approach and a more complex algorithm based ML principles. They were implemented in C++ and tested in a separately developed simulator called *SMRTCTRL* (3).

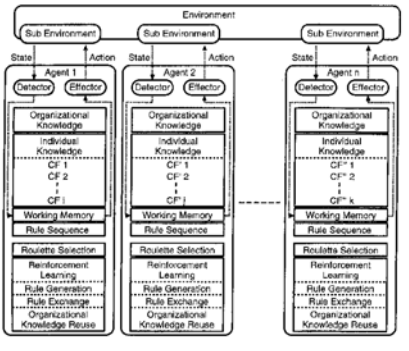
2 Related Work

The Art Gallery Problem is the real-world problem of guarding a gallery with a minimum number of guards, viewing the whole gallery. The mathematical problem was proved to be NP-hard (4).

The Learning Classifier System (LCS) developed by Holland in the 1970s is based on techniques found in evolutionary computing and reinforcement learning. It tries to learn which actions to perform based on the input received. LCS uses a population of binary rules on which a genetic algorithm (GA) is used to



(a) Holland's LCS (5)



(b) Takadama's OCS (6)

select and alter these (7). The GA evaluates rules by its fitness, which is based on a payoff received similar to reinforcement learning approaches.

An LCS-based algorithm receives *inputs* from the environment at every iteration step. Based on this the algorithm selects the best/fittest action. The possible *actions* depend on the (physical) context of the system. The *reinforcement* should reflect physical properties, e.g. a mobile robot might get reinforcement according to the distance moved towards the goal. An LCS is a rule-based system, with the *rules* usually in the form of "`IF state THEN action`". The GA used is operating on the whole rule-set every few time steps using roulette wheel selection and replacing rules with newly created ones.

Organizational-learning Oriented Classifier System (OCS) takes an *organizational learning* approach and adapts it to machine learning in multi-agent systems. The method is described as *"learning that includes four kinds of reinterpreted loop learning"* (6) and tries to generate a *hybrid solution*, aiming to overcome restrictions of the Michigan and Pittsburgh approaches (8). The system, sketched in Fig. 1b, uses reinforcements and evolutionary learning.

Vector-Based Formations. The use of vectors is often found in reactive behaviors for robots, with this representation generally known as a *vector-force-field*. These approaches, generally, do not need a model of the environment, it is enough to sense obstacles. (9) presents an approach to use vector-force-fields to control the movement of robots in a given formation, while avoiding obstacles and progressing towards a target. It concentrates on the subtask of controlling and maintaining the formation by using forces calculated from the other robots, leading to the robot being pulled into the formation or pushed away. The paper uses position and heading of each robot to calculate 2 independent forces.

3 Placement Using Force Vectors

A vector-based approach, with vectors representing attractive and repulsive forces, is used to place the robots close to the target area. These forces are calculated and a combination of these is used to generate the movement for each robot separately. First the attractive forces to the mid-point of the target area and to the other robots are calculated, then repulsive forces are added, where the values are given by the coverage in the direction of the other robot. From this the coverage optimization is an emergent property of this algorithm.

4 Placement Using Machine Learning

ML techniques, based on the work of Takadama on OCS (6), are used to calculate the actions to be taken to optimally place the robots for area coverage. The system consists of autonomous *agents*, with local LCS-based implementations. Each agent is able to recognize the *environment* and its local state and is able to change the environment due to a chosen action. The algorithm is implemented in

a class encapsulating the learning algorithm and facilitates an object-orientated approach for representation of the agents and the environment. Each agent has its local memory, which is used to create, store and update rules, also called *classifiers* (CFs). These rules are used to select the most suitable action for the current state. The agents apply reinforcement learning (RL) and GA techniques for rule management. For this they detect the environment state and decide on an action based on the state. Each agent updates its own rule-set.

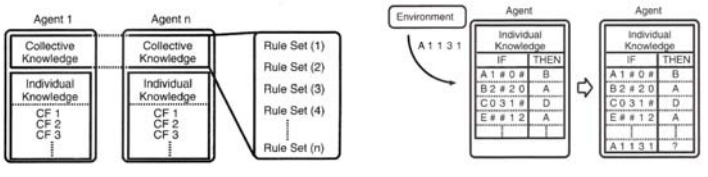
4.1 Mechanisms

The OCS algorithm uses 4 separate forms of *learning mechanisms* presented and described in this section. Pseudo-code is added to sketch the implemented algorithm and its main components as described in (6). The main OCS loop is shown in Listing 1.

Listing 1. The OCS algorithm

```
procedure OCS
    iteration := 0
    Collective_Knowledge_Reuse
    while iteration < MAX_ITERATION
        reset the problem to the starting position
        iteration := iteration + 1, step := 0
        while not solution_converges
            Detector()
            Rule_Generation, Rule_Exchange
            Effector()
            step := step + 1
        end
        Reinforcement_Learning, Collective_Knowledge_Reuse
    end
```

Collective Knowledge Reuse. The reuse of good solutions (e.g. role specialization) is helping to reduce the learning time as well as solving problems that need *organizational knowledge*. It uses a central set of agent behaviors, as is depicted in Fig. 1a. This mechanism (Listing 2) is useful when no solutions can be found without organizational knowledge, but due to time constraints was not implemented in the test runs.



(a) Collective Knowledge Reuse (b) Rule Generation

Fig. 1. Illustration of 2 Learning Mechanisms in OCS (10)

Listing 2. `Collective_Knowledge_Reuse` **procedure**

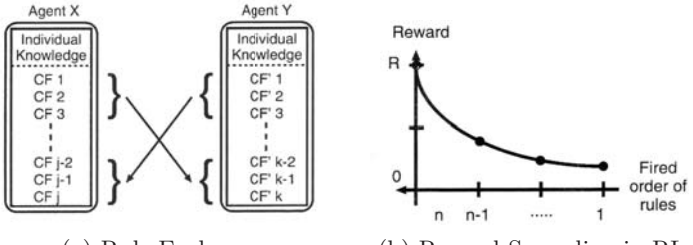
```
procedure Collective_Knowledge_Reuse
    if iteration = 0
        use stored collective knowledge
    else if solution is the best
        if collective knowledge is set already
            delete stored collective knowledge
        store current rule-sets as collective knowledge
```

Rule Generation (Fig. 1b) is used to create *new* CFs when no CFs in the rule-set of an agent, i.e. the *Individual Knowledge Memory*, match the current sub-environment state detected. The generation of new CFs allows to explore other solutions to the problem, that have not yet been tested. These new CFs are created, with a random action value and a *condition* based on the state detected. The strength is initialized to `START_STRENGTH`. The weakest CF is dropped if the defined maximum of CFs is reached (Listing 3).

Listing 3. `Rule_Generation` **procedure**

```
procedure Rule_Generation
    for all agents
        if no classifier matched
            if number of classifiers = MAX_CF
                delete CF with lowest strength
            create a new CF & set strength to START_STRENGTH
    end
```

Rule Exchange is used to generate new behaviors in the agents and therefore explore new areas and CF combinations. The CFs of two, randomly chosen, i.e. non-elite selection, agents are exchanged, at every `CROSSOVER_STEP`. The rules are sorted by strength and the worst CFs of one agent are replaced by the best CFs of the counterpart and vice-versa (Listing 4). The strength values of the rules exchanged are then reset to the starting value. The rule exchange is depicted in Fig. 2a.



(a) Rule Exchange (b) Reward Spreading in RL

Fig. 2. Illustration of 2 Learning Mechanisms in OCS (10)

Reinforcement Learning (RL) (Fig. 2b) updates the *strength* of each classifier in the *Rule Sequence Memory* in the case of a converging result. The received reward is spread, based on a simple exponential function, over all the fired CFs, i.e. the CFs in the *Rule Sequence Memory*.

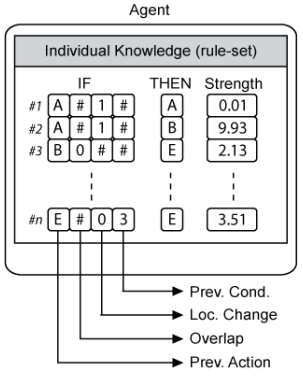
Listing 4. `Rule_Exchange` procedure

```
procedure Rule_Exchange
    if mod(step, CROSSOVER_STEP) = 0
        for a random pair of agents
            order rules by strength
            for the weakest CROSSOVER_NUM CFs
                if strength < BORDER_ST
                    delete CF
            copy CROSSOVER_NUM strongest from other agent
            reset strength to START_STRENGTH
```

4.2 Classifier

The classifiers are the same as the rules in LCS, simple *if-then* constructs composed of three parts: (i) the *condition*, (ii) the *action* and (iii) the *strength* or fitness value. The *condition*, or *if* clause, of the CF contains various information about the environmental state. The *action*, or *then* part, defines the action to be executed by the agent if the condition part matches the current environment. The third part, the *strength* value, defines how good the CF is, which effects selection and therefore the agent's behavior.

Condition. The condition part of a classifier is representing the knowledge of the current situation. It is a list of input values, represented by the ternary alphabet $\{0, 1, \#\}$, though extended in some cases. Fig. 3 shows the structure of the CFs and the condition in the OCS algorithm. The condition contains: (i) The `PREV_ACTION` executed by this agent, (ii) a flag representing an existing



Listing (5): Code for the CFs

```
enum eCFCondElements {
  PREV_ACTION = 0, OVERLAP,
  LOC_CHANGE, PREV_CONDITION
};

//...
char cond[CDPARNR]={'#','#','#','#'};
// init others
cond[PREV_CONDITION] =
  prevOverlap - '0' + cond[OVERLAP];
if(cond[PREV_CONDITION] != '0' &&
   cond[LOC_CHANGE] == '1')
  cond[PREV_CONDITION]++;
prevOverlap = cond[OVERLAP];
```

Fig. 3. The structure of the rules, aka classifiers (CFs). *based on (6).*

OVERLAP with another agent, (iii) a flag representing a LOC_CHANGE and (iv) a number calculated and indicating the PREV_CONDITION.

Actions. The actions of the agents are 7 simple motions, derived from a simple mobile robot. The actions are: MOVE FORWARD, MOVE BACKWARD, MOVE RIGHT, MOVE LEFT, ROTATE RIGHT, ROTATE LEF and STAY.

4.3 Memories

The memories at every agent store various information about the *classifiers*. In the memory CFs can be created, stored and updated. The four memories of each agent are:

(i) *Organizational Knowledge Memory*, also referred to as *Collective Knowledge Memory*, stores a global set of rule-sets. This knowledge is shared by all agents and allows for role specialization and classifier reuse.

(ii) *Individual Knowledge Memory* stores the current rule-set of every agent. This is where the knowledge of the individual agent is developed over the various iterations. FIRST_CF number of rules are generated and initialized at the start, then during the learning operation new rules are added or exchanged.

(iii) *Working Memory* is used for operations in the agents, e.g. selection of matching CFs, a temporary list is available to be filled with CFs.

(iv) *Rule Sequence Memory* stores the rules selected at every iteration.

4.4 Environment Interactions

The *environment interactions* are modeled and are implemented in two functions: (i) for retrieving the current state, `Detector()` and (ii) reacting by executing an action, `Effector()`.

Detector() This function detects the current state of the (sub-)environment at every agent, with the main objective to create a new CF condition representing the state. This is modeled into the 4-field vector, described above, where the PREV_CONDITION is calculated by adding the last iterations overlap field with the current. If one of them is set, also the value of the current LOC_CHANGE flag is added.

5 Results

5.1 Simulation Results

A *standard initial configuration* (SIC) was used to be able to compare the results achieved in the test runs. This configuration reflects a marsupial society just deployed using: a *mother-ship* (Motherbot), which is, for these two control algorithms, not moveable and has a circular coverage area (5x5, 25 cells), as well as four controllable *child-bots* (Marsubots), each possessing an elliptic coverage area (5x7, 27 cells). The robots are placed with one cell in between each other. The *target area* is defined as a circular area, 69 (9x9) cells in size, and placed 22 and 11 cells away in X and Y axis respectively.

Vector-Based Simulation Results. The vector-based control approach was mainly aimed to be tested with circular target areas and therefore does not implement a directional force vector, hence the robot actions are reduce to the four motions (*forward, backward, left, right*). For the simulation runs no obstacles were inserted and no terrain interaction was performed on the sensing areas. The simulation also assumed that all actions can be completed within one time step and that the robots can be controlled simultaneous, which turned out to be difficult (see SMURFS project below).

The vector approach generated higher repulsive forces, when switched to elliptical, due to the lack of a directional understanding, and hence not a full coverage was found. This tests, with the elliptic sensor areas, showed that the algorithm converges to a stable formation after 83 actions, generating a combined coverage of only 12%, or 8 of the 69 target area cells. This could be overcome by using a better scaling factor and a directional component.

An interesting effect observed was that robots with smaller sensing areas are moving closer to the target, and by doing so, drive the robots with larger areas away from it. This is explained by the target force calculation, which leads to larger forces in the direction of the target for robots with smaller coverage (since there is a smaller repulsive force).

Machine Learning. All ML algorithm test runs use the SIC and classifier-action pairs selected through learning to control the robots' motion. After a convergence in the results, the robots are reset to the SIC and another iteration of ML is performed. Here all 7 robot motions were used.

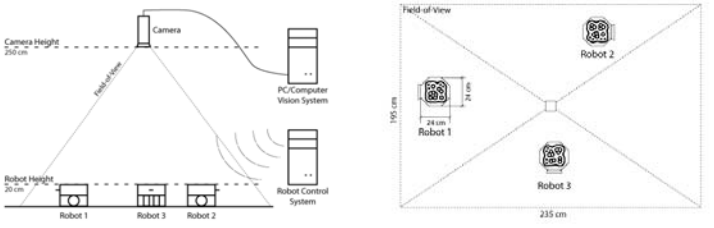
Throughout the iterations the agents and therefore the global solutions are increasingly optimizing the coverage problem, with the the best coverage a bit over 26%, which is better than the vector-based approach. It can be seen that the results are varying quite a bit but an overall trend to increased coverage is visible. The results show that the randomized learning approach does overall yield a good coverage with a good distance and that the best solution improves over time.

The test run used the following settings for its OCS learning to obtain the result: Each agent has between 15 (at start, `FIRST_CF`) and 30 (`MAX_CF`) classifiers and exchanges 7 at every crossover step (`CROSSOVER_CF_NUM`). These classifiers are initialized with a strength of 1.0 (`START_STRENGTH`) and the border strength (for crossover) is set to 1.15 (`BORDER_ST`). The test run had a maximum number for 500 iterations (`MAX_ITERATIONS`), each with a maximum of 300 steps (`MAX_STEPS`), but the average number of steps per iteration was only 68.3. A rule exchange between two random agents was done every 10 steps (`CROSSOVER_STEP`). The average reward per iteration was only 1.86 (with the maximum at 6.3).

As with the vector-based approach for the simulation no obstacles were inserted and the simulation assumed that all actions can be completed within one time step. A few hundred test runs were done, but due to the inherent randomness (in the initial rule generation) the results vary quite a bit, for example, the final coverage ranges between 0% and 47%. Obtaining the results was harder than anticipated and they are also not as good as expected.

5.2 Project *SMURFS* at IJCAI

To allow testing of the algorithm, robots from the Society of Multiple Robots (SMURFS) project were used, which were designed to be cheap and able to reconfigure. The robots were using LEGO Mindstorms NXTs for control and 4 motors for actuation. More information about the robots can be found at (11). A webcam and visual tracking system using fiducials (12), nicknamed *Gargamel*, was used to provide localization for the robots, which were then controlled via the *SMRTCTRL* simulator using Bluetooth. A simple sketch showing the experiment arrangement can be seen in Figure 4.



(a) Camera Setup

(b) Field of View of the Camera, birds-eye view

Fig. 4. The *Gargamel* Tracking System for the *SMURFS* robots

Experimental Findings. The simple control, based on the vector-based placement algorithm, was controlling the robots using the freely available *nxtlibc* library. The robots were controlled sequentially, due to limitations of the library, which did not provide multi-threading support. The discretization in the simulator was changed to closer represent the robots and the field-of-view of the camera, but even with that the robots could not be controlled very precisely. Though some effort was put into the action-programs, for example, the move forward program, it could not be ensured that the robot would move exactly one field when executed. This discrepancy was seen especially at the borders of the vision system. Videos of the system controlling the robots can be found online[1]. Figure 5 compares the motion of the robots as seen by the simulator *(boxes)* and in real-life *(circles)*. The data presented is the average of 3 test runs performed with the final revision. The position of the target is marked with a **T**. The graph shows the robots performing 12 actions (per test run), of which 3 are rotations. To visualize the orientation and rotation small lines are added to one side of the square or circle. Discretization errors at the center can be seen with all robots. The robots' `MOVE_FORWARD` motion were tested and yielded $20.1 - 20.5cm$, with the first movement usually a bit shorter ($19.2 - 19.6cm$), probably due to the initial orientation of the castor wheels.

[1] At our webpage (`http://smrt.name/ijcai`) and on YouTube (e.g. `http://www.youtube.com/watch?v=RnP70SaH4iw`). *(July 15, 2009)*

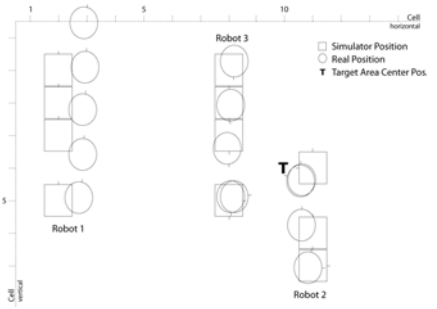


Fig. 5. Robot Movements: Real-World vs. Simulator

Table 1. A comparison of the results of the two algorithms

	Vector Approach		**OCS**		
	(circular)	(elliptic)	(average)	(best)	(worst)
Coverage	100 %	12 %	26.58 %	46 %	0 %
Steps	29	83	64.03	17	289
Distance [px]	230	880	1235.3	582	1843

In the first test runs a discrepancy between sent action and the outcome of the motion was seen. For example, a `MOVE_FORWARD` action would lead, due to drift, also to a movement to a side (this can also be seen in the motion of robot 1 in Figure 5); this was then reduced as much as possible.

6 Conclusions

The *vector-based approach* leads to a converging coverage in most cases, though in some situations singularities or oscillations occur. It is not able to handle terrain interaction and performs best with circular sensor areas. The *organizational-learning oriented classifier system (OCS)* approach was more complicated to implement but allows for terrain interactions and different elevations as well as obstacles. It though is not always converging and is, in the simple case, outperformed by the lightweight vector approach (see Table 1). The system is still in its very early stages and needs to be evaluated further. Especially the connection between the parameters (for example, number of classifiers, strength values and of course, `MAX_ITERATIONS`, `MAX_STEPS`, `CROSSOVER_STEP`, just to name a few), but also the choice of the actions, the classifiers (and its structure) and the definition of the reward and fitness function need to be researched further. For real-world experiments the algorithm has to perform faster and as decentralized as possible, which is currently not the case.

To make the OCS approach work and yielding good results more time than planned was needed. In the end it worked better than the vector-based approach

in an environment having different elevations and obstacles present. This is because of the classifiers, which allow the OCS approach to generate two solutions for a given input condition, for example, one that moves the robot to higher ground and a second one that keeps the robot on the same height. The fitness for the two solutions will not be the same, which means one of the CFs is chosen with higher probability, therefore the environment interaction leads to different solutions, which is not the case for the vector-based implementation. The conclusion would though still be that the lightweight, vector-based control algorithm might be the more feasible approach to use, although also this approach needs some further research to work in the environment and scenario described in this thesis.

At the IJCAI robotics workshop a simple vector-based control of the robots was shown. Due to time constraints only a very basic implementation was shown, but we are confident that with more time a more precise control using a multi-threaded Bluetooth library could be implemented.

7 Future Work

Apart from the placement problem discussed above, another interesting issue is the change of formation due to changing conditions. Some thoughts about possible extensions for the approaches to tackle this problem. Because of the limited time and increasing complexity this problem was though not solved in this project. The idea was to extend the ML approach to be able to handle these situations. This could be done by adding an algorithm, that works with the results of the OCS learner presented here and uses the optimal configurations found to generate a transition matrix. This matrix would contain which formation to switch to, according to some optimizing criteria, if one robot fails.

References

[1] Menezes, R., Martins, F., Vieira, F.E., Silva, R., Braga, M.: A model for terrain coverage inspired by ant's alarm pheromones. In: ACM Symp. on Applied computing, pp. 728–732 (2007)
[2] Gage, D.W.: Command control for many-robot systems. In: The 19th Annual AUVS Technical Symposium, pp. 22–24 (1992)
[3] Leitner, J.: Multi-robot formations for area coverage in space applications. Master's thesis, Helsinki University of Technology, Espoo, Finland (2009)
[4] Aggarwal, A.: The art gallery theorem: its variations, applications and algorithmic aspects. PhD thesis, The Johns Hopkins University (1984)
[5] Bull, L., Kovacs, T. (eds.): Foundations of Learning Classifier Systems. Studies in Fuzziness and Soft Computing, vol. 183 (2005)
[6] Takadama, K., Terano, T., Shimohara, K., Hori, K., Nakasuka, S.: Making organizational learning operational: Implications from learning classifier systems. Computational & Mathematical Organization Theory 5(3), 229–252 (1999)
[7] Holland, J.H.: Adaptation. In: Progress in Theoretical Biology (1976)

[8] Goldberg, D.E.: Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley Longman, Amsterdam (1989)
[9] Schneider, F.E., Wildermuth, D., Wolf, H.L.: Motion coordination in formations of multiple mobile robots using a potential field approach. Distributed Autonomous Robotic Systems 4, 305 (2000)
[10] Takadama, K., Nakasuka, S., Shimohara, K.: Robustness in organizational-learning oriented classifier system. Soft Comput 6(3-4), 229–239 (2002)
[11] Leal Martínez, D.: Reconfigurable multi-robot society based on lego mindstorms. Master's thesis, Helsinki University of Technology (2009)
[12] Kaltenbrunner, M., Bencina, R.: Reactivision: A computer-vision framework for table-based tangible interaction (2007)

Challenges of the Multi-robot Team in the GUARDIANS Project

Lyuba Alboul[1], Joan Saez-Pons[1], Jacques Penders[1], and Leo Nomdedeu[2]

[1] Centre for Robotics and Automation, Sheffield Hallam University, Sheffield, UK
{L.Alboul,J.Saez-Pons,J.Penders}@shu.ac.uk
[2] Computer Engineering and Science, University Jaume I (UJI), 12071 Castellón, Spain
leo.nomdedeu@gmail.com

Abstract. The GUARDIANS multi-robot team is being designed to be deployed in a large warehouse to assist firefighters in the event or danger of a fire. The large size of the environment together with development of dense smoke that drastically reduces visibility, represent major challenges in search and rescue operations. The GUARDIANS robots act alongside a firefighter and should provide, among others, the following tasks: to guide or accompany the firefighters on the site while indicating possible obstacles and locations of danger and maintaining communications links. In order to fulfill the aforementioned tasks the robots need to be able to exert certain behaviours. Among the basic behaviours are capabilities to unite in a group - generate a formation - and navigate on the site while keeping this formation. The basic control model used to generate these behaviours is based on the so-called social potential field framework, which we adapt to fulfill specific tasks of the GUARDIANS scenario. All of the tasks can be achieved without central control, and some of the tasks can be performed even without explicit communication among the robots. We discuss advantages and shortcomings of our model and present samples of implementation on ERA-MOBI robots, commonly referred to as Erratics.

1 Introduction

The GUARDIANS[1] (Group of Unmanned Assistant Robots Deployed In Aggregative Navigation by Scent) project is an FP6, EU funded project, which aims at developing a team (swarm) of heterogenous autonomous robots to assist fire-fighters in search and rescue operations in an industrial warehouse in the event or danger of fire [9].

The challenge of the GUARDIANS project is to apply the team of robots for performing tasks in a real-life situation.

The GUARDIANS scenario has been chosen after consulting with South Yorkshire Fire and Rescue Service, UK, referred further to as SYFIRE. They indicated that indus-

[1] GUARDIANS runs from 2007 to 2010, and involves the following partners: Sheffield Hallam University (coordinator), Robotic Intelligence Lab, Jaume-I University, Spain; Heinz Nixdorf Institute, University of Paderborn, Germany; Institute of Systems and Robotics, University of Coimbra, Portugal; Space Application Services, Belgium; K-Team Switzerland; Dept. of Electrical and Electronics Engineering, TOBB University of Economics and Technology, Turkey; Robotnik Automation, Spain; and South Yorkshire Fire and Rescue Service, UK.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 112–125, 2009.

trial warehouses in the emergency of fire are a major concern to them. Searching for victims is dangerous due to the several, interrelated, factors. Firstly, enormous dimensions of the warehouses already represent a challenge for a search, which only aggravates by the expected low visibility when smoke develops. Next are the time constraints; one is being the amount of oxygen in the breathing apparatus of a firefighter which suffices for about 20 minutes, another one is the crawling speed if smoke has been developed (approximately 12m a minute) - firefighters can proceed about 240m with a full tank. Taking into account that they have to negotiate the 20 minutes of air between getting in and getting out, the maximum advance they can make is only 120m which is less than the largest dimension of the modern warehouses. Another issue related to the time constraint is such phenomenon as *flashover*, which can occur also very quickly [3].

However, as SYFIRE pointed out that apart from the smoke, the warehouse is, in general, in a normal and orderly state. This implies that ground is easily passable and therefore no particular restrictions on robots motion are imposed; even wheeled mini robots are suitable.

The multi-robot team in the GUARDIANS projects consists mostly of mini-robots Khepera III and middle-sized Erratic robots, presented in Figure 1.

(a) (b) (c)

Fig. 1. Team robots in GUARDIANS, (a) Khepera III (K-TEAM), (b) ERRATIC robot (Videre Design). (c) Robot Guardian.

These robots are intended to be applied in some, possibly large, quantity. An exception is the robot called Guardian, developed by the partner Robotnik Automation. This robot can be a member of a team, but also can perform certain task where a more powerful robot may be needed, such as carrying tools for firefighters.

The paper is organised as follows. Section 2 gives a short introduction to the team robotics, with a focus on the GUARDIANS multi-robot team cooperation and tasks to be performed. Section 3 is dedicated to the description of the basic control model in the GUARDIANS, that provides necessary navigation behaviours required from a heterogenous group of robots in the GUARDIANS scenario. Section 3 proceeds with a description of implementation of the algorithms based on the presented control model, on the Erratics robots, and indicates the encountered challenges. Section 4 discusses in brief current work and concludes the paper.

2 Team Robotics

Team or Collective robotics are divided in two major streams: *accidental or non-intentional cooperation* and *intentional cooperation* [16]. Conventionally, the Swarm robotics paradigm is used for non-intentional cooperation; cooperation just happens and emerges from the group behaviour without being made explicit. Intentional cooperation can be described as combining particular behaviours aiming at an explicit goal. Robots interacting with people can comprise both aspects, whereas people, in general, interact intentionally with robots.

2.1 GUARDIANS Robot Team Cooperation

The GUARDIANS robots team should exercise certain cooperation in order to fulfil the tasks assigned to them. The tasks can be roughly split into two main categories. The tasks of the first category provide direct assistance to fire-fighters, such as guiding a firefighter, accompanying them and indicating them possible obstacles and locations of danger. The second category comprises the so-called supportive tasks that can be fulfilled without a human squad-leader, such as deployment on the site, positioning as beacons and maintaining communication. Some tasks of both categories are overlapped, such as searching and navigating the environment; the main difference is that in the first category the robots act within an immediate vicinity of the human, and therefore their sensor range covers only a relatively small area of the environment, whereas in the second category of tasks the robots can disperse in the site and therefore the perception of the environment is more global.

In both categories both non-intentional and intentional cooperation are applied. Therefore, some developments from the Swarm robotics are used. Swarm robotics research is distinguished by the following criteria [17]: a swarm consists of (i) a large number, of (ii) homogenous, (iii) autonomous, (iv) relatively incapable or inefficient robots with (v) local sensing and communication capabilities.

The GUARDIANS group of robots does not comply directly to this definition. First of all, the group consists of non-homogenous robots (different either by physical parameters, or by their functionality), and human agents can be also part of the group. Secondly, the number of robots in the group may be not very large in particular if robots accompany a firefighter.

However, some characteristics of a swarm are present as well. The GUARDIANS group does not have a predetermined size, and due to huge dimensions of a warehouse a large number of robots may be required to fulfil tasks in the second category (criterion (i)). Communication with the outside might not be possible and the human being will be busy ensuring their own safety, thus autonomy (criterion (iii)) is a requirement. A single robot cannot do much in a large warehouse (criterion (iv)) and as communication cannot be guaranteed the robot cannot but rely on local information (criterion (v)).

Swarm robotics is also often divided into so-called communicative-less and communicative robotics. The former case, in general, means that ‘communication’ is assumed to be implicit, i.e. robots react to each other without explicitly exchanging messages, whereas in communicative swarm robots can exchange information. In GUARDIANS

both types of swarm robotics are used; some more details are given in Section 3 and Section 4.

The GUARDIANS project uses developments of swarm robotics field, whenever is appropriate and in what follows the term 'swarm' is also used to describe corresponding behaviours. Surveys on recent advances and the state-of-the-art in swarms can be found in [4,18,10] and a web database on swarm robotics related literature has been compiled on the dedicated site[2].

In this paper we focus mostly on basic navigation behaviours of multi-robot or human-robot teams, which have to be achieved without central and on-line control. The behaviours described are needed in both categories of GUARDIANS robots' tasks, and they are essential when robots directly assist the firefighter. These behaviours, generally speaking, can be also achieved without explicit communication and therefore can be still applicable when communication links are severed. In this case we can speak of non-intentional cooperation. The generated global behaviour is relatively independent of the number of robots in the team, that makes the team also robust to failures of individual robots. These behaviours can be enhanced if the robots communicate, and thus cooperation becomes intentional. We briefly touch upon this enhancement at the end of Section 3 and in Section 4.

2.2 GUARDIANS Team Description

In the GUARDIANS scenario main performers are robots, humans and obstacles, which we identify as classes of GUARDIANS agents. The classes are: 1) the class of robots r_i, $i = 1,2,\ldots,n$; 2) the class of humans (fire-fighters) h_j, $j = 1,2,\ldots,m$; and 3) the class of obstacles o_k, $k = 1,2,\ldots,l$.

The class of robots, which may be *heterogenous*, can be split in several sub-classes of *homogenous* robots and robots may be either *holonomic* or *non-holonomic*. The agents are situated in a domain $D \subset \quad ^2$. In a real-life situation of fire fighting, humans in general act in groups of two: one person takes the role of the leader and the second is the follower. However, we assume that only one human being is present and the human takes the role of leader. Nevertheless the tasks of the robot team is not just to follow the human but also to assist him/her to navigate safely and prevent the human from colliding with obstacles. To a certain extent, robots take the role of the second firefighter acting as a reference unit. The human does not communicate to the robots and is in this context beyond control and performs two basic behaviours: standing still or moving. The robots have to organize themselves into some kind of formations either surrounding or following the firefighter and maintain this formation throughout.

Robots and humans are referred further to as *active agents*, and obstacles as *passive* correspondingly.

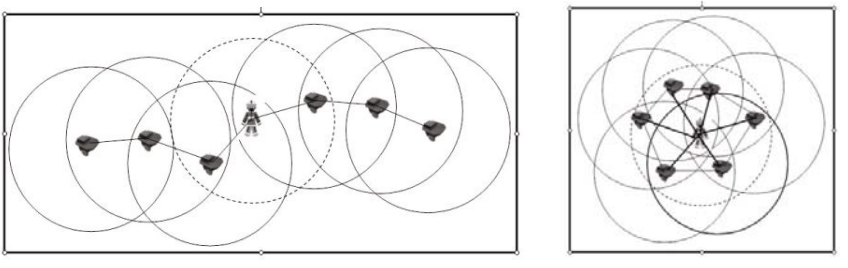
The robots act *independently* and *asynchronously*. We also assume that they are *oblivious*, meaning that they do neither remember observations nor computations performed in previous steps contrary to the assumptions made in [6]. However, this assumption can be relaxed in order to produce more stable behaviours (see Section 4). The sensing range of each robot may vary from zero to infinity. We refer to the sensing

[2] http://swarm-robotics.org/

range of a robot as its *visibility domain*. In the current section the *field of view* of each robot is supposed to be 360 degrees, resulting in a *circular* visibility domain. Let us note that a robot can have several visibility domains each for each sensor installed on a robot. However, we can select one visibility domain as the main one and do all the reasoning with respect to this domain.

We assume that each robot can recognise humans and distinguish robots from obstacles and humans. In computational simulations this is done by indicating the class of an agent, for example, by assigning a specific flag to the agents of the same class. In practice, this can be achieved in various ways. Depending on the sensors a tracking system can be developed, focussing on the characteristics of the stepping feet (of a human) [13]. Other techniques (for communicating robots) which are being developed and tested in the GUARDIANS consortium, include the use of ultrasonic sensors, radio signal intensity, and IR. In our implementation trials the robots are able to localise themselves and the other robots in their visibility domains by using a rough map of the environment provided to them. We do not involve here explicit interaction between a robot team and a firefighter, as human-robot interface development does not belong to the basic behaviours of the robot teams. We refer the reader to related papers of the GUARDIANS consortium members [12,11].

2.3 Human-Multirobot Team Formations



(a) (b)

Fig. 2. Two examples of human-multirobots formations. (a) Maximal formation, (b) Minimal formation.

In the GUARDIANS scenario, formations are defined as groups of agents establishing and maintaining a certain configuration without a predetermined shape (opposite to the assumption expressed, for example, in [7,2]) but without spreading too much from each other. One of requirement of the GUARDIANS (human)-multirobot formation is its adaptability: formation can be stretched and deformed when obstacles are in the close vicinity since the firefighter has to be protected and escorted all times. Considering a group of agents as a graph (network) where each agent represents a node, and agents are interconnected via their visibility domains, we can define formation as follows:

Definition 1. *The GUARDIANS formation represents a connected graph, where nodes are robots or a human and edges are virtual links between the nodes, with a property*

that each edge is situated in the intersection of the visibility domains of nodes which it connects.

The definition implies that the distance r_i between neighboring agents (either a robot or a human) does not exceed the certain value d_{max}. d_{max} can be defined to be smaller or equal either to the (smallest) radius of the visibility domains or reaction domains of agents. This definition is similar to the definition of the formation given in [8].

Neither initial positions, nor final positions of agents are predefined. To some extent, this definition complies with the definition proposed in [5], where the group determines autonomously the most appropriated assignment of positions in the formation.

The definition of formation given above can be specified further.

Indeed, both configurations presented in Figure 2 comply with Definition 1. Both configurations can be useful for GUARDIANS scenario. The one on the left can occur when a group passes a narrow passage, and the one the right may be desirable in an open space. The connected graph that describes a formation may contain loops.

We, therefore, define a *degree* g, $g = 1,...,n-1$ of a formation by the minimum number of the visibility domains that contain a spanning tree of the graph of a formation. We put $g = \infty$ if there are agents without virtual links in their visibility domains (essentially, it means that there is no formation). In Figure 2 a *maximal* ($g = n-1$) and *minimal* ($g = 1$) formations are depicted. In the latter case the 'visibility' domain of the firefighter is depicted by a dashed line and this visibility domain contains the most 'compact' spanning tree.

3 Control Model

In order to achieve formation generation and their maintenance we apply robot-robot and robot-fire-fighter avoidance/attraction as well as robot-obstacle avoidance. Our method is based on the social potential field framework, which was introduced by Reif and Wang [15].

The method for generating navigation behaviour patterns in mixed human-robot groups in complex environments has been initially discussed in [1].

We define **Robot-Human, Robot-Robot and Robot-Obstacle Potential Functions**. The robots have to avoid collision with the human and obstacles, and at the same time to be able to approach and keep the human within their sensor range. While robots 'see' the fire-fighter they will execute repulsion behaviour among themselves; however if a group of robots have lost a fire-fighter in their visibility domain, we would like that robot do not disperse and therefore attraction behaviour is executed towards the robots in a robot's visibility domain. Therefore the aforementioned functions are defined as follows:

1. Robot-human potential function P_{Human} between the robot r_m and the Human H is:

$$P_{Human}(d^H_{r_m}) = \frac{1}{(k_{hrr}(d^H_{r_m} - w_{hrr}))^2} + \frac{1}{(k_{hra}(d^H_{r_m} - w_{hra}))^2} \quad (1)$$

where k_{hrr}, k_{hra}, w_{hrr} and w_{hra} are scaling parameters, and $d^H_{r_m}$ is the distance between the robot r_m and the human H.

2. Robot-Robot Potential function P_{Robot} between the robot r_m and the robot r_i is, in the presence of the human in a robot visibility domain, is defined

$$P_{Robot}(d_{r_m}^{r_i}) = \frac{1}{(k_{rr}(d_{r_m}^{r_i} - w_{rr}))^2} \tag{2}$$

where k_{rr} and w_{rr} are scaling parameters and $d_{r_m}^{r_i}$ is the distance between the robot r_m and the robot r_i. Obviously $d_{r_m}^{r_i} = d_{r_i}^{r_m}$. In this case only the repulsion term is present. If a robot loses the human, then P_{Robot} has a similar form to P_{human}.

In the presence of a human we assume that robots avoid each other, by exerting on each other the repulsive force $IR_{(m,i)}$, the magnitude of which is determined given by the derivative $P_{rr}(r_{mi})$ of $P_{Robot}(d_{r_m}^{r_i})$ with respect to $d_{r_m}^{r_i}$.

In the absence of a human in the visibility domain of a robot, the force acting on the robot by other robots in its visibility domain becomes a combination of attraction and repulsion similar to the potential function between a robot and the human in order to avoid spreading robots in the site. The corresponding function is:

$$P_{Robot}(d_{r_m}^{r_i}) = \frac{1}{(k_{rr}(d_{r_m}^{r_i} - w_{rr}))^2} + \frac{1}{(k_{ra}(d_{r_m}^{r_i} - w_{ra}))^2} \tag{3}$$

where k_{rr}, k_{ra}, w_{rr} and w_{ra} are scaling parameters, and $d_{r_m}^{r_i}$ is the distance between the robot r_m and the robot r_i.

3. Robot-Obstacle Potential Function P_{Robot} is defined between the robot r_m and the obstacle O_s as

$$P_{Obstacle}(d_{r_m}^{O_s}) = \frac{1}{(k_{ro}(d_{r_m}^{O_s} - w_{ro}))^2} \tag{4}$$

where k_{ro} and w_{ro} are scaling parameters and $d_{r_m}^{O_s}$ is the distance between the robot r_m and the obstacle O_s. We assume that robots avoid the obstacles and therefore do not introduce the 'attraction' term.

The social potential function P_{Social} of r_m is defined as the sum of the aforementioned potential functions:

$$\begin{aligned} P_{Social}(X_{r_m}) &= P_R^O(\mathbf{X}_{r_m}) + P_{r_i}^{r_j}(\mathbf{X}_{r_m}) + P_r^H(\mathbf{X}_{r_m}) \\ &= \sum_{s=1}^{S} P_{Obstacle}(d_{r_m}^{O_s}) + \sum_{j=1, j\neq i}^{M} P_{Robot}(d_{r_m}^{r_i}) + P_{Human}(d_{r_m}^{H}) \end{aligned} \tag{5}$$

The artificial force $\overrightarrow{F}_{Arti}(X_{r_m})$ which is 'acting' on robot r_m is, therefore, computed as the sum of gradients of corresponding potential functions:

$$\overrightarrow{F}_{Arti}(X_{r_m}) = \overrightarrow{F}_{Arti\_Obstacle}(\mathbf{X}_{r_m}) + \overrightarrow{F}_{Arti\_Robot}(\mathbf{X}_{r_m}) + \overrightarrow{F}_{Arti\_Human}(\mathbf{X}_{r_m})$$

Table 1. Values of the parameters used in the potential functions employed for simulation

Potential Function	Parameter	Value
Robot-Obstacle	k_{ro}	$= 5.00$, $w_{ro} = 0.49$
Robot-Robot	k_{rr}	$= 2.00$, $w_{rr} = 0.98$
	k_{ra}	$= 2.00$, $w_{ra} = 4.00$
Robot-Human	k_{hrr}	$= 5.00$, $w_{hrr} = 0.82$
	k_{hra}	$= 2.00$, $w_{hra} = 4.00$

Parameters. The parameters of all the employed potential functions are shown in the table 1.

This selection is loosely based on the specifications and characteristics of the considered system. We use the dimensions of an Erratic robot (given in Table 2), but it can be easily adapted to other types of robots.

Table 2. Basic parameters of the ERA-MOBI robot

ER-AMOBI	Parameter Value
Dimensions	$L = 40$cm, $W = 41$cm, $H = 15$cm
Maximum Speed	$2ms^{-1}$
Sensors	Laser Range Finder-Hokuyo(range 4m)

For example, the robot's size determines the value of the contact distance, i.e. for the robot-obstacle potential function w_{ro} represents the distance at which the edges of the robot and the obstacle would come into physical contact ($w_{ro} = 0.49$). The same selection criteria applies to the robot-obstacle (w_{rr},w_{ra}) and robot-human (w_{hrr},w_{hra}) contact distances. The parameter k_{ro} in the potential function (4) determines at which distance the repulsive potential starts pushing the robot away from the obstacle. Choosing $k_{ro} = 5$ means that the robot will not start avoiding the obstacles until approximately $d_{R_m}^{O_s} = 1.5$ meters. The same selection criteria has been applied for the parameters of the remaining potential functions.

Algorithm. The pseudocode of the algorithm that uses the Social Potential forces approach for the implementation in the real-world scenario can be seen in Algorithm 1. Each robot calculates the resultant social potential force (F_x, F_y) which dictates the velocity of each robot (v_x, v_y). We use a discrete-time approximation to the continuous behaviour of the robots, with time-step Δt. The speed of the robots is bounded to a maximum velocity V_{max}. The output of the *compute motion* algorithm is the direction and the speed of the robot.

Our algorithms have been tested using and Player/Stage software[3] that allows their direct application to real robots. Simulation results comply with our theoretical considerations regarding formation generation and keeping, and show that our algorithms are

[3] http://playerstage.cvs.sourceforge.net/viewvc/playerstage/papers/

Algorithm 1. Compute motion

for all robots but current robot i **do**
  determine the distance r to robot j
  determine the spherical coordinate θ to r_j
  $netForce = SocialPotentialForce(r)$
  $F_x \Leftarrow F_x(netForce) \times \cos(\theta)$
  $F_y \Leftarrow F_y(netForce) \times \sin(\theta)$
end for
$\Delta v_x \Leftarrow \Delta t \times F_x$
$\Delta v_y \Leftarrow \Delta t \times F_y$
$v_x \Leftarrow v_x + \Delta v_x$
$v_y \Leftarrow v_y + \Delta v_y$
if $\|v\| > V_{max}$ **then**
  $v_x \Leftarrow (v_x \times V_{max})/\|v\|$
  $v_y \Leftarrow (v_y \times V_{max})/\|v\|$
end if
$speed \Leftarrow \|v\|$
$direction \Leftarrow \tan^{-1}(v_y/v_x)$
move the robot with speed and direction

robust and capable to deal with teams of different sizes and failure of individual agents, both robots and humans.

Stability Considerations. The presented control model has been analysed for stability. The stability analysis is based on geometric concepts which avoids heavy computation while providing qualitative proofs of attainability of desired formations under certain conditions. Some results on stability analysis were presented in [1]. The main conclusions are the following (sufficient conditions for formation maintenance):

1. In the absence of obstacles the robot are always gathered around the human, forming a minimal formation, if at the initial step the robots and human are in formation according to our definition;
2. In the presence of obstacles, if the human (at rest) and robot agents are in formation at any step, all robots will gather around the human. The deg_{fin} of the final formation does not exceed the deg_{init} of the initial one;
3. If the human moves and robots are in formation at any step, the robot team will follow the human.

In the cases 2) and 3) some robots may be lost due to the fact that an obstacle will appear in their visibility domain which may break the formation. In Section 4 we present an adaptation of our framework by introducing some, still very limited, 'memory' to a robot.

3.1 Examples of Implementation

We have tested our algorithms on Erratic's mobile platforms. Four Erratic platforms equipped with: 1) On board computer equipped with wi-fi; and 2) Hokuyo laser-range

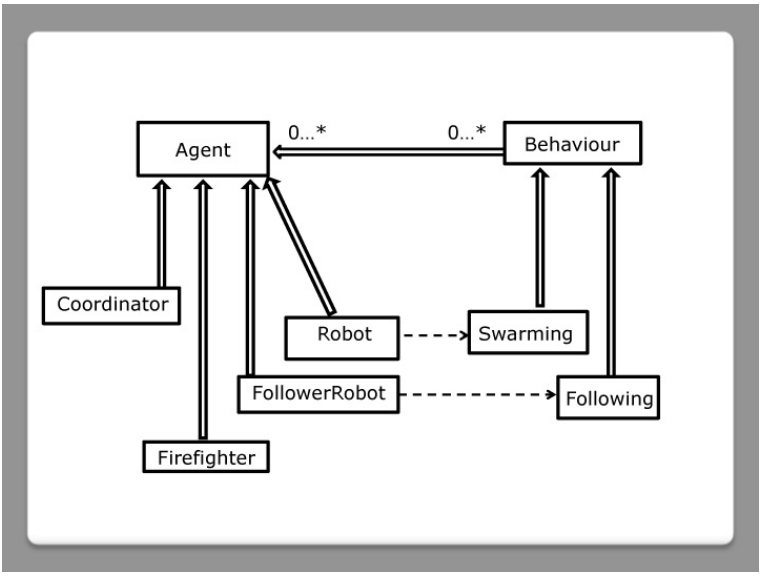


Fig. 3. Agents chart used in demo's

finder, 'participated' in the tests. One of the robot 'played' the role of a firefighter. The main goal of the implementation was to demonstrate that robots are able to generate a formation and keep the formation while follow a human (or a leader robot).

The implementation of our algorithms in the real-world scenario with the ERRATIC robots represented a challenging issue. Most of the efforts focussed on achieving a reliable way of detecting the components of the multi robot human team without using any sort of tracking system. The considered solution implied to design an architecture environment capable of implementing different robot behaviors (aggregation and following), handle communication, run distinct robot navigation algorithms (localization and collision avoidance), define different agent types, interact with the hardware involved (actuators and sensors), interface with the users and everything combined with different software platforms (Player, Javaclient and JADE).

In order to mimic relative robot detection and distance estimation that is still under validation, robots were provided with a map of the environment in which they localise themselves by using the Adaptive Monte-Carlo localisation method.

JADE (Java Agent Development Environment)[4] was used to take care of the agent's life-cycle and other agent-related issues. JADE provides a runtime environment and agent communication and management facilities for rapid and robust agents-based developments. In our demonstration we have developed 4 different types of agents, as shown in Figure 3, each one having a clear role in the demo. Note that agents here are different from our the classes of agents determined in Section 2.

[4] http://jade.tilab.com/papers-exp.htm

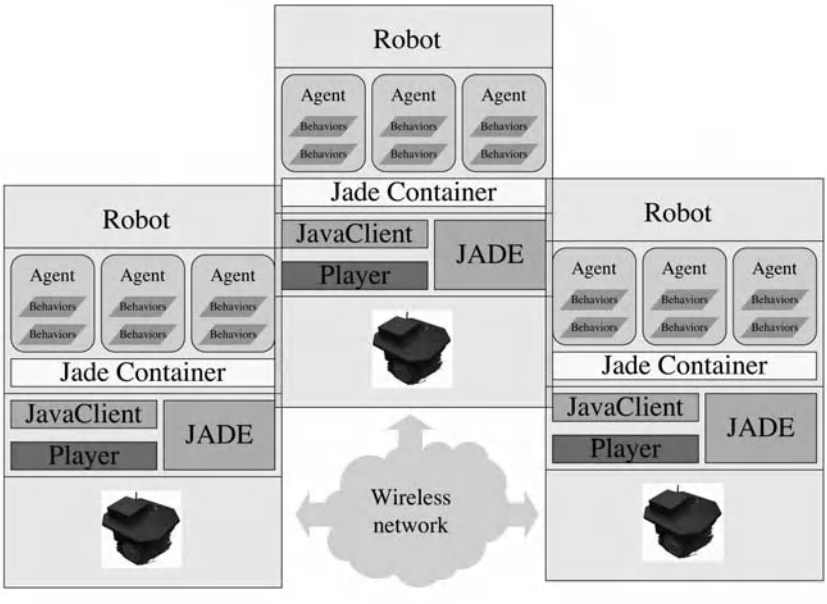


Fig. 4. Software components used for demo's

(a) (b)

Fig. 5. Two snapshots of experiments on formation generation and maintenance using Erratic's . (a) Formation generation, (b) Formation maintenance.

Each agent is composed of a set of behaviours that determines how this agent acts or reacts to stimuli. For our demo we have developed several communication, swarming, and following behaviours, and assigned them in different ways to different agent types to get a set of multi-functional agents. By doing so, we are able to share the robots and human poses through the whole team, allowing swarming techniques to take advantage of these essential data.

In Figure 4 we can see the combination of software pieces that plays in our team. Player, from Player/Stage, acts as a Hardware Abstraction Layer, allowing us to forget

specific hardware problems. JavaClient allows us to connect to the Player server from a Java environment, while JADE provides us the ability to use Agents. In terms of runtime, Agents, and their behaviours, run on top of an agent container provided by the JADE, making use of the JavaClient to access Player facilities.

Some of implementations were demonstrated during the evaluation of the GUARDIANS project's progress in Brussels in January 2009, and were met enthusiastically by the audience. In Figure 5 snapshots of video of the experiments on formation generation and keeping on a group of Erratic's robots are presented. The robots with a flag 'plays' the role of the firefighter.

4 Work in Development

The control models, based on artificial potential fields, have drawbacks such as getting in local minima. Our control model is not an exception, however we established the important condition that would prevent undesirable local minima: the robots have to be in formation, as defined in Section 2, at any step. It means that a robot may not 'sense' the leader/human/goal at any step, but a chain (path) must exist consisting of 'formation' edges, that connects the robot to the leader/human/goal. Some authors realise this (without explicitly formulating neither sufficient nor necessary conditions). However, in order to avoid local minima, all robots are either assumed to be able to sense the leader or its equivalent at any step, or be able to reproduce the previous steps of the leader [14,6]. This leads to extensive computation and higher complexity of the corresponding algorithms.

We presently work on a method which only slightly increases necessary computations. We allow a robot to 'remember' the force between a robot and the leader at the previous step. A robot at each step calculates the force/direction between itself and the leader and remember this force till the next step. The force which was calculated at the step before the previous one the robot 'forgets'. Therefore, at each step a robot has two 'human-robot' forces in its memory. If a robot senses the leader it applies the current force, if the robot is lost, it applies the force at the previous step. An informal idea behind this approach is that a robot may lose the leader only if an obstacle or another robot 'blocks' the 'view' of the leader. If the obstacle is the reason of losing the leader, it means that the leader turns along the obstacle, otherwise it will not disappear. Therefore, if a robot continues to follow the previous direction, it may achieve a point where it will sense the leader again. If robots communicate than this method can be applied to several 'lost' robots as well. In this case the robots do not attract to each other, but follow only the direction calculated at the previous step. While the formal justification of this method is still under consideration, we have tested this new approach on a group of three robots (one being a leader) in a simulated setting. Snapshots of a simulation in Player/Stage is presented in Figure 6. Currently we are working on the extension of the method to robots team of arbitrary sizes. We are also developing a framework that will combine the described basic behaviours with other behaviours, such as wall following, and allow easy 'switch' between behaviours.

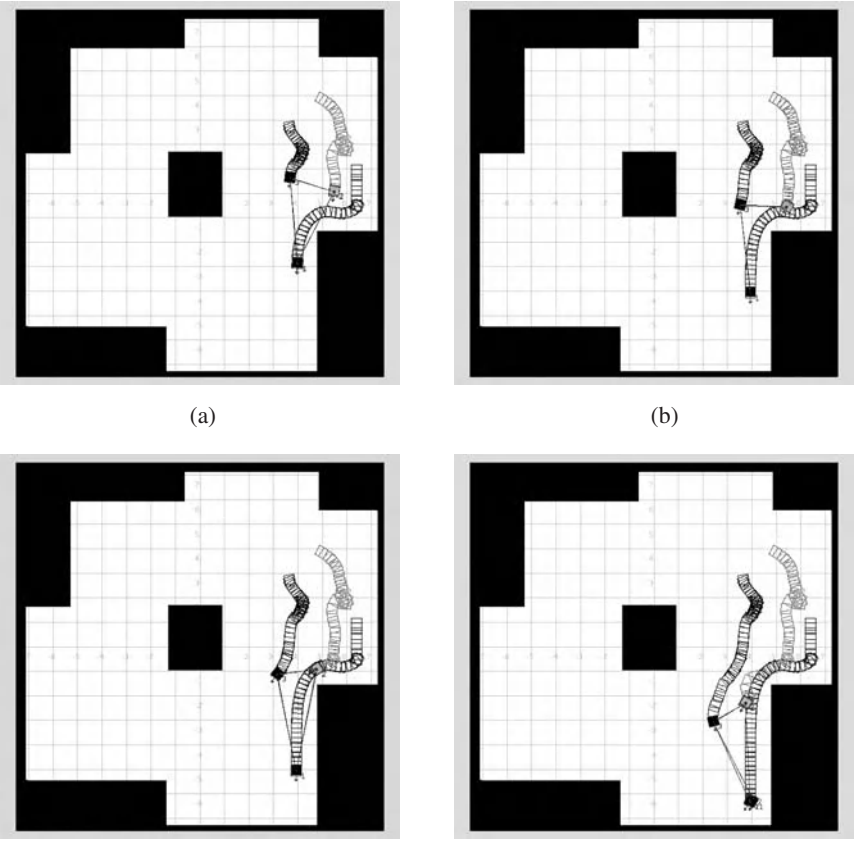


Fig. 6. Behaviour pattern of a robot team by using the adapted algorithm. (a) Robots are moving in a triangular formation, following the leader; (b) One of the robot-followers lost the leader; (c) The robot 'found' the leader; (d) The robots resumed moving in a triangular formation.

Acknowledgements

The authors wish to acknowledge that the GUARDIANS project is a cooperation between staff at all project partners, as well as Master students, Chathura Wijaya Sri Dolahamuna, Carlos Bataller Narbó, Manuel Espinosa Prieto, Miquel Cerezo Gandía and José María Blasco Igual, who worked on different aspects of our control model. The contributions of all of them have been essential to the development of the project and this paper. The EU FP6 framework provided financial support.

References

1. Alboul, L., Saez-Pons, J., Penders, J.: Mixed human-robot team navigation in the guardians project. In: International Workshop on Safety Security and Rescue Robotics, SSRR 2008 (2008)
2. Baldassare, G., Nolfi, S., Parisi, D.: Evolving mobile robots able to display collective behaviour. In: Hemerlijk, C. (ed.) International Workshop on Self-Organisation and Evolution of Social Behaviour, pp. 11–22. ETH, Zurich (2002)
3. Clark, W.: Firefighting Principles and Practices (1991)
4. Dorigo, M., Sahin, E.: Special issue on swarm robotics. Autonomous Robots 17(2-3) (September 2004)
5. Lemay, M., et al.: Autonomous initialization of robot formations. In: Proceedings of Robotics and Automation. ICRA 2004, pp. 3018–3023 (2004)
6. Fazenda, P.V., Lima, U.: Non-holonomic robot formations with obstacle compliant geometry. In: Proc. of IAV 2007 (2007)
7. Gazi, V., Passino, K.M.: Stability analysis of social foraging swarms. IEEETSMC: Part B 34(1), 539–557 (2004)
8. Pappas, G.J., Tanner, H.G., Kumar, V.: Leader-to-formation stability. IEEE Transactions on Robotics and Automation 20, 443–455 (2004)
9. Roast, C., Penders, J., Alboul, L., Cervera, E.: Robot swarming in the Guardians project. In: ECCS 2007. European Conference on Complex Systems 2007 (2007)
10. Kumar, V.J., Leonard, N.E., Morse, A.S. (eds.): Cooperative Control: 2003 Block Island Workshop on Cooperative Control. LNCIS, vol. 309. Springer, Heidelberg (2005)
11. Naghsh, A.M., Roast, C.R.: Designing user interaction with robots swarms in emergency settings. In: Proceedings of the 5th Nordic Conference on Human-Computer interaction: Building Bridges (2008)
12. Gancet, J., Tanoto Naghsh, A.M.: Analysis and design of human-robot swarm interaction in firefighting. In: Proc. of IEEE Conference of Robot-Human Interactions, ROMAN (2008)
13. Nomdedeu, L., Sales, J., Cervera, E., Alemany, J., Sebastia, C., Ilzkovitz, M., Penders, J., Gazi, V.: An experiment on squad navigation of human and robots. In: Penders, et al. (eds.) Robotics for Risky Interventions and Surveillance of the Environment, Universitat Jaume I (2008)
14. Fiorelli Ögren, E., Leonard, N.E.: Cooperative control of mobile sensor networks: Adaptive gradient climbing in a distributed environment., vol. 49, p. 1292
15. Reif, J.H., Wang, H.: Social potential fields: A distributed behavioral control for autonomous robots. Robotics and Autonomous Systems 27(3), 171–195 (1999)
16. Rybski, P., Larson, A., Lindahl, M., Gini, M.: Performance evaluation of multiple robots in a search and retrieval task. In: Workshop on Artificial Intelligence and Manufacturing, pp. 153–160. AAAI Press, Menlo Park (1998)
17. Sahin, E.: Swarm robotics: From sources of inspiration to domains of application. In: Swarm Robotics: SAB 2004 International Workshop, Santa Monica, CA, USA, July 17, 2004. LNCS, pp. 10–20. Springer, Heidelberg (2005) (Revised Selected Papers)
18. Sahin, E., Spears, W.M. (eds.): Swarm Robotics 2004. LNCS, vol. 3342. Springer, Heidelberg (2005)

Modelling and Control of a Train of Autonomous Electric Vehicles

Kamel Bouibed[1], Abdel Aitouche[2], and Mireille Bayart[1]

Laboratoire d'Automatique Génie Informatique et Signal (LAGIS)
[1] Polytech'Lille Engeneering School, 59650 Villeneuve d'Ascq, France
{kamel.bouibed,mireille.bayart}@polytech-lille.fr
[2] High School Engineer (HEI), 59046 Lille, France
abdel.aitouche@hei.fr

Abstract. While driving, the position and velocity of the vehicles in front of one greatly influence the way he/she drives the car. The purpose of this paper is to design a stable train of vehicles using longitudinal control, with a constant spacing between each vehicle. The method proposed here makes use of relative spacing and velocities between the vehicles to determine any changes (if any) that are required in the condition of the following vehicle to ensure a safe and comfortable journey. We will use a second order sliding mode control to maintain the platoon stable and the inter-vehicular spacing to the safe distance. Simulation results will be given to show the effectiveness of the method.

Keywords: Robucar, Platooning, Dynamic Model, Inter-vehicular Distance, Spacing Deviation, Autonomous vehicles.

1 Introduction

In the present day and age, with the number of vehicles playing on the roads increasing exponentially, the number has easily exceeded the capacity of the present infrastructure. A few solutions to this problem are the creation of new roads or for optimal utilization of the present resources that are available to us. However, owing to spatial, environmental and financial constraints, construction of new roads is not feasible everywhere. Hence it is high time for optimizing the use of vehicles on the roads with, as little space as possible, being wasted while commuting from one place to another at a decent speed.

This has led to research in the field of Artificially Intelligent (AI) cars in the present times. These have been shown to have a greater efficiency than human-driven vehicles be it speed, control or prevention of accidents. This is because various human factors such as human errors and reaction times are considerably reduced in the case of AI cars.

This can be achieved by creating a train of vehicles (at least three vehicles) in which the leader is given a specific route on which it's followed by the vehicles that are trailing behind. When it comes to creation of a stable platoon of vehicles, the prerequisite is to have a constant speed and distance [1] between the vehicles while at

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 126–134, 2009.

the same time providing traveling comfort to the people in the car. The basic principle that is followed when it comes to this is the exchange of data between the vehicle that is following and the vehicle immediately in front of it to determine the motion of the vehicle that is following. In this paper, a constant spacing policy is proposed with a constant distance between each vehicle of the platoon being maintained throughout the journey, along with a deviation value in case of errors or malfunction [2].

In the following system, the relative spacing and the velocity are used to determine the behaviour of the following vehicle. The relative spacing between any two vehicles is regularly compared to a set value that is kept as the constant spacing that is allowed between any two vehicles. Should the value of the spacing change from the constant value [3], the following vehicle will execute the necessary maneuvers so as to bring the spacing back to the constant value.

The paper is organized as follows: In Section 2 the problem statement is mentioned and the necessary conditions explained. Next, the dynamic longitudinal vehicle model is illustrated in Section 3. This is followed by the vehicle following strategy, and the sliding mode (SM) control design, are explained in Section 4. In Section 5, the results of the simulation carried out on a computer along with its discussion are given so as to prove the viability of the specified control law.

2 Problem Statement

To have a stable platoon of vehicles, each vehicle must maintain a constant spacing between the vehicle in front as well as the vehicle behind it. The factors that play an important role in determining this are the vehicles performance, reaction times, the road surface etc.

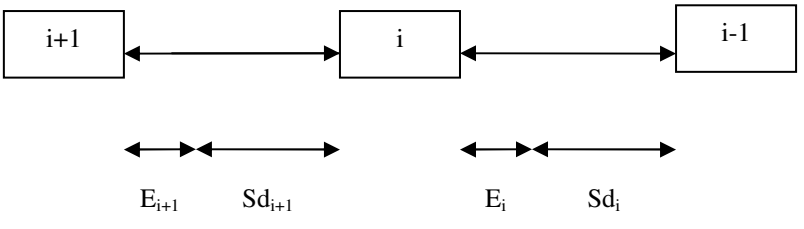


Fig. 1. Inter-vehicular distance

In this paper, the constant inter-vehicular distance is measured by taking the difference of the positions of the centre of gravities of the vehicles, that is:

$$Sd_i = x_i - x_{i-1} - L \tag{1}$$

Where, 'x_i' is the 'x coordinate of the centre of gravity of the i^{th} vehicle'

'x_{i_1}' is the 'x coordinate of the centre of gravity of the i-1^{th} vehicle'

'L' is the 'length of the vehicle'

Along with this, a spacing deviation has been taken in case of malfunction or any other error that might creep in, given by *E*.

For the system to be stable, the following conditions must be satisfied:

1. The spacing deviation must be regulated to zero with the passage of time, that is:

$$\lim_{t\to\infty}\left(E_i(t)\right)=0 \tag{2}$$

The relative velocity between the i^{th} vehicle and the $i\text{-}1^{th}$ when the leading vehicle moves at constant velocity must be regulated to zero, that is:

$$\lim_{t\to\infty}\left(v_i(t)-v_{i-1}(t)\right)=0 \tag{3}$$

for constant v_{i-1}.

3 Vehicle Dynamic Model

RobuCar is a 4x4 electrical vehicle, with four electromechanical wheel systems, produced by the Robosoft society, with each motor part being actuated by a DC motor delivering a relative important mass torque. Front and rear steering are controlled by their respective electric steering jack. Measurement of variables such as actuators, wheels velocities and rear and front steering angles can be done with the help of the incremental encoders.

While modeling a RobuCar, the following assumptions must be taken into consideration:

- The pitch phenomenon is not taken into consideration.
- The road is supposed to be uniform and horizontal throughout so that the suspension dynamics may be neglected.
- Other factors like electromechanical systems, longitudinal, lateral and yaw dynamics are already modeled.
- Steering angles are supposed to be small and hence can be neglected.
- The front and rear lateral forces are taken as equal.

The model of a RobuCar can be said to comprise of the following dynamics:

- Wheel Dynamics
- Longitudinal and Lateral Dynamics
- Front and Rear Steering Dynamics (Not considered here owing to the steering angle being negligible)
- Yaw Dynamics

3.1 Wheel Dynamics

For the i^{th} vehicle, let ω_{ij} be the front (f) or rear (r) velocity of wheel *j* and U_{ij} represent the control input torque T_i applied straightly to wheel *j*. *j = {1,…4}*

$$\begin{cases} \dot{\omega}_{i1f} = \dfrac{1}{J_{i1}}[-f_{i1}\omega_{i1f} + R_i Fx_{i1} + U_{i1}] \\ \dot{\omega}_{i2f} = \dfrac{1}{J_{i2}}[-f_{i2}\omega_{i2f} + R_i Fx_{i2} + U_{i2}] \\ \dot{\omega}_{i3r} = \dfrac{1}{J_{i3}}[-f_{i3}\omega_{i3r} + R_i Fx_{i3} + U_{i3}] \\ \dot{\omega}_{i4r} = \dfrac{1}{J_{i4}}[-f_{i4}\omega_{i4r} + R_i Fx_{i4} + U_{i4}] \end{cases} \tag{4}$$

3.2 Longitudinal and Lateral Dynamics

Taking v_{ix} and v_{iy} as respectively the longitudinal and lateral velocities, and using the fundamental principles of dynamics, the longitudinal and lateral dynamics can be obtained as,

$$\begin{cases} \dot{v}_{ix} = \dfrac{1}{M_i}[(Fx_{i1} + Fx_{i2} + Fx_{i3} + Fx_{i4}) - f_{ix}v_{ix}] \\ \dot{v}_{iy} = \dfrac{1}{M_i}[2(Fy_{i1} + Fy_{i2}) - f_{iy}v_{iy}] \end{cases} \tag{5}$$

3.3 Yaw Dynamics

Applying the Kinetic Moment's Theorem at the centre of gravity of the ith vehicle, we obtain:

$$\dot{\Gamma}_i = \frac{2a_i}{Iz_i} Fy_{i1} - \frac{2b_i}{Iz_i} Fy_{i2} + \frac{d_i}{2Iz_i}(Fx_{i1} - Fx_{i2} + Fx_{i3} - Fx_{i4}) - \frac{f_{i\Gamma}}{Iz_i}\Gamma_i \tag{6}$$

Where Γ_i and $\dot{\Gamma}_i$ are respectively yaw velocity and yaw acceleration. Fx_{ij} and Fy_{ij} are respectively longitudinal and lateral tire/road contact forces. They can be estimated by an empirical model [5] or by using sliding mode observers [6].

4 Vehicle Following Strategy

In order to achieve a stable train of vehicles, there exist many strategies. The basic principle behind all these strategies consists of a leading vehicle that follows the specified trajectory whereas the rest of the vehicles behind follow it. Some of the methods for achieving this can be by inter-vehicle data transmission, use of Real Time Kinetic (RTK) GPS system [5], by checking the behaviour of the vehicle immediately in front [6] or by having the leader under manual control with the rest of the vehicles following it [7].

This paper adopts the method of checking the behaviour of the vehicle immediately in front. The leading vehicle is given a trajectory that it follows. The rest of the vehicles that are behind it mimic the behaviour of the leading vehicle by checking various parameters such as the inter-vehicular distance or the relative velocity between the

vehicles. These values are then compared to a pre-ordained constant value. Should the measured value deviate from the corresponding value, the necessary corrective measures will be automatically applied in the vehicle's behaviour.

In this paper, the inter-vehicular distance is constantly measured and compared with a constant value that is obtained taking into consideration all the factors that come into play while driving a car such as vehicular capabilities, reaction times, road surfaces etc. Also, to prevent slinkiness [4] and other oscillatory defects, the speeds of all the vehicles are kept the same. Should the inter-vehicular distance deviate from the specified constant, the following vehicle executes the necessary maneuvers to bring the distance to the safe value. For example, should the value go below the safety distance, then there is a chance of a crash between the vehicles should any problem occur. Therefore, the following vehicle will decrease its speed i.e. decelerate until the distance between the vehicles comes back to the safe value, after which it again accelerates to the speed of the leading vehicle so as to keep the system moving.

Also, a deviation from the safe value of the inter-vehicular distance is also taken in case of malfunctioning of the sensors or the vehicle. This ensures more than enough space for the following vehicle to take evasive measures and prevent any mishap from happening.

A second order sliding mode technique [8] is used for longitudinal control. This is achieved by applying the algorithm not directly on the control input u but on its derivative. This method is adapted for non-linear system, and permits a considerable reduction of the inherent chattering phenomenon [9]. This is achieved by applying the algorithm not directly on the control input U but on its derivative. Besides sliding modes are robust to sensor noise which is very important for experiments knowing that several sensors are required for longitudinal control.

Maintain the inter-vehicular distance Sd_i constant means to maintain the distance between the wheel ij of the i^{th} vehicle and its homologue of the $i+1^{th}$ vehicle Sw_i constant too. This also means to maintain the difference between the velocities of the two wheels at zero. So we write:

$$Sd_i = C_1 \Rightarrow Sw_i = C_2 \Rightarrow R\omega_{i1f} - R\omega_{(i+1)1f} = 0 \tag{7}$$

where C_1 and C_2 are constants.

From (7), we have the following equation:

$$\omega_{i1f} - \omega_{(i+1)1f} = 0 \tag{8}$$

System inputs U_{ij} are T_{ij} but in an experimental objective it is easier to use wheel angular acceleration input which is easier to measure. Then in order to control the vehicle from a leader speed profile, the desired input of each wheel of the vehicle is:

$$u_{ij} = \alpha_{ij} = \frac{1}{J_{ij}}[-f_{ij}\omega_{ij} + R_i Fx_{ij} + T_{ij}] \tag{9}$$

where $\alpha_{ij} = \dot{\omega}_{ij}$ is the angular acceleration of the wheel j of the vehicle i.

The sliding surface chosen for the control the wheel *i* of the followed vehicle *j* is of the form:

$$S_{ij} = \omega_{(i-1)j} - \omega_{ij} \tag{10}$$

The control input of the wheel *j* of the vehicle *i* appears at the first derivative of S_i. So the system has a relative degree of 1. The Twisting Algorithm [8] is then applied to the vehicle model to simulate an automatic vehicle following control. Tis algorithm is:

$$\dot{u}_{ij} = \begin{cases} -u_{ij} & if \quad |u_{ij}| > |u_{ij\_eq}| \\ -K_m sign(S_{ij}) & if \quad S_{ij}\dot{S}_{ij} > 0 \; and \; |u_{ij}| \le |u_{ij\_eq}| \\ -k_m sign(S_{ij}) & if \quad S_{ij}\dot{S}_{ij} \le 0 \; and \; |u_{ij}| \le |u_{ij\_eq}| \end{cases} \tag{11}$$

where u_{ij_eq} is the equivalent control which is computed such that $\dot{S} = 0$. The gains k_m an K_m are determined in order to fulfil the four conditions of applications of Twisting Algorithm which is a second order sliding mode algorithm that converges in finite time [9, 10]. These four conditions summarized below are needed to limit variables like vehicle (or wheel) speed and acceleration.

$$\begin{cases} K_m > k_m > 0 \\ k_m > 4\dfrac{C_m}{s_0} \\ k_m > \dfrac{C_0}{c_m} \\ K_m > \dfrac{C_m k_m}{c_m} + \dfrac{2C_0}{c_m} \end{cases} \tag{12}$$

where C_m, c_m, s_0 and C_0 are four constants defined by the assumptions of the work.

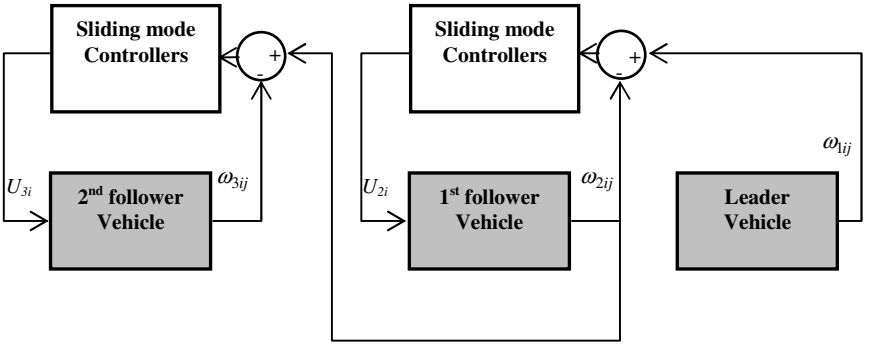


Fig. 2. Block diagram of a platoon with three vehicles with SM controllers

5 Simulation

To show the effectiveness of the method, we have simulated a platoon of three vehicles using Matlab Simulink. Each vehicle is represented by his dynamic model given previously. We assume that the leading vehicle follows autonomously a desired trajectory. At time *t=0*, the three vehicles start with a spacing of *5m* and with a constant velocity. All wheels have the same velocity at startup. To turn left, one increase the velocity of the front right wheel of the leader from *15rad/s* to *25rad/s* between the instants *t=50s* and *t=135s* as it is shown in figure 3.(a). The other two vehicles follow autonomously the leader keeping the spacing at *5m* as it is shown in figure 3.(d).

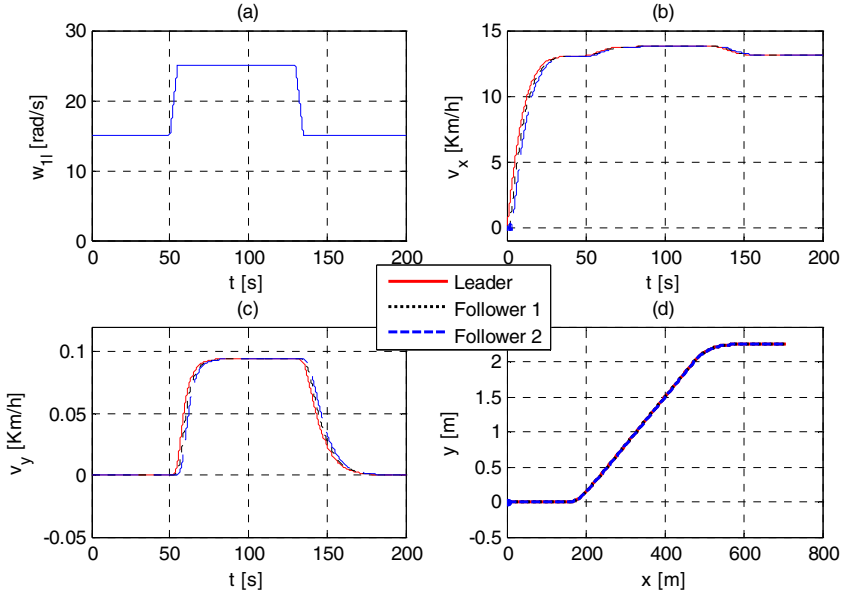


Fig. 3. (a): Imposed front right wheel velocity of the leader, (b) and (c): longitudinal and lateral velocities of the three vehicles respectively, (d): trajectories of the three vehicles

Figure 3. (b) shows the longitudinal velocities of the three vehicles. We note that they increase at time t=50s, the same time when the imposed velocity of the front right wheel oh the leading vehicle increase. The lateral velocities, figure 3. (c), are null before increasing the velocity of the wheel, but after that, we observe that these velocities are different from zero.

Figure 4 shows the velocities of the four wheels of each vehicle. They follow the desired velocities imposed for the leader.

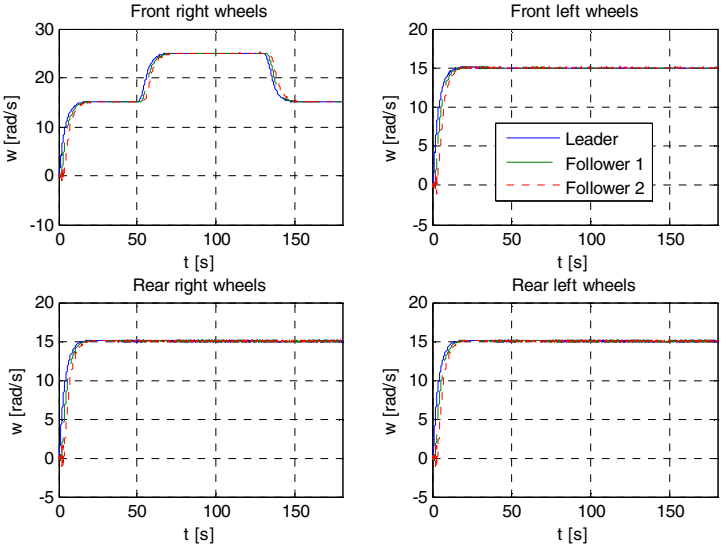


Fig. 4. Wheel velocities of the three vehicles

Table 1. RobuCar Parameters

Symbol	Description	Value
J_i	Inertia moments coefficients on electromechanical system	$3\ kgm^2$
f_i	viscous friction coefficients on electromechanical system	$0.02\ Nms$
R	Wheel ratio.	$0.35m$
D	Half-width of the tire / road contact area	$0.1m$
a	Distance between the centre of gravity and the front axle of the vehicle	$0.4m$
b	Distance between the centre of gravity and the rear axle of the vehicle	$0.8m$
d	Distance between the two front wheels (or both rear wheels)	$1m$
M	Total mass of the vehicle	$350kg$
Iz	Moment of inertia of the vehicle around the yaw axis	$82kgm^2$
f_x f_y f_r	Tire/road friction coefficients	$f_x=0.19kgms^{-1}$ $f_y=0.01kgms^{-1}$ $f_r=0.01kgms^{-1}$

6 Conclusion

The purpose of this paper, which was to create a train of vehicles, in which the leader was given a trajectory whereas the rest of the vehicles behind it would mimic the behaviour of the vehicle in front, has been achieved, as is evident from the simulation results. Also, the constant spacing as well as the same vehicle speeds is maintained throughout the journey.

Acknowledgments. This research was supported by the research department of the Region Nord Pas de Calais and the Centre National de Recherche Scientifique (CNRS) in France which are gratefully acknowledged.

References

1. Iannou, P.A., Chien, C.C.: Autonomous Intelligent Cruise Control. IEEE Transactions on Vehicular Technology 42(4) (1993)
2. Martinez, J.-J., Canudas-de-Wit, C.: A Safe Longitudinal Control for Adaptive Cruise Control and Stop-and-Go Scenarios. IEEE Transactions on Control Systems Technology 15(2), 246–258 (2007)
3. Martinez-Molina, J.J., Canudas de Wit, C.: A Warning Collision System based on an Interdistance Reference Model. In: 16th IFAC World Congress, Prague (2005)
4. Swaroop, D., Karl Hedrick, J., Choi, S.B.: Direct Adaptive Longitudinal Control of Vehicle Platoons. IEEE Transactions on Vehicular Technology 50(1) (2001)
5. Bom, J., Thuilot, B., Marmoiton, F., Martinet, P.: A Global Control Strategy for Urban Vehicles Platooning relying on Nonlinear Decoupling Laws. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2875–2880 (2005)
6. Zhang, Y., Kosmatopoulos, E.B., Ioannou, P.A., Fellow, IEEE, Chien, C.C.: Autonomous Intelligent Cruise Control Using Front and Back Information for Tight Vehicle Following Maneuvers. IEEE Transactions on Vehicular Technology 48(1) (January 1999)
7. Daviet, P., Abdou, S., Parent, M.: Platooning for Vehicles and Automatic Parking by Scheduling Robotic Actions. In: International Symposium on Robotics and Manufacturing, Montpellier, France (1996)
8. Levant, A.: Sliding order and sliding accuracy in sliding mode control. International Journal of Control 58(6), 1247–1263 (1993)
9. Emelyanov, et al.: Second order sliding modes in controlling uncertain systems. Soviet Journal of Computer and System Science 24(4), 63–68 (1986)
10. Levantovsky, L.V.: Second order sliding algorirhms: their realization in dynamics of heterogeneous systems, Moscow. Institute for System Studies, pp. 32–43 (1985)

Organization and Operation of Electronically Coupled Truck Platoons on German Motorways

Ralph Kunze, Richard Ramakers, Klaus Henning, and Sabina Jeschke

RWTH Aachen University,
Center for Learning and Knowledge Management and
Department of Information Management in Mechanical Engineering (ZLW/IMA),
Dennewartstr. 27, 52068 Aachen, Germany

Abstract. One opportunity to manage the increase of freight transportation and to optimize utilization of motorway capacities is the concept of truck platoons. With the aid of Advanced Driver Assistance Systems, trucks are electronically coupled keeping very short gaps (approx. 10 meters) to form truck platoons on motorways. This contributes to optimisation of traffic flow and reduction of fuel consumption advantaged by slipstream driving. In this paper, a brief introduction into these truck platoons is given as well as a short overview about the elements of the automation-, information- and automotive-technology of the experimental trucks. The paper focuses on the Driver Information System which helps truck drivers to organize and operate these platoons. A generic software architecture for the Driver Information System of the platoon system is presented, which guarantees the development of a modern, flexible, extensible and easily configurable system, especially for Human Machine Interfaces of Advanced Driver Assistance Systems.

Keywords: Truck Platoons, Advanced Driver Assistance System, Driver Information System, Software Architecture, Electronically Coupled, HMI.

1 Introduction

The continuing traffic increase in Europe during the last years poses a huge challenge, especially for transit countries such as Germany. Due to the increase of freight transportation, the maximum road capacity in several countries worldwide is nearly reached.

In some countries in Asia and the Pacific, the road density increased up to 100% between 1990 and 2008. More than half of the Asian and Pacific countries had to face an increase of over 20% in that period of time [1]. In Europe a growth rate of 35% in road freight transport was detected between 1995 and 2004. Furthermore, an increase of over 55% in road transportation is expected between the years 2000 and 2020 [2]. In the year 2003, the European Commission stated that every day 7.500 kilometers of the European road system are being blocked by traffic jams [3].

Additionally, the integration of the new European member countries in combination with German's centricity within Europe provides another challenging component for Germany's national traffic planning. Environmental pollution, safety risks and a

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 135–146, 2009.

loss in efficiency for the economy are only some of the effects that result from these factors. Similar problems are known and discussed worldwide.

1.1 The Approach

One possibility to meet the rising transport volume on roads is the modal shift to other types of transportation (e.g. rail, shipping). Another opportunity lies in the optimisation of the road-side traffic flow by driving assistance systems. Since the 90s, Advanced Driver Assistance Systems (ADAS) for trucks have been on offer, including pre-adjustment of speed and distance to the front vehicle. The combination of an Adaptive Cruise Control (ACC) together with an Automatic Guidance (AG) leads to autonomous driving. Here, a precondition is a computerised engine- and brakes-management in connection with an automated transmission.

The difference between platooning and autonomous driving makes the necessity of a leading vehicle. Following trucks can go far distances without any manual engagement by the driver as long as another ahead-driving vehicle exists. Nevertheless, each truck must be assigned with a truck driver at all times due to legal rules and regulations. Within platoons, smaller distances between the vehicles (up to 10 meters) can be realized. These truck platoons contribute to an optimization of traffic flow of up to 9% and a reduction of fuel consumption of up to 10% due to slipstream driving [4].

1.2 Objective

The development of a generic software architecture for a Driver Information System (DIS) for the organization and operation of truck platoons is the objective of this paper. This paper focuses on the DIS of the platoons system and how a generic software architecture of a DIS as a human machine interface (HMI) for ADAS should be designed. Consequently, the system architecture of the ADAS will be not the subject of this paper. A detailed overview of the system architecture can be found in Henning et al. [5].

2 The Scenario “Driver Organized Truck Platoons”

The Project KONVOI is based on the scenario “Driver Organized Truck Platoons” (Figure 1) which was developed in the Project “Operation-Scenarios for Advanced Driving Assistance Systems in Freight Transportation and Their Validation” (EFAS) [6]. The development and evaluation of the practical use of truck platoons is the objective of the project KONVOI, which was funded by German’s Federal Ministry of Economics and Technology. The Project KONVOI is an interdisciplinary research project with partners of the RWTH Aachen University, industry and public institutions, which ended after a duration of 49 months at 09/05/31. With the assistance of virtual and practical driving tests, by using experimental vehicles and a truck driving simulator, the consequences and effects on the human, the organization and the technology have been analyzed [5].

In the scenario “Driver Organized Truck Platoons”, the platoons can operate on today’s existing motorways without extending the infrastructure and the driver has the

permanent control of the autonomous driving procedures [6]. The creation of a platoon depends on the initiating driver who delivers the necessary data about time and place of meeting, the destination, as well as the required truck telemetric data (loading weight, engine power etc.) with the help of a Driver Information System. Because no schedules have to be generated like they have to be in rail traffic, the high flexibility of truck transportation is not lost. After activating the ADAS, it automatically shows a selection of the best matching platoons, informs the driver and prepares the participation to the selected platoon. The DIS acts as human machine interface of the platoon system and helps the truck driver to plan the route and guides the driver to the meeting point [7].

The driver has to initialize and respectively confirm all of the platoon maneuvers in order to build and to dissolve the platoon. As soon as the final position in the platoon is reached, an automated longitudinal guidance with a target distance of 10 meters between the trucks and a lateral guidance is possible. On one hand, this target distance was chosen because the short distance prevents most drivers from driving between the platoons. On the other hand, the short distance causes slipstream effects, which can lead to a reduced fuel consumption of about 10%.

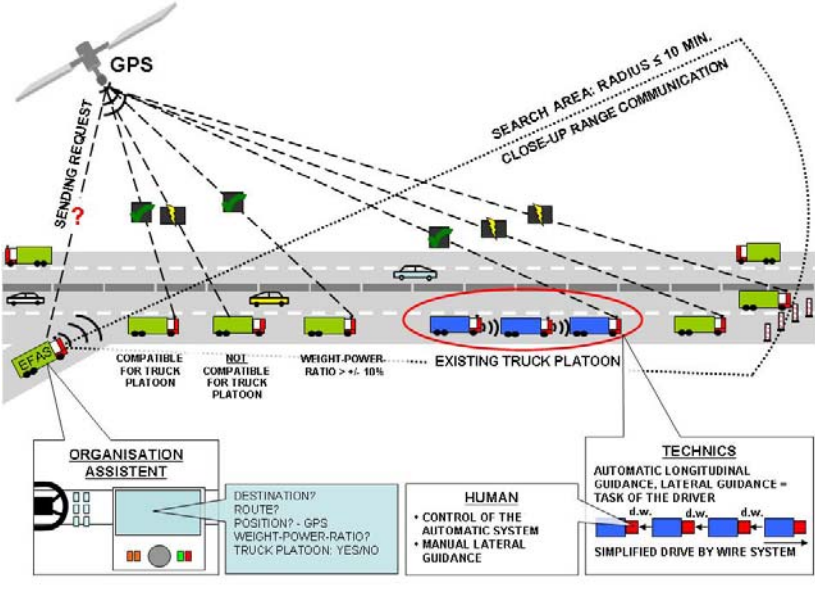


Fig. 1. Scenario 1 – "Driver Organized Truck Platoons" [6]

Since road markings are needed for the lateral guidance, the platoon system is exclusively developed for the use on motorways. Because of a limitation for most trucks at approx. 50 mph, the speed of the trucks on motorways differs only slightly. Therefore, the truck platoons are operated at a speed between 37 and 50 mph. This speed can be managed safely at 10 meters distance by the KONVOI-System.

3 The Platoon System

In order to realize different platoon sizes, four experimental vehicles have been equipped with the required automation-, information- and automotive-technology (Figure 2). The main components for the implementation of the system architecture in the experimental vehicles are the actuators (steering and power train), the sensors (object registration in close-up and far range, recognition of lane), the vehicle-vehicle-communication (WLAN), the automation unit (coordination of the different vehicle states), the control unit (adaptive cruise control and automatic guidance) and the driver information system (human-machine interface, organization assistant, GPS and 3G) [5]. The longitudinal guidance of the ADAS is based on a LIDAR distance sensor, a CMOS-Camera and a RADAR-sensor. The distance sensors are used to determine the distance in longitudinal direction and the lateral offset to the leading vehicle. The vehicle-vehicle-communication transfers necessary vehicle data from all platoon members, which are required for the ACC to realize the target distance of 10 meters. In all trucks, a target acceleration interface is implemented, which automatically calculates the drive-train and the management of the different brakes in the vehicles. The acceleration is either calculated autonomously for each vehicle or deduced from the data which is transferred via the vehicle-to-vehicle-communication.

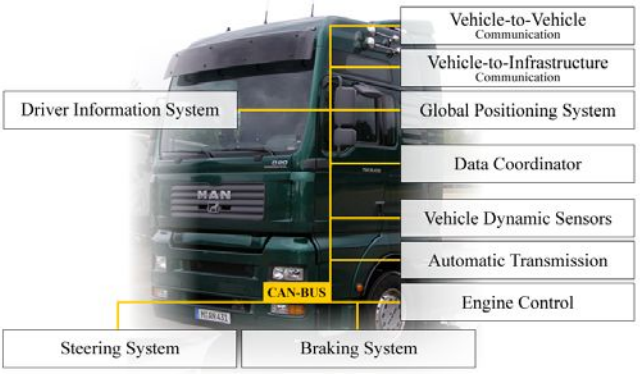


Fig. 2. Automation-, Information- and Automotive-Technology of an Experimental Vehicle

Every experimental vehicle is attached with cameras which are able to identify the traffic lane, thus determining the position of every truck within the traffic lane. An electronically accessible steering system has been integrated additionally. A steering actuator on the base of an electric motor delivers the necessary steering moment for the automated guidance of the trucks [7].

With the help of the Driver Information System, the truck driver plans his route, selects economic platoon participants as well as initialises and respectively confirms the platoon manoeuvres in order to build and to dissolve the platoon.

The platoon organisation is realised on a central server with a data-mining-algorithm under consideration of economic aspects [8]. For this task, the DIS has to send the time schedule, route plan and GPS position of the truck with a vehicle-infrastructure-communication via G3 to the central server (Figure 3).

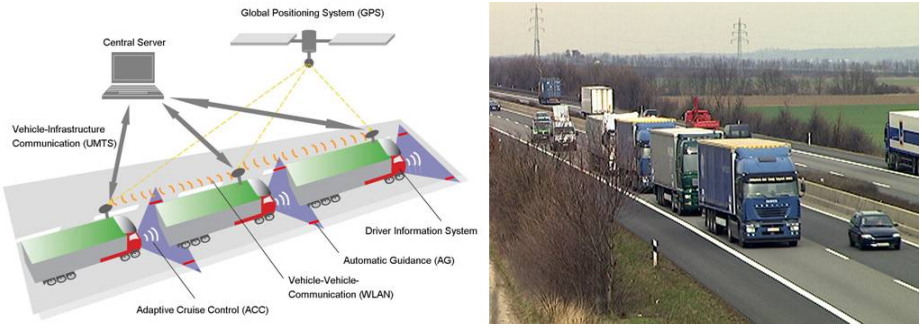


Fig. 3. The Platoon System [9] and Test Run on German Motorways (March 2009)

4 Driver Information System (DIS)

The DIS is not only the HMI of the platoon system, it also acts as information manager between the truck driver and the ADAS as well as the central server. Via touch screen, the user input of the truck driver is processed by the DIS, and in dependency of the user input the data is given to the technical systems. Vice versa, the DIS processes the data of the technical systems and informs the driver in all platoon phases about the current platoon situation (Figure 4).

During the platoon organization, the driver has to enter the DIS settings, planning his route and time schedule and choosing a suitable platoon from an offer list. This data is preprocessed by the DIS and sent to the central server via 3G. The most important data for the central server is the route information of each truck as well as the platoons chosen by the driver. Additionally, the DIS gets a list with suitable platoons sorted by economic criteria from the central server. Furthermore, the central server immediately informs the DIS about any alterations within the planned platoons.

Two further tasks of the DIS are the navigation to the planned meeting points/destinations and the warning of danger areas such as road constructions, bridges, motorway junctions and tunnels. The platoons have to be dissolved manually by the truck driver ahead of these areas.

During the platoon drive, the drivers have to initiate and respectively confirm all platoon maneuvers (connecting, dissolving and lane change). The control signals from the driver to the ADAS are sent through CAN-Bus. The DIS permanently informs the driver about the actual state of the platoon. This applies to the manual as well as to the automated platoon drive.

4.1 Requirements Specification for the Software Architecture

In the following, the requirements of the DIS software architecture are indicated with the letter “R” (for requirement) and a consecutive number.

Modularity, Extensibility, Flexibility and Configurability (R1)
The usual demands made to software architectures for an HMI are modularity, extensibility, flexibility and configurability. In this context, modularity means that certain

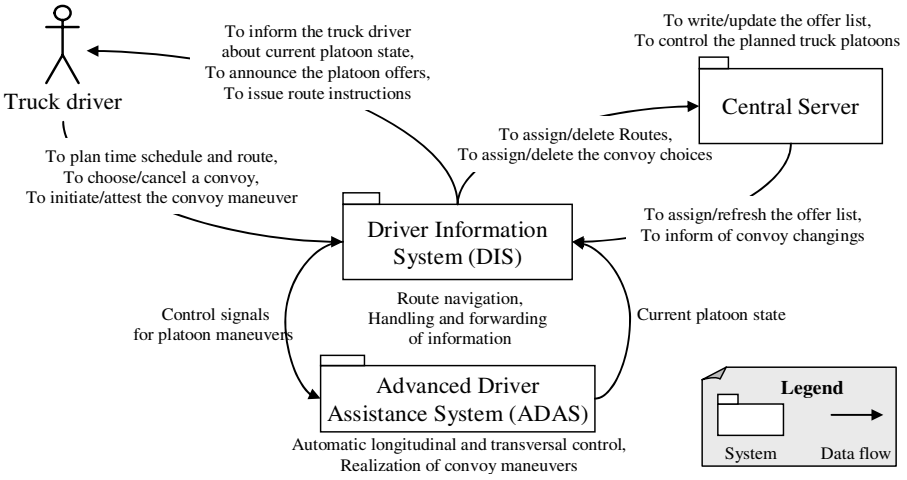


Fig. 4. The DIS as the Information Manager [9]

functionalities are combined in well-defined software components. These components have to be self-explanatory and exchangeable. The extensibility of the software architecture has to be as flexible as possible in order to allow additional functions subsequently and easily. The whole software system – particularly the graphical design of the HMI – is fast to configure and change with a configuration file to avoid alterations of the source code.

Robustness and Reliability (R2)
The automobile sector has especially high demands on technical systems in vehicles concerning robustness and reliability [10]. Therefore, the software architecture has to ensure the robustness and reliability through adequate safety mechanisms and functions.

Information and Data Management (R3)
The DIS as an information manager has to handle and process a large quantity of data as well as the transmitting of processed data to the corresponding technical systems. Therefore, the system architecture has to support an internal communication, in order to support the different software components with the required data. Furthermore, the software architecture has to support different functions to manage the data in the internal memory as well as in data bases.

External Communication with System Environment (R4)
The DIS is capable to communicate with their system environment. Accordingly, an external communication with appropriate communication interfaces has been implemented. Furthermore, the ease extensibility (also for other technologies e. g. WLAN, Flexray) has to be guaranteed by the software architecture (cf. requirement R1).

Interaction with the User (R5)
The DIS is the HMI of the platoon system. The relevant input is made by the driver on the user interface (e.g. touch screens). The system architecture handles the user input through the user interface and gives the effects of the users' manipulation back to the user interface so that the user can assess the system state.

Table 1. Requirements for the Software Architecture

R1	Modularity, Extensibility, Flexibility and Configurability
R2	Robustness and Reliability
R3	Information and Data Management
R4	External Communication with System Environment
R5	Interaction with the User

4.2 Design of the Software Architecture

The software architecture comprises software components, the externally visible properties of those components and the relationships between them [11]. The choice of an architecture pattern, as the base for the software architecture, is a fundamental one and also one of the first decisions to make [12]. The decision for an architecture pattern was made on the basis of the defined requirements (chapter 4.1).

The claimed interaction with the user (R5) classifies the DIS as an Interactive System. The DIS has to process all the incoming data (R3) and has to exchange this data with the system environment, thereby relying on the external communication interfaces (R4).

Compared to the required aspects of modularity, extensibility, flexibility and configurability (R1), the DIS can be regarded as an Adaptive System, which must run robust and reliable in the car (R2).

Consequently, the defined requirements can be summarized as the claim for a modular, extensible, flexible and configurable HMI, whereby especially the aspect of extensibility accounts for the modularity and flexibility of the software architecture. The denotation of the DIS as a HMI makes the importance of a structured human-machine interaction obvious, so finally the Model-View-Controller architecture was chosen for the development of the DIS as an Interactive System [12].

Model-View-Controller (MVC) architecture
The component concept of the DIS is build on the basis of the Model-View-Controller (MVC) pattern introduced before. In some cases, the functionality of the software components is slightly different from the ones in the MVC literature. The architecture – following the MVC architecture pattern – is based on three core components, a Controller-, a View- and a Model-Component [15].

The Model-component serves as a collection of abstract data structures which administrates the data. Unlike the MVC pattern, where the Model-component is responsible for manipulating the dates, the Model-component in the adapted MVC pattern has solely administrative tasks. These tasks are primarily the access control and thread backup. The other components (Controller and View) are unknown to the Model-component.

Along the lines of the general MVC pattern, the View-component displays the information provided by the Model-component. Data manipulation or other alterations by user interaction are communicated to the View-component by the Controller-component. The View-component knows the Model- as well as the Controller-component in order to readout data about user interaction. Simultaneously, it administrates the various GUI-components. These comprise graphical elements, which visualize selected data from the Model-component for the user.

The logic and functionality of the software is located in the Controller-component. This component evaluates user interactions and manipulates data. In addition, the Controller informs the View-component about data manipulation or special requests of the user leading to a new or modified notation of a GUI-component (e.g. change of notation, deactivation of functionality). The Controller-component also administrates Logic- and Communication-components. The Controller extends the program logic on the one hand, and on the other hand enables the possibility of external communication (e.g. CAN, 3G).

4.3 Implementation of the Software-Architecture

Figure 5 shows a simplified UML class diagram of the DIS software architecture. The components Model, View, Controller, Logic, GUI, Comm and ModelData are realized through several classes. To avoid complexity, a well arranged overview is given in the figure, by excluding the attributes, methods and multiplicities of the classes.

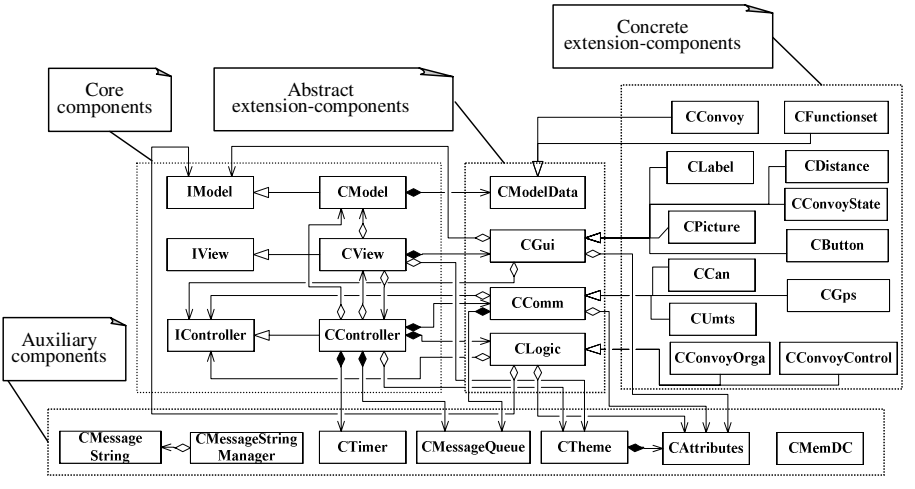


Fig. 5. UML class diagram of the DIS software architecture [9]

Basically, the software architecture is divided into core components, abstract extension-components, concrete extension-components and auxiliary components. This concept enables - next to the chosen software architecture - a high flexibility concerning the extensibility with program logics, data models and design elements. In addition, dynamic libraries are used to uncouple the core source code of the concrete extension-components from the main program of the DIS. Also, a static library for the DIS is used, where all

necessary definitions (e.g. classes, auxiliary classes) for the implementation of the software system are included. In the following, the realization of the different software components will be explained.

Core Components
The core components inherit from interface definitions (IModel, IView and IController), so that some functionalities are concealed to the extension- and auxiliary components and access is only allowed to designated functionalities. The overall functionality (according to the visibility of the methods of the classes) is only known to the core components among themselves and can only be used by them.

The internal communication between the Controller- and the View-components are done with an interprocess communication. Asynchrony, messages are saved into message queues until the recipient retrieves them. Both core components have their own message queue.

The DIS has to handle different threads during the runtime. In such a multithread-application, thread safety is very important. A piece of software code is thread-safe if it functions correctly during simultaneous execution by multiple threads [15].

In the software architecture, the message queues as well as the model-components are thread-safe implemented. For this purpose the synchronization mechanism CSingleLock and CMultiLock from the Microsoft Foundation Classes (MFC) of Microsoft was used. A lock is used to ensure that only one resource respectively one critical section in a software component can be used by a thread. The other threads have to wait – due to the closed locks – until the critical sections are opened for the next thread.

Abstract Extension-Components
An abstract extension-component is an abstract class, i.e. a class not completely implemented concerning the method definitions. In an abstract class some methods are defined, other methods – so called "pure virtual functions" – are not declared. More precisely this means that from an abstract class no object can be derived. In the DIS software architecture the Model-component is extended with ModelData-components, to add, remove or read out data. The View is extended with the abstract class CGui. The Controller is extended with two abstract extension-components: The class CComm for the external communication (e.g. UMTS, CAN) and the class CLogic for the core functionality of the software system. The class CComm also has a message queue, which was described in the previous section. All abstract classes have to be extended with concrete classes.

Concrete Extension-Components
Concrete extension-components are concrete classes which extend the software system through logic functionalities (Logic-components: e.g. the functionality to organize and operate truck platoons), data (ModelData-components: e.g. to manage the convoy state) and views (GUI-components: the different elements for the graphical user interface, for instance pictures, buttons, labels etc.). To reach this goal, the concrete extension-components inherit from the abstract extensions components and complete the non-defined methods of the abstract classes. The abstract extension-components provide a quantity of code for the internal communication and processing within the software

system, so that during the development of a concrete extension component, the relevant part of the component can be focused. Moreover, this procedure secures that every concrete extension component makes the required interfaces available. Most of the concrete extension-components are implemented in Dynamic Linked Libraries.

Auxiliary Components
The DIS software architecture provides a set of auxiliary components in a static library. Static libraries, unlike dynamic ones, are not linked with the program during runtime of the application, but already during compilation. As shown in figure 3, the static library includes the interfaces of the core components (IModel, IView, IController), the abstract extension components (CLogic, CComm, CGui, CModelData) and some auxiliary classes, for instance to support Double-Buffering (CMemDC, CDC), thread safety (CMessageQueue), manipulating text strings (CString) and timer functions (CTimer).

Configurability of the Software System with XML-Files
The design of the software architecture intends the configuration of the software system with a XML configuration file. In this XML-file the configuration of the extension-components, the behavior of the software systems by user interaction, the functionality of the core components and the design of the graphical user interface is specified.

4.4 Trial Implementation

The V-Model has been subject of the development process of the KONVOI-System. The KONVOI-System had been developed according to final specification. Internal and external reviews have been implemented during the development to evaluate parts of the KONVOI-System, such as the software-architecture of the DIS. In the very beginning of the development process a preliminary hazard analysis had been taken place. A variety of systems test were carried out during the project to prove the robustness and reliability of the system. Additionally thread safety had been implemented. Finally, the KONVOI-system had to meet the IEC 61508 standard. The IEC 61508 “Functional safety of electrical/electronic/programmable electronic safety-related systems” is an international standard, which include a Failure Mode and Effects Analysis (FMEA) and a Fault Tree Analysis (FTA).

The development of the KONVOI-System is accompanied by a trial implementation in four steps. First, the whole platoon system and all of its components are completely implemented into a truck driving simulator at the RWTH Aachen University. An integrative part of the simulator is a software- (sil) and hardware-in-the-loop (hil) development environment. All test procedures (unit, integration, system and acceptance tests) as well as the optimization of the different system components were carried out in the simulator. This helps to increase the speed and quality assurance during the development. In addition to the technical development and testing the second step includes another work package: the examination of both the acceptance of the ADAS and the arising stress of the truck drivers. In a third step, the trial implementation on test tracks has taken place and the trial implementation on motorways has taken place in the second half of the project. After this, an evaluation phase on motorways in real

traffic with all four experimental trucks followed to confirm the effects determined by simulation and test drives.

For the system tests the DIS was first implemented into a truck driving simulator which was used as a test environment for the module, integration and system tests. The tests with 30 truck drivers from the freight forwarding companies of the project consortium showed a consistently positive assessment and proved the functionality as well as the reliability of the software system. After a sufficient testing of the ADAS in the driving simulator, the platoon system was tested with experimental vehicles on test tracks. Already thousands of miles were driven with the platoon system on test tracks during the trial runs. The development team worked for more than 60 days on test tracks and proved successfully that the platoon system runs absolutely safe and free from errors.

After the trial runs on test tracks were successfully finished in December 2008 the test runs on German motorways started in March 2009 (Figure 3). The test runs were completely realized on public motorways to measure the effects of the platoon system on the traffic flow, the economic efficiency and the acceptance of the truck drivers.

5 Conclusion

In this paper, an introduction into electronically coupled truck platoons with Advanced Driver Assistance Systems was given. Furthermore, the requirements for the software architecture of the Driver Information System as the HMI and information manager of the platoon system were derived and transferred into a software design. Finally, the implementation of the architecture was described in detail.

The presented software architecture fulfills all the fundamental demands for the development of interactive software systems in the automotive sector. Hereby, the architecture ensures the user interaction between the driver and the technical systems as well as the data processing between the different system components in the vehicle. This architecture guarantees a modern, flexible, extensible and easily configurable system, especially for HMI of driver information and assistance systems. Due to its interactive and adaptive characteristics, the presented architecture could be moreover seen as a generic software architecture framework for Driver Information System of platoon systems.

References

1. United Nations, Economic and Social Commission for Asia and the Pacific: Statistical Yearbook for Asia and the Pacific. United Nations Publication, Bangkok (2008)
2. Commission of the European Communities: Keep Europe moving – Sustainable mobility for our continent. Brussels (2006)
3. Commision of the European Communities: Europe at a crossroad - The need for sustainable transport. Manuscript of the European commission, Brussels (2003)
4. Savelsberg, E. (ed.): Lastenheft für elektronisch gekoppelte Lkw-Konvois, vol. 22(21). VDI, Reihe, Düsseldorf (2005)
5. Henning, K., Wallentowitz, H., Abel, D.: Das Lkw-Konvoisystem aus den Perspektiven Informations-, Fahrzeug- und Automatisierungstechnik. In: Mechatronik 2007 - Innovative Produktentwicklung. Hrsg. v. VDI/VDE-Gesellschaft Mess- und Automatisierungstechnik, pp. 133–147. VDI, Düsseldorf (2007)

6. Henning, K., Preuschoff, E. (eds.): Einsatzszenarien für Fahrerassistenzsysteme im Strassengueterverkehr und deren Bewertung, vol. 12(531). VDI, Reihe, Düsseldorf (2003)
7. Friedrichs, A., Meisen, P.: A Generic Software Architecture for a Driver Information System to Organize and Operate Truck Platoons. In: Vortragsveroeffentlichung, International Conference on Heavy Vehicles (HHVT 2008), Paris, May 19-22 (2008)
8. Meisen, P., Henning, K., Seidl, T.: A Data-Mining Technique for the Planning and Organization of Truck Platoons. In: Proceedings of the International Conference on Heavy Vehicles (2008)
9. Friedrichs, A.: A Driver Information System for Truck Platoons, vol. 12(673). VDI, Reihe, Düsseldorf (2008)
10. Wietzke, J., Tran, M.: Automotive Embedded Systems. Springer, Berlin (2005)
11. Balzert, H.: Grundlagen der Informatik. Spektrum (2005)
12. Buschmann, F., Meunier, R., Rohnert, H.: Pattern-Oriented Software Architecture, A System of Patterns. John Wiley, Chichester (1998)
13. Posch, T., Birken, K., Gerdom, M.: Basiswissen Softwarearchitektur. dPunkt, Heidelberg (2007)
14. Reenskaug, T.: Thing-Model-View-Editor, An example from a planning system. In: Xerox PARC technical note (May 1979)
15. Budszuhn, F.: Visual C++-Programmierung mit den MFC. Addison-Wesley, Reading (2002)

Navigation Method Selector for an Autonomous Explorer Rover with a Markov Decision Process

Simon Le Gloannec and Abdel-Illah Mouaddib

GREYC, UMR 6072
Université de Caen Basse Normandie
14000 Caen, France

Abstract. Many high level navigation methods exists for autonomous rovers. Usually, fixed methods is choosen for the rover to follow its trajectory. In some cases, a rover can have several embedded navigation methods, but there is currently no automatic approach to select the best method. In this paper, we propose an automatic navigation method selector using Markov Decision Process (MDP) framework. Navigation method are modelized in the MDP transitions functions. Results we achieved were promissing.

1 Introduction

In the robotic application fields, robots often have several actuators and sensors in order to interact with the environment (cameras, telemeters, GPS, wheels, arms...). Some high-level methods are then implemented in the robot's kernel to produce complex behaviours. For example, some robots recognize objects (methods) with a camera (sensor) in order to follow or grab them (behaviour). Many methods can also produce the same behaviour with different sensors.

We consider here the problem[1] of a robot that have to follow a given trajectory. The trajectory is composed of points that the rover must approach as best as possible. There exist many methods to go from one point A to another point B. A robot can move forward by using only its odometer, or by following a wall detected with a telemeter. These methods produce more or less good results but they are more or less difficult to get start. Following a wall by using the telemeter has a cost because at the same time, the robot can not use it to perform an other task. The main problem in this context is to **select the appropriate method** to perform the desired behaviour, i.e. follow the trajectory. In our case the methods are : move towards an object, move along a wall and move without help except the robot odometer. Some objects in the environment can then help the robots : walls, trees, houses, other vehicules. Methods are implemented to use these objects in order to follow this path. We do not present these methods in details but we are concerned with the module that permits the robot to select the best method at the right time.

[1] This work is supported by DGA (Délégation Générale pour l'Armement) and THALES company.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 147–156, 2009.

Markov Decision Processes (MDP) [Bel57, Put94] are used to modelize this selection problem [LMC08],[MZ00]. Even if most of the time the policy computation is "*only*" quadratic in the problem state space, the state space is generally huge. Moreover, if a new useful object appears in the environment, the policy must be recomputed online. We tackle this two problems in this paper with an acyclic directed graph in the MDP.

The first part of this article shows the autonomous rover context. Then we modelize the problem into an MDP. We describe the algorithm that produces the policy. In the same section, we show how to recompute a part of this policy when a new object appears in the environment. Finally we present some experimental results

2 The Target Application

This work is developped in a project in collaboration with THALES company under grant of DGA. The project consists of an autonomous rover has to follow a given trajectory. This path is defined as a set of points (see Fig. 1). The objective for the rover is to pass throw these points or as near as possible. It has three high-level methods to move to the next point. The first method is to use an odometer that is quite imprecise. The second method is to recognize a particular object in front of the rover, it will then move toward this object during some time [CBC+05], [SRC06], [BM07]. The third method is to detect a wall on the side. The rover has to move along the wall.

The three methods end up with in the same behaviour which is to move toward a particular point. The first method is less precise than the others but the environment does not always provide good objects to use. Between some points, the method to apply is fixed. For example, during the first part of the trajectory depicted in Fig. 1, the rover must move along the wall. In some region, no methods are selected (third part in Fig. 1). We propose in this case an automatic navigation method selector, We formalize this problem with an MDP. In addition, the rectangles offers a simple strucutre to perform easily computation since it allows us to transform the space as a grid. If the rover is far from it's path, we consider it is in a failed state. We define a rectangle around each pair of points. Beyond these rectangles, the rover is in a failed state.

We also suppose that the rover never turn back. This assumption is important because it permits us to have an acyclic structure problem. We will show in the next section how to use this acyclic property to solve the problem more efficiently.

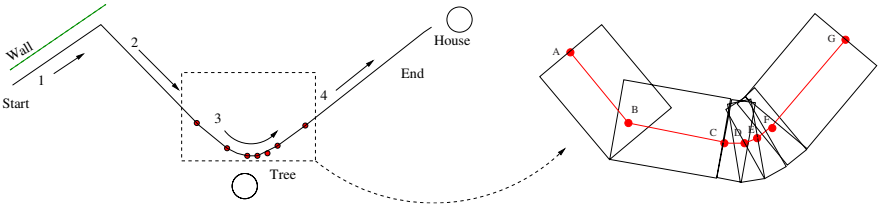


Fig. 1. The rover must follow a trajectory and select navigation methods

3 Formal Framework

The problem is formalized in a Markov Decision Process MDP model. We calculate a policy with this model that indicates the action to select given a particular state.

3.1 Background : MDP Model

An MDP is a tuple S, A, T, C where

- S is a finite set of states,
- A is a finite set of actions,
- T is a transition function. It gives the probability of being in the next state given the selected action and the current state,
- C is a cost function. This cost includes many features like distance and changing action.

Actually a state is the rover position combined with its orientation. $S = \mathcal{R}^2 \times]-\pi, \pi]$. But not all states are accessible according to the previous assumption. We transform the state space into an acyclic graph that is much easier to solve with an MDP. As shown in Fig. 1 (right side), we transform every single segment of the path into a rectangular shape area. States that are outside these boxes are failed states. All the failed states are considered as a unique failed state.

At this point the transformation is not finished. We combine three geometrical tranformations (rotation, scale, translation) in order to obtain a simple grid for each rectangular shape. The grid ratio Δ transforms a $l \times L$ rectangle into a $M \times N$ matrix where $M = L/\Delta$ and $N = l/\Delta$. An α rotation is needed to straighten up the grid. Finally, the left lower corner is moved to the origin (translation). We do this transformation for each rectangle.

The graph is not yet acyclic. To reach this, we only keep angles that are not moving backward. We select N_ω angles (for example five in Fig. 2) appropriate to our target application. The state space is now composed by n $M \times N \times N_\omega$ matrix. The state space is acyclic : while moving forward, the rover cannot not enter the same state twice.

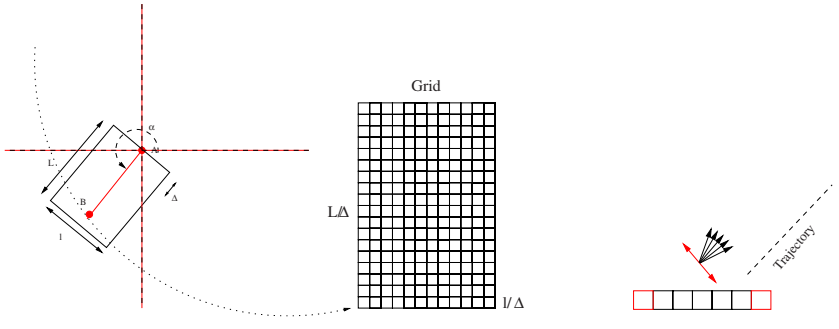


Fig. 2. Transformation from a continuous state space into a 3D discrete $M \times N \times N_\omega$ matrix

There are 3 kinds of actions

1. a_m consists of moving along the path without help (except an odometer),
2. a_{o_i} consists of moving towards the object i,
3. a_{w_j} consists of moving along the wall j.

In each rectangle/grid, the rover can only follow one of two possible objects or one of two walls. Thus, the rover can perform 5 actions in a given grid $a_m, a_{o_1}, a_{o_2}, a_{w_1}, a_{w_2}$, where o_1, o_2, w_1, w_2 are respectively object 1, object 2, wall 1 and wall 2. The transition function represent the uncertainty model the uncertainty reliable to the rover movement method. When the rover helps itself with an object or a wall we consider the transition as perfect (i.e. deterministic)

$$T(\langle x, y, \omega\rangle, a_o, \langle x + v.\cos\omega, y + v.\sin\omega, \omega\rangle) = 1. \tag{1}$$

where v represent the rover speed.

Without helping, the Transition is uncertain. The rover can deviate from $\delta\omega$

$$T(\langle x, y, \omega\rangle, a_m, \langle x + v.cos(\omega), y + v.sin(\omega), \omega\rangle) = 1 - 2p \tag{2}$$

$$T(\langle x, y, \omega\rangle, a_m, \langle x + v.cos(\omega + \delta\omega), y + v.sin(\omega + \delta\omega), \omega + \delta\omega\rangle) = p \tag{3}$$

$$T(\langle x, y, \omega\rangle, a_m, \langle x + v.cos(\omega - \delta\omega), y + v.sin(\omega - \delta\omega), \omega - \delta\omega\rangle) = p \tag{4}$$

where v represents the rover velocity.

The cost function C is a sum of four atomic cost functions:

1. the distance cost $C_d(\langle x, y, \omega\rangle)$ measures the distance between the current state and the ideal path
2. the cost of using an action a $C_a(a)$ occurs each time the rover performs the action.
3. the changing cost $C_c(a, a')$ is paid each times de rover changes its current action
4. the distance $C_i(\langle x, y, \omega, n\rangle)$ that we will discuss later on.

In this context, T and C are sligthly different from the usual definition because we have to take the cost of the action change into account. We will detail all costs in the next section.

To solve the MDP we have to calculate a value function. This Value function V is as usually calculated with the Bellamn Equation. But in our context, we have to take the change cost into account. Thus, in order to memorize the last action, we calculate Q values on pairs of (state, action). The objective here is to minimise the cost.

$$Q(s, a) = \min_{a' \in A} \sum_{s \in S} C(s, a, s', a') + T(s, a, s').Q(s', a') \tag{5}$$

We compute the Cost value to the last line of the last grid. Then, the calculation is done backward from the last grid to the first one. In the last grid, the algorithm computes the cost value to the last line. Then, the entire Q-Value function is calculated backward on the grid according to the previous equation. The policy is finally saved in a file. Then, the algorithm step back to the previous grid until the first grid is reached.

$$\pi^*(s) = argmin_a Q(s, a) \tag{6}$$

Only grid G_{i+1} remains in memory during the calculation of grid G_i's values. All previous ones are freed after being saved in a file. The state space is strutured on ordered grids where each grid is a subpace that can be considered solely for computation. This is an advantage of having an acyclic graph structure. The program does not need a lot of memory, since the memory space is an important issue of the trooper rover from THALES. Some states are sometimes declared two times when many rectangles overlap each other. In this case, the algorithm does consider the latest state, i.e. the one in the grid G_{i+1}.

Once the policy is entirely calculated, the rover only has to read the files in order to select the best action.

3.2 Cost Functions

The cost function is divided into four parts.

1. the distance cost $C_d(\langle x, y, \omega \rangle)$ measures the distance between the current state and the ideal path
2. the cost of using an action a $C_a(a)$ occurs each time the rover performs the action.
3. the changing cost $C_c(a, a')$ is paid each time the rover changes its current action
4. the distance cost $C_i(\langle x, y, \omega, n \rangle)$.

C_d measures the distance cost between the current state and the ideal state. Equation 8 calculates this cost (also see Fig. 3). This measure is divided into two parts. It measures first the euclidian distance from the straight line Δ_i in relation to the rectangle width l. Secondly, it measures the angle deviation from the current ideal orientation α_i. There is no penalty when the angle brings the rover closer to the straight line Δ_i. Thus, equation 8 is applied to the left part of the rectangle only. We introduce a β weighting factor (generally equal to $0,5$) in order to give more importance to the distance or to the deviation.

$$C_d(\langle x, y, \omega \rangle) = \beta . \frac{d((x, y), \Delta_i)}{l/2} + (1 - \beta) . \frac{max(0, \omega - \alpha_i)}{\alpha_{max}} \tag{7}$$

In order to avoid to use the odometer that is quite imprecise, we introduce a using cost for each action. $C_a(a)$ is a constant positive integer. When the rover activates an high-level method like move toward an object, it monopolizes an effector like a camera. It

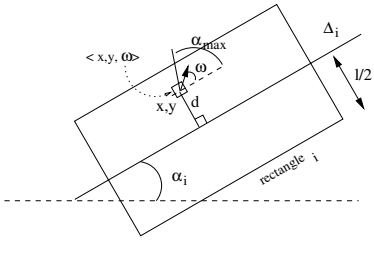


Fig. 3. Distance cost in rectangle i

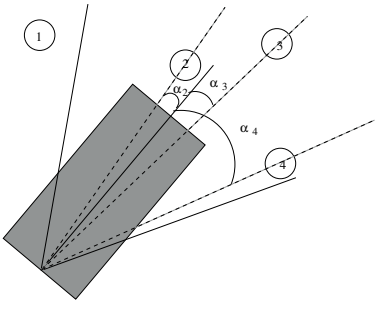


Fig. 4. Objects eligibility

also takes some time to locate the target. To take this into account, we define a change cost $C_c(a, a')$ as a 2 dimensionnal matrix.

Finally all costs are added together to form the global cost. Factors w_d, w_a, w_c (and w_u) weight this sum, they could be adapted during a learning step.

3.3 Object Eligibility

The rover can only move toward an unique object at a given time. But it has many sensors (or camera) ready to capture an object. Thus in each rectangle of the path, we have to select 0, 1 or 2 objects. This selection is made like it is shown in Fig. 4. All objects outside a $\pi/6$ angle are excluded from the selection. Then we keep the two closest objects if there are any. In this Fig., object 2 and 3 will be selected.

The Wall selection is quite similar. We select the closest wall from the left/right side of the rover if it is not too far from the trajectory (e.g. three meters).

3.4 Including Uncertainty in the State Description

While moving with the help of the odometer, the rover has an uncertain position. The longer it moves without assigning a helping object, the more uncertain it will be about its the position. To modelize this, we define N_u levels of uncertainty. Thus we add a variable to the state. It becomes $\langle x, y, \omega, u\rangle \in \mathcal{R} \times \mathcal{R} \times]-\pi, \pi] \times [1, N_u]$. Each uncertainty level corresponds to an elliptical area E that represents where the rover can be. A normalized gaussian curve indicates the position probability ($P(\langle x, y, \omega\rangle) \rightarrow]0, 1]$). In Fig. 5, there are 4 level of uncertainty. At the 4^{th} level, the rover could be in a failed state. Each time the rover moves with the odometer, the uncertainty increases to the next level ($s_1, \ldots, s_4$). If the uncertainty level is equal to the maximum N_u, the level stays at this maximum. Elliptical range and their associated gaussian are predefined (as the number N_u).

With this uncertainty level, we can introduce the uncertainty cost. It is the probability for being in a failed state mulitplied by the failed state cost.

$$C_u(\langle x, y, \omega, u\rangle) = V_{FAIL} \int_{x,y,\omega}^{E\times]-\pi,\pi]} P(\langle x, y, \omega\rangle = FAIL).dx.dy.d\omega \tag{8}$$

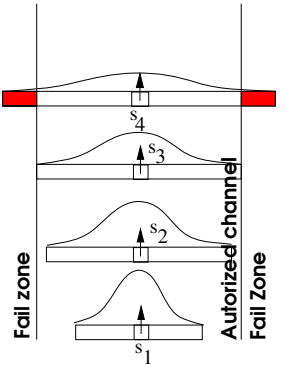


Fig. 5. Uncertainty in the model

Where FAIL is the failed state and V_{FAIL} a constant value. It has an high cost. When the rover uses an high-level method, uncertainty automatically decreases to 1. The object and the walls give a good localization to the rover. This cost is in our opinion more realistic that the cost C_a. But it also increases the algorithm complexity.

3.5 Complexity

The algorithm we presented previously is linear in the state space. The result of this policy is a policy that the rover must follow once it is calculated. Thus the complexity depends on the path total lenght L_T, the channel width l, the grid ratio size Δ, the angle number, the uncertainty levels and the action number (here 5). This leads to the following complexity :

$$L_T * l/\Delta^2 * N_\alpha * N_i * A \tag{9}$$

In order to have a very precise policy, we can increase N_α and N_i or decrease Δ. This model is interesting because it's adaptive. The policy can be calculated very quickly if necessary. Drawback of MDP model is their lack of adaptivity when th environment changes after the policy has been calculated. In this paper, we can take some changes in the environment into account without calculating the policy entirely.

4 Dynamic Environment

During the exploration mission, the path never changes. But sometimes, new objects or walls that the rover had not detect yet could appear and be very helpfull. Generaly, problems modelized with MDP model calculates a definitive policy that indicates the rover the action to select during execution time. But here, the graph induced by the MDP is acyclic. Thus, if something change the policy or the value function in the i^{th} rectangle of the path, only the i^{th} first rectangle of the path are affected. Moreover, if this change appears during the mission, the change affects the rectangle from j to i if the

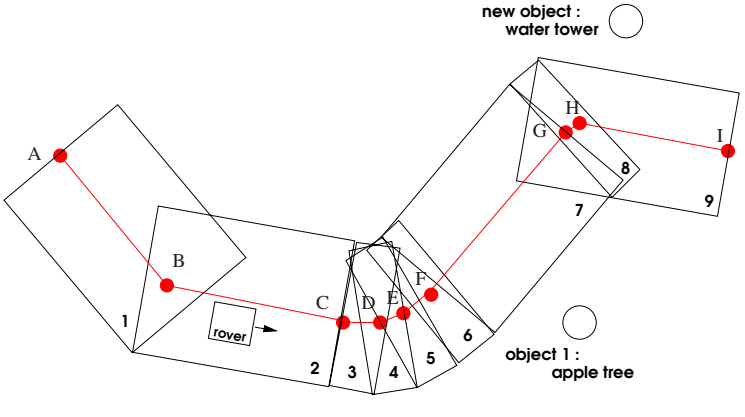


Fig. 6. A new object in the environment does not affect the entire policy

rover is currenctly in rectangle j. Fig. 6 explains this principle. The rover in rectangle 2 has detected a water tower in the distance. This object can not be usefull for now. But it will be for rectangles 4, 5, 6, and 7. Then it has to calculate a new policy from rectangle 7 to rectangle 2.

When a new interesting object is detected, we first search the first rectangle where this object is eligible. If no rectangle is affected, the policy does not change (this object was not so interesting). Otherwise, we load the policy in rectangle i+1 and calculate the new policy from i to j very quickly. Then, the rover selects new actions according to the new policy.

5 Experimental Results

We develop a C++ program that calculates the MDP policy on a given trajectory. We have included a graphical interface that shows the policy and also the Q-Value functions.

5.1 Interface

This is a very helpful tool to simulate the rover's behaviour. With this graphical interface, you can : load a trajectory from a file, see the corresponding channel and rectangle around this trajectory, add a new object or a new wall, calculate the Q-Values and the policy, see the Q-Values and the policy, change the parameters Δ, $N_\alpha \ldots$

A trajectory file is a set of points. After being loaded into the main application, the user can see the corresponding channel inside which the rover has to move (Fig. 7).

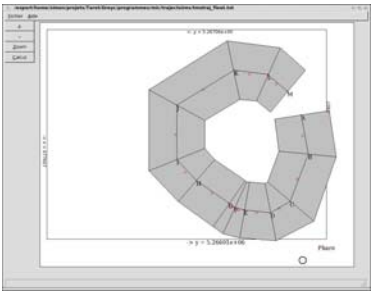


Fig. 7. A trajectory and the corresponding channel

We define a color gradient green, yellow, orange, red that represents values. Green corresponds to 0 while red corresponds to V_{FAIL}. Value functions are painted for different reasons.

1. showing the policy value function, i.e. the value of the best action for each state.
2. it also shows all QValue function values, in order to see the values for a particular function.

The goal is to minimize the cost. Thus the typical result is to have a red on the channel's border and green in the middle. The beginning of the path is more red than the end because the state's values are an accumulation (or sum) of atomic using costs (plus other costs). The value function decreases slowly from the beginning to the end of the path.

The parameters can also be changed directly in the interface. When Δ is very small, there are lots of states. In this case the solution comes slowly. The cost function weighting factors $w_d, \ldots w_u$ can be changed before calculating the policy. It has a great influence on the value function. We tuned this parameters in order to have a "good" value

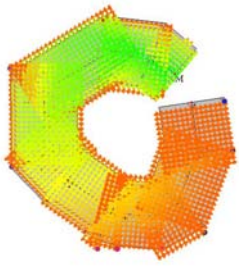


Fig. 8. A value function without object

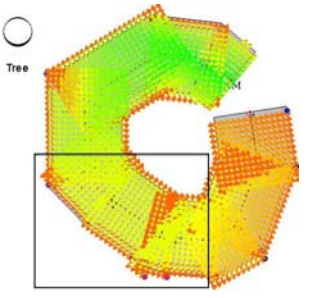


Fig. 9. With an Object

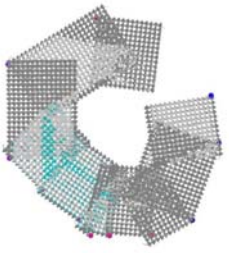


Fig. 10. Policy

function. We do not want to have a completely red value function neither a completely green function but a gradation of colors. Without helping objects , we obtain the value function shown in Fig. 8.

In Fig. 9, the rover follow the tree in the dotted rectangle area. We see that the values are lower in this region than in Fig. 8. The policy of the first case (Fig. 8) is not depicted because there is only one action : trust the odometer.

5.2 Running Time and Memory Requirements

We measure time and memory required for some cases.

Total length L_T (meter)	1672,64	1672,64	3092,43	3092,43
Channel width l (meter)	20	20	20	20
Grid ration Δ	10	1	1	1
Number of angle N_α	21	21	21	5
State space	8.10^5	4.10^6	$7,3.10^6$	$1,7.10^6$
running time (s)	1	62	119	27

We used a 2.60GHz CPU to perform these tests. The function that take the more time is the state's transformation. The memory requirement is not so huge because grids are saved into file one per one during the calculation.

5.3 Behaviour Analysis

We add some objects in order to see whether the behaviour change or not. In this example, we add an object in the rigth up corner. It can help the rover at the beginning of the turn. Effectively, it does. The value function turns into green in the rectangle where this object is elligible. As a result, the corresponding policy also change : the rover move towards this object when it is on the right part of the rectangles. It does not move towards it on the left side because in this case it would come off [or leave] the channel. In

Fig. 10, the rover moves towards an object when the states are blue. Otherwise, it just follows the path without help.

6 Conclusion

In this paper we consider the problem of an autonomous rover that have to follow a path. This is developped in a project in collaboration with THALES company under the grant of DGA. The rover has many methods to direct itself. It can move towards object, follow a wall side by side or use his imprecise odometer. We showed how the rover can select the best method to use in order to follow the trajectory. This selection is based on a Markov Decision Process framework. A policy is calculated off-line. On-line, the rover reads the instructions. MDP framework often suffer from a lack of adaptivity, but we found a method that transforms the problem into an acyclic structure. Then, the policy calculation time is linear in the state space. This allows us to consider dynamic changes during running time : it is now possible to compute a policy with a new environment quickly. Some experiments have been developped to test running time and memory requirement. The behaviour of the rover is interesting : it helps itself with objects in order to follow the path. This automatic selection method is under integration on a six-wheel Trooper rover soon equiped with these behaviours.

References

[Bel57] Bellman, R.E.: Dynamic Programming. Princeton University Press, Princeton (1957)

[BM07] Benhimane, S., Malis, E.: Homography-based 2D Visual Tracking and Servoing. The International Journal of Robotics Research 26(7), 661–676 (2007)

[CBC+05] De Cabrol, A., Bonnin, P., Costis, T., Hugel, V., Blazevic, P.: A new video rate region color segmentation and classification for sony legged robocup application (2005)

[LMC08] Gloannec, S.L., Mouaddib, A.-I., Charpillet, F.: Adaptive multiple resources consumption control for an autonomous rover. In: Bruyninckx, H., Přeučil, L., Kulich, M. (eds.) Proc. European Robotics Symposium 2008 (EUROS 2008), pp. 1–11. Springer, Heidelberg (2008)

[MZ00] Mouaddib, A.-I., Zilberstein, S.: Optimizing resource utilization in planetary rovers. In: Proceedings of the 2nd NASA International Workshop on Planning and Scheduling for Space, pp. 163–168 (2000)

[Put94] Puterman, M.L.: Markov Decision Processes: Discrete Stochastic Dynamic Programming. John Wiley & Sons, Inc., New York (1994)

[SRC06] Segvic, S., Remazeilles, A., Chaumette, F.: Enhancing the point feature tracker by adaptive modelling of the feature support. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3952, pp. 112–124. Springer, Heidelberg (2006)

Discrete-Time Adaptive Sliding Mode Control of Autonomous Underwater Vehicle in the Dive Plane

Baoju Wu[1,2], Shuo Li[1,2], and Xiaohui Wang[1]

[1] Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China
[2] Graduate School of the Chinese Academy of Sciences, Beijing 100049, China

Abstract. This paper presents a discrete-time adaptive sliding mode controller for an autonomous underwater vehicle (AUV) in the presence of parameter uncertainties and external disturbance. The controller makes the system stable in the presence of system uncertainties and external disturbances. The proposed algorithm has a time varying sliding surface which is obtained by parameter estimation method. The presented algorithms have been applied to the problem of depth control of an AUV. Resulting performances have been tested by simulation.

Keywords: AUV, discrete-time, sliding mode control, adaptive control.

1 Introduction

The performance of the control of the autonomous underwater vehicle (AUV) is affected by several elements such as: uncertainty in the model knowledge, presence of hydrodynamic effects, coupling effects between horizontal and vertical planes, system nonlinearity and so on. An adaptive sliding mode control approach is proposed. First, a direct sliding mode control scheme is introduced. This controller is further incorporated into the adaptive control strategy. This approach combines the advantages of traditional adaptive control with sliding mode control without knowing the accurate dynamic model of the system. The introduced controller compensates the structured and unstructured uncertainty of the vehicle and the environment. It does not require an accurate model of the vehicle dynamics, and ensures robust performance in the presence of disturbances and dynamic uncertainty.

Traditional sliding mode control algorithms are always designed in continuous mode, but control algorithms are always ran in discrete time implementation that may lead the system to chatter along the desired sliding mode and even to instability. Therefore, the discrete-time sliding mode method is needed to counter this problem. An adaptive parameter identifier is used to estimate the discrete model.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 157–164, 2009.

The control system architecture is shown in fig.1.

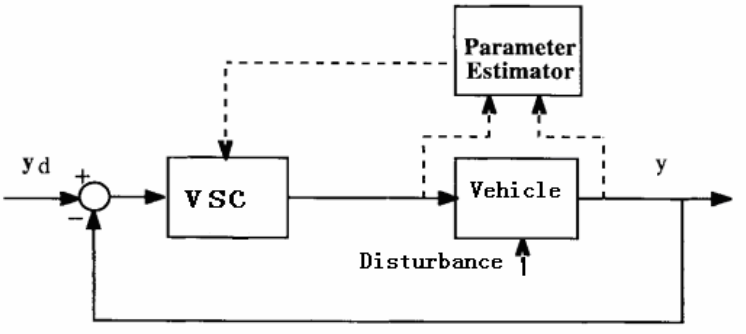


Fig. 1. Control system architecture

2 Dynamic Model

It is supposed that the vehicle is operating around a desired forward speed. Without regards to the nonlinear components, the equations of motion can be reduced to the following simple forms:

According to Newton theory:

$$\begin{aligned} m(\dot{w} - u_0 q) &= Z \\ I_y \dot{q} &= M \end{aligned} \tag{1}$$

The heave force Z and moment M can be described as follows:

$$Z = Z_{\dot{w}}\dot{w} + Z_{\dot{q}}\dot{q} + Z_w w + Z_q q + Z_\delta \delta \tag{2}$$

$$\begin{aligned} M &= M_{\dot{w}}\dot{w} + M_{\dot{q}}\dot{q} + M_w w + Mq - mg(z_G - z_B)\sin\theta + M_\delta \delta_s \\ &\approx M_{\dot{w}}\dot{w} + M_{\dot{q}}\dot{q} + M_w w + M_q q - W\overline{BG_z}\theta + M_\delta \delta \end{aligned} \tag{3}$$

At the stable point we have $\theta_0 = q_0 = \phi_0 = 0$ so:

$$\begin{aligned} \dot{\theta} &= q \\ \dot{z} &= -\theta u_0 + w \end{aligned} \tag{4}$$

The heave speed w is very slow, so the dynamics can be written as:

$$\begin{bmatrix} \dot{q} \\ \dot{\theta} \\ \dot{z} \end{bmatrix} = \begin{bmatrix} \dfrac{M}{I_y - M_{\dot{q}}} & -\dfrac{\overline{BG_z}w}{I_y - M_{\dot{q}}} & 0 \\ 1 & 0 & 0 \\ 0 & -u_0 & 0 \end{bmatrix} \begin{bmatrix} q \\ \theta \\ z \end{bmatrix} + \begin{bmatrix} \dfrac{M_\delta}{I_y - M_q} \\ 0 \\ 0 \end{bmatrix} \delta_s \tag{5}$$

The dynamic model for control in the vertical plane yields the general form of state equation in discrete time:

$$x(k+1) = Ax(k) + Bu(k) + d \tag{6}$$

where

$$A = \begin{bmatrix} -a_1 & -a_2 & 0 \\ 1 & 0 & 0 \\ 0 & -u_0 & 0 \end{bmatrix} \tag{7}$$

$$B = \begin{bmatrix} b \\ 0 \\ 0 \end{bmatrix} \tag{8}$$

$$x = [q, \theta, z]^T \tag{9}$$

u is the elevator angle, and d represents the system disturbance vector. A control law can be designed using this simplified equation.

From the state space equation (6) we can deduce the input and output model of pitch angel θ:

$$y(k+2) = -a_1 y(k+1) - a_2 y(k) + bu(k) + d$$
$$y = \theta \tag{10}$$

The coefficients, a1, a2 and b verify according to the vehicle speed. Due to the difficulty of modeling the system disturbances d and handling the coupling effects, a robust controller is needed to control the vehicle in the presence of system uncertainties.

3 Sliding Mode Controller Design

The discrete-time sliding mode control is briefly described in this section. A single-input multiple-output discrete system is described by:

$$x(k+1) = A(k)x(k) + B(k)u(k) \tag{11}$$

Where x(k) is a n dimensional state vector and u(k) is a scalar control input. It is assumed that the matrix $B(k)$ is a known constant B and the system matrix $A(k)$ is modeled as a system model A_0 with uncertainties ΔA at time step k.

If the system uncertainty satisfies the matching condition ΔA is reduced to:

$$\Delta A = Bd(k)^T \tag{12}$$

Where $d(k) = [d_{k1}, d_{k2}, ..., d_{kn}]$. It is assumed that the uncertainties are bounded and the maximum bound of each component is known. Define a vector

$$D^T = [D_1, D_2, ..., D_n] \tag{13}$$

Where each D_i is a positive constant satisfying the following condition:

$$D_i > | d_{ki} | \text{ k=1,2 to n} \tag{14}$$

If the desired trajectory r(k) is given at each time step, the error vector will be obtained as

$$e(k) = x(k) - r(k) \tag{15}$$

The switching hyperplane s(k) is defined as

$$s(k) = G^T e(k) \tag{16}$$

Where the G^T is designed such that the system is stable on the hyperplane $s(k) = 0$.

It is supposed that the control input is composed of equivalent control and switching control. The switching inputs are exclusively available in the outside of an equivalent control region, which is a sector near the switching hyperplane.

By applying the condition:

$$s(k+1) = s(k) \tag{17}$$

which is analogous to the sliding mode condition $\dot{s} = 0$ in the conventional sliding mode controller, the equivalent control reduces to

$$u_e(k) = \alpha^{-1} G^T \Delta r(k+1) - \alpha^{-1} G^T (A_0 - I) x(k) \tag{18}$$

$$\text{Where } \alpha = G^T B \text{ and } \triangle r(k+1) = r(k+1) - r(k) \tag{19}$$

A switching control law for the system to track a reference trajectory and to keep the state on the hyperplane in the presence of the uncertainties is suggested as follows:

$$u(k) = \alpha^{-1} G^T \triangle r(k+1) - \alpha^{-1} G^T (A_0 - I) x(k) + k \operatorname{sign}(s(k)) \; k > D^T | x | \tag{20}$$

where the first two terms on the right side is the equivalent control, and the last term is the switching control for robustness improvement.

Proof:

The sliding mode existing condition can be stated as follows:

$$s(k)\Delta s(k+1) < 0 \tag{21}$$

$$\begin{aligned} \Delta s(k+1) &= s(k+1) - s(k) \\ &= G^T \Delta A x(k) + G^T BK \text{ si } gn(s(k)) \\ &= G^T Bd(k)x(k) + G^T BK \text{ si } gn(s(k)) \\ &\le G^T B(D^T x(k) + K \text{ si } gn(s(k))) \end{aligned} \tag{22}$$

When $k > D^T \mid x \mid$ we have

$$s(k)\Delta s(k+1) < 0 \tag{23}$$

The upper limit of each switching gain guarantees the stability of the discrete-time quasi-sliding mode control.

A saturation function is used to reduce the control chatter and transient performance of the system.

$$sat(s/\Delta) = \begin{cases} 1, s > \Delta \\ ks, \mid s \mid < \Delta \\ -1, s < \Delta \end{cases} \quad k = 1/\Delta \tag{24}$$

4 Self-tuning Sliding Mode Control

The algorithm designed above assumes that the coefficients a1, a2 and b are known before and change slowly. This condition is not always satisfied. When coefficients uncertainties are present in the plant, the performances of the sliding mode control system may chatter badly and even becomes unstable.

The performances of the system can be improved by updating the control law parameters through on-line estimation.

We use least squire method in this paper with dead zone modification.

$$\hat{\theta}_k = [a_1, a_2, b] \tag{25}$$

$$\hat{\theta}_{k+1} = \hat{\theta}_k + K_k \alpha(k)(y_{k+1} - q_{k+1}\theta_k) \tag{26}$$

$$K_k = P_k q_{k+1}^T (\rho + q_{k+1} P_k q_{k+1}^T)^{-1} \tag{27}$$

$$P_{k+1} = (I - K_k q_{k+1}) P_k / \rho \tag{2}$$

$$\alpha(k) = \begin{cases} 1, | y_{k+1} - q_{k+1}\theta_k | > g \\ 0, | y_{k+1} - q_{k+1}\theta_k | < g \end{cases} \tag{29}$$

Dead zone parameter g can be chosen according to the level of sensor noise。

The parameters of vertical dynamic model (10) can be updated using $\hat{\theta}_k$。

5 Simulation

Simulation has been made to verify the effectiveness of the controller presented above. The disturbances are caused by the variation of forward speed, sensor noise, coupling forces, and the current. The switching gains and the disturbances are chosen as $G^T = [1, 0.15, 0.0001]$ and $D^T = [1,1,1]$. The estimator parameter are selected as ρ =0.99, $\alpha(k)$ =1. The switching gain K is 0.15.

Figs.2,3 show the step responses of the vehicle with the adaptive sliding mode controller in the presence of system uncertainty and environment disturbance. The desired depth is 10 meters. The total simulation time is 50 seconds, and at the 25th second a parameter variation of 25% with respect to the nominal value has been applied to all the parameters appearing in the model (30) and a 0.5 meter sudden variation to depth is simulated.

Figs. 2 depicts the desired depth and real depth response. Solid line denotes the desired value and dashed line denotes the real value.

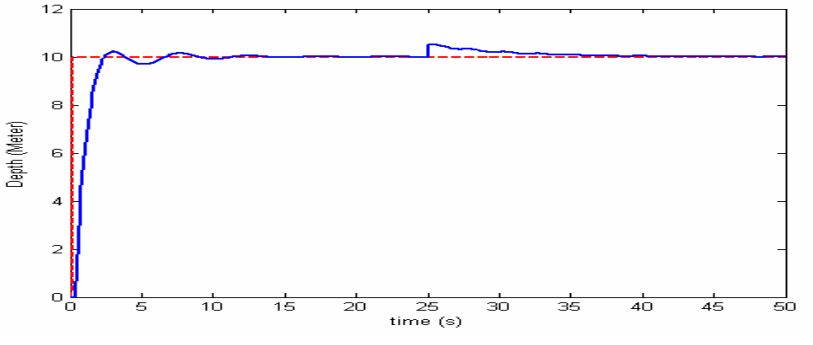


Fig. 2. Vehicle depth response

Figs. 3 depicts the theta angle response. Solid line denotes the desired value and dashed line denotes the real value.

Figs. 4 shows the elevator input.

Figs. 5 shows the parameter tracking error.

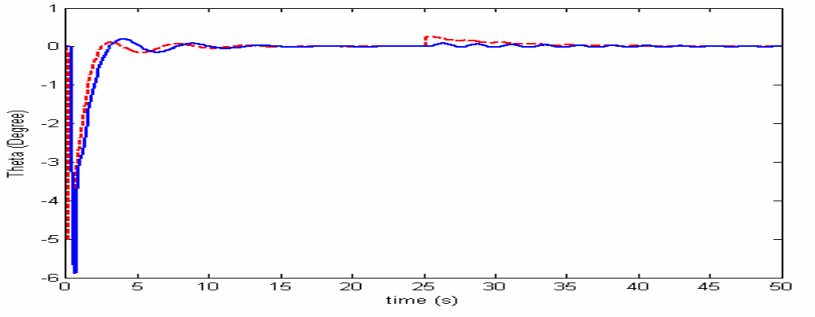


Fig. 3. Vehicle theta angle response

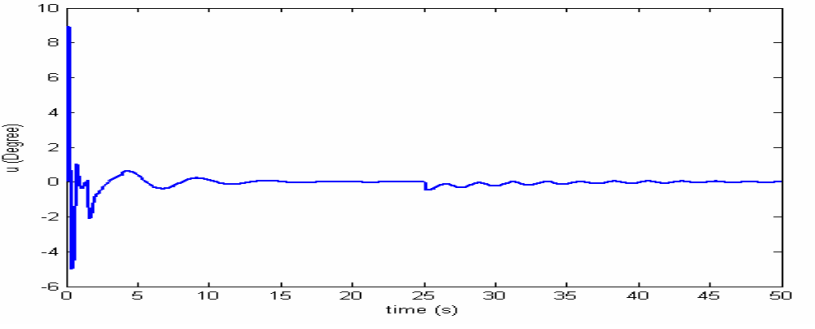


Fig. 4. Vehicle elevator input

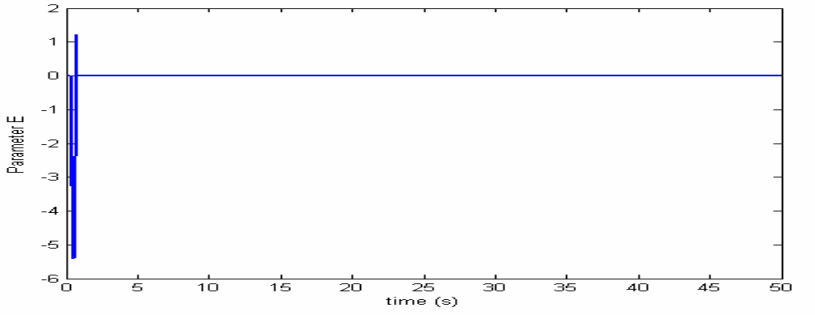


Fig. 5. Parameter estimating error

The figures show that the controller generates stable control signals, which are robust to uncertainty of system parameters and environment disturbance.

6 Conclusion

The adaptive parameter identifier estimates the input-output model, and the sliding mode controller compensates external disturbances. No state feedback is needed, but only output measurements are fed to the controller. The robustness of the algorithm is ensured both by the sliding mode mechanism and by the adaptive characteristics of the control law. The controller provides robust performance of coordinated control in the presence of uncertainties about the dynamics and the hydrodynamic disturbances.

Extensive computer simulations are performed to verify the feasibility and efficiency of the proposed control algorithm. The results show the estimated parameters are converged and the robustness and accuracy are obtained.

References

[1] Healey, A.J., Lienard, D.: Multivariable Sliding Mode Control for Autonomous Diving and Steering of Unmanned Underwater Vehicles[J]. IEEE Journal of Oceanic Engineering 18(3), 327–339 (1993)

[2] Antonelli, G., Chiaverini, S., Sarkar, N., et al.: Adaptive Control of an Autonomous Underwater Vehicle: Experimental Results on ODIN. IEEE Transactions on Control Systems Technology 9(5), 756–765 (2001)

[3] Lee, P.-M., Hong, S.-W., Lim, Y.-K.: Discrete-time quasi-sliding mode control of an autonomous underwater vehicle. IEEE Journal of Oceanic Engineering 14(3), 388–395 (1999)

[4] Letizia, M., Corradini, Orlando, G.: A Discrete adaptive variable structure controller for MIMO Systems, and Its Application to an Underwater ROV. IEEE Transactions on Control Systems Technology 5(3), 349–359 (1997)

[5] Cristi, R., Fotis, A.P., Healey, A.J.: Adaptive sliding mode controll of autonomous underwater vehicles in the dive plane. IEEE Journal of Oceanic Engineering 15(3), 152–160 (1990)

[6] 松, . [M]. 辽 科 术 (2000)

[7] Healthy, A., Lienard, D.: Multivariable sliding mode control for autonomous diving and steering of unmanned underwater vehicles. IEEE Journal of Oceanic Engineering 18(3) (1993)

[8] Ljung, L.: System Identification: Theory for the User. Prentice-Hall, Englewood Cliffs (1987)

[9] Clarke, D.W., et al.: Generalized Predictive Control-Part1 The Basic Algorithm. Automatic 23(2), 137–148 (1987)

[10] 泰. 艇 [M]. 业 (6), 68 (1981)

Emergent Behavior Control Patterns in Robotic Collectives*

Razvan-Dorel Cioarga[1], Mihai V. Micea[1], Vladimir Cretu[1], and Daniel Racoceanu[2]

[1] Department of Computer and Software Engineering (DCSE), "Politehnica" University of Timisoara, 2, Vasile Parvan Blvd., 300223 - Timisoara, Romania
[2] Image Perception, Access & Language (UMI CNRS 2955), International Research Unit (CNRS, NUS, I2R A*STAR, UJF), Singapore, 1 Fusionopolis Way, #21-01 Connexis South Tower, Singapore 138632
{razvan.cioarga,mihai.micea,vladimir.cretu@cs.upt.ro}@cs.upt.ro, daniel.racoceanu@ens2m.fr

Abstract. This paper focuses on the implementation and evaluation of a set of integrated models for the representation of emergent behavior control patterns in robotic environments. The models have been validated on a custom developed emergent behavior simulator and tested using the CORE-TX (COllaborative Robotic Environment - the Timisoara eXperiment) prototype platform. Four metrics (pheromone intensity, path affinity, reachability and liveness) are introduced and used to evaluate the performance of the proposed control patterns. Experimental results for an environment which employs ant colony behavior patterns in obstacle avoidance applications show that the emergent behavior of the robotic collective is triggered by a number ranging from 9 to 11 entities. The results are also consistent with the theoretical model-based predictions. When doubling the number of entities, the performance of the system can be further increased by 19.3%. On the other hand, a high concentration of entities has been noted to affect the emergent behavior of the system and, thus, its performance, mainly due to the interaction overhead. An upper bound to the number of individuals has been computed, based on a set of parameters which model each particular application. The experimental validation of the proposed behavior control patterns endorses them as a good framework for the analysis and development of complex applications which require collaborative and distributed intelligence, perception and operation.

1 Introduction

The evolution of information processing equipment and the need of adaptable digital control systems generate an extraordinary spreading of digital equipment in all fields of life including military applications, service industries, space exploration, agriculture, mining, factory automation, health care, waste management, disaster intervention

* This work is supported by the Romanian Ministry of Education and Research, through the grant PNCDI II ID-22/2007-2010 and, in parts through the grant PNCDI II PDP-2306/2007-2010.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 165–173, 2009.

and domestic applications. Most of such equipment is needed to be more or less autonomous, meaning that some degree of intelligence must be embedded into them. The trend of miniaturizing electronic devices often leads to the necessity of small, relatively independent entities which can interact with each other and with other entities.

The field of emergent behavior is mainly based on the study of the natural world in order to retrieve simple natural algorithms that can be applied in other fields of interest. The simplicity of these patterns and the interaction between them can solve some of the most important problems in embedded systems: energy consumption, communication protocols, efficiency and reliability features, obstacle avoidance algorithms etc. Emergent phenomena are usually unpredictable, representing non trivial results of simple interactions between simple entities. As a result, there are several combinatorial complexity problems for which finding an optimal solution is difficult or sometimes impractical because of the high dimensions of the search field or because of the computational requirements involved by the existing constraints.

This paper describes the implementation and evaluation of a set of integrated models for the representation of such emergent phenomena in robotic environments. These mathematical models have been validated both by using the Mathematica calculation software package and a custom developed emergent behavior simulator called eBMS (Emergent Behavior Model Simulator). They have also been tested using the CORE-TX (COllaborative Robotic Environment - the Timisoara eXperiment) prototype platform [1].

2 Related Work

An algorithm which uses a population of agents which can cooperate in order to reach their goal (solving combinatorial optimization problems) is presented in [2]. The algorithm uses Ant Colony Optimization (ACO) techniques. The pheromone is implemented using shared memory; also it decays over time. In [3], an algorithm based on ACO is successfully used for solving the resource-constrained project scheduling problem. In [4], another ACO based algorithm is used for solving Single Machine Total Tardiness Problem (SMTTP).

In order to solve combinatorial complexity problems heuristic techniques are preferred, to generate solutions which are not always optimal ([5], [6], [7]). The meta-heuristics used are optimizations based on ant colonies studies (Ant Colony Optimization [8]) and global convergence (Global Convergence [9]) in which case a great number of emergent behavior systems are exponentially converging to the required result, the convergence rate increases with the number of entities involved. Another extremely important domain of applicability of emergent behavior patterns is represented by NASA's future spatial missions involving new exploration technologies based on swarm emergent behavior [10], [11].

Although these meta-heuristics are sustained by a significant mathematical basis, there are no formal descriptions of the emergent behavior patterns used, just some partial representations mostly seeking the global effect of the patterns. The triggering of the emergent behavior patterns is not sufficiently presented. Also, there are no models to be used to calculate the conditions which are necessary to trigger the emergent behavior patterns.

3 The Emergent Behavior Model

The nondeterministic emergent behavior of natural systems if formed by the behavior of simple agents which follow a limited set of simple rules, based on their local view of the system. The individual agents are not aware of the global state neither of the system nor of the final functional purposes of the system.

The pheromone exchange is one of the most used ways of communication in emergent natural systems. Ant colonies are a very important example of a decentralized system that has a pheromone based communication. Individual ants form a local image of the whole system based on the pheromone exchange between them and the ants they meet. Different types of information correspond to different types of pheromones; ants also extract information from the intensity of the pheromone. By extracting information from pheromones, the ants determine the behavior of the neighboring ants and thus determine their own role without establishing any dialogues or any other type of negotiation.

One of the most frequent class of applications which can be approached with ant-inspired emergent behavior patterns is resource gathering with obstacle avoidance. This will be used in the following paragraphs to exemplify some of the principles of the emergent behavior model.

The study of real-world ant colonies reveals that ants are capable of identifying the path between their colony and the food source without using any visual indicators. The ants deposit certain amounts of pheromones on the objects of interest or on their path, thus marking them for further use. Usually, ants follow the path with the greatest intensity of pheromones. As seen in Fig. 1, if an obstacle obstructs the path, the ants try to go around the obstacle in search of the lost pheromone path until they reach it on the other side of the obstacle. Considering that all the ants move at the same speed, the ants which, by chance, have chosen in the beginning the shortest path, will arrive on the other side and to the goal faster. By doing this, the ants deposit larger amounts of pheromones of the shortest path, thus forcing the rest of the ants to follow that path.

The emergent behavior model which is presented in this article is based on the study of ant colonies and formalizes a system of independent entities which have limited

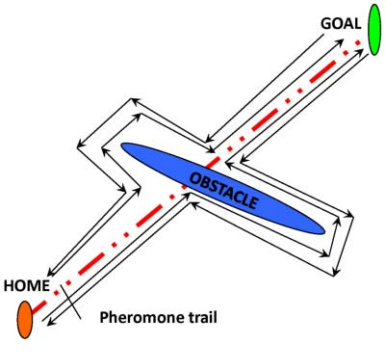


Fig. 1. Obstacle on the pheromone path

knowledge of their surroundings (local knowledge of the environment) and may interact which each other. Some fields of application for this model are: obstacle avoidance, exploration, resource gathering etc.

This model can be used to calculate the minimal number of entities which are necessary to trigger the emergent behavior, the time it takes for those entities to develop the emergent behavior and the time needed for the completion of the goal. The model is based on a series of application-dependent parameters, such as the entry rate, path affinity, goal and so on.

The system of entities which is represented by the model is an open system in which entities enter the system at a certain rate R (i.e. at a certain number per time unit). N_{E} represents the current number of entities within the system at a given time.

The number of entities which are currently on a given path is denoted as N. In the resource gathering with obstacle avoidance example, there numbers associated to each existing paths: N_{S} for the shortest path and N_{L} for the longest path.

Each path that an entity may use has a value, A, indicating the degree of preference of that entity for the respective path. That value is called the path affinity. For the resource gathering with obstacle avoidance class of applications, there is an affinity calculated for each of the existing paths: A_{S} for the shortest path and A_{L} for the longest path.

Assuming that the affinities of the paths are equal in the obstacle avoidance example, the entity may choose any path randomly with a given probability P. This paramter is also associated to each of the path: P_{S} representing the probability to choose the shortest path and P_{L} representing the probability to choose the longest path.

A certain amount of time t is necessary in order to avoid the obstacle by any entity. The parameter t_{S} is the time it takes one entity to go around the obstacle on the shortest path and t_{L} the time it takes one entity to go around the obstacle on the longest path. The path affinities A_{S}, respectively A_{L} will be affected only after t_{S}, respectively t_{L} seconds as those are the minimal times that one given entity can go around the obstacle and increase that path's affinity. This established the model's initial conditions.

Depending on the application which is modeled, the *goal* function, Γ, may have different interpretations. In obstacle avoidance applications, *goal* is arriving to a set of *target* coordinates with a minimal number of entities and

$$\Gamma = ((x_c - x_g)^2 + (y_x - y_g)^2)^{-1} \tag{1}$$

where(x_{c}, y_{c}) represent the current coordinates and (x_{g}, y_{g}) the target coordinates.

In exploration applications, *goal* is covering the most of the enviroment with the minimal number of entitites. Resource gathering applications, which are a combination of the two types presented above, have a *goal* function which also implies that the entity returns to a *home* positions.

The model represents N_{S} as a recursive function which is calculated at each step i from a function of its previous values.

Then

$$N_S(i) = \Gamma * \begin{cases} N_S(i-1) + \lambda, & A_S(i-1) > A_L(i-1) \\ N_S(i-1), & A_S(i-1) < A_L(i-1) \\ N_S(i-1) + \lambda, & A_S(i-1) = A_L(i-1), P_S > P_L \\ N_S(i-1), & A_S(i-1) = A_L(i-1), P_S \leq P_L \end{cases} \tag{2}$$

with the initial conditions

$$N_S(0) = 0, N_S(1) = \begin{cases} 1, P_S > P_L \\ 0, P_S \leq P_L \end{cases}, N_S(2) = \begin{cases} 0, P_S > P_L \\ 1, P_S \leq P_L \end{cases} \tag{3}$$

where

$$\lambda = \frac{\mathrm{d}N_e}{\mathrm{d}t} r(1, P_S, 0, P_L) \ . \tag{4}$$

and λ represents the number of entities which may choose the shortest path in the unit of time and r(a, P_a, b, P_b) randomly picks either option a or option b with the occurrence probabilities of P_a, respectively P_b.

Similarly to (2),

$$N_L(i) = \Gamma * \begin{cases} N_L(i-1), & A_S(i-1) > A_L(i-1) \\ N_L(i-1) + \rho, & A_S(i-1) < A_L(i-1) \\ N_L(i-1), & A_S(i-1) = A_S(i-1), P_S > P_L \\ N_L(i-1) + \rho, & A_S(i-1) = A_L(i-1), P_S \leq P_L \end{cases} \tag{5}$$

with the initial conditions

$$N_L(0) = 0, N_L(1) = \begin{cases} 0, P_S > P_L \\ 1, P_S \leq P_L \end{cases} N_L(2) = \begin{cases} 1, P_S > P_L \\ 0, P_S \leq P_L \end{cases} \tag{6}$$

where

$$\rho = \frac{\mathrm{d}N_e}{\mathrm{d}t} r(0, P_S, 1, P_L) \ . \tag{7}$$

and ρ represents the number of entities which may choose the longest path in the unit of time.

Using a set of helper functions, (2) becomes

$$N_S(i) = N_S(i-1) + \lambda l(P_S - P_L) eq(A_S - A_L) + \lambda l(A_L - A_S) \tag{8}$$

where $l(x)$ is 1 if $x > 0$ and 0 otherwise and $eq(x)$ is 1 if $x = 0$ and 0 otherwise.

Similarly, (5) becomes

$$N_L(i) = N_L(i-1) + \lambda l(P_L - P_S) eq(A_S - A_L) + \lambda l(A_S - A_L) \tag{9}$$

In real life situations some of the entities may be lost in the environment due to a number of possible failures so the number of entities which are necessary for the trigerring the emergent behavior is calculated from minimizing the function:

$$N_S + N_L + \zeta = N_E \tag{10}$$

where ζ represents the number of entities which are lost in the enviroment.

The amount of time necessary for achieving the emergent behavior can be calculated by integrating (10) with the time variable, t.

The emergent behavior model is composed of the N_E, A, N, t, P parameter together with (8), (9) and (10).

4 Validation of Model

The model presented in the last chapter has been formally validated using the mathematical, computational software package know as Mathematica. The model has also been simulated using a specially developed software simulator, called eBMS.

4.1 Formal Validation of the Model Using Mathematica

Case1: Real Life Example. In the first case the parameters t_S and t_L have very different values(the old pheromone path is more than a half away from the middle section of the obstacle). Also the probability of choosing either paths are very similar (as they would be in real life situations). The model parameter values are $t_S = 4$, $t_L = 10$, $P_S = 0.5$, $P_L = 0.5$. As shown in Fig. 2 all the entities rapidly decide to follow the shortest path. A number of 9 entities is neccessary to trigger this emergent behavior.

Case 2: Worst Case Scenario. The second case is the most disadvantageous case presented, in which the parameters t_S and t_L have similar values(the old pheromone path crosses almost the middle of the obstacle). Also the probability of choosing the "wrong" path is much greater than the probability of choosing the "good" path. All these conditions mean that becase the first entity who has chosen the shortest path will arrive on the other side almost at the same time with the first entity who has chosen the longest path, in this time interval a greater number of entities would have chosen the longest path, thus increasing its affinity more. The model parameter values are $t_S = 9$, $t_L = 10$, $P_S = 0.1$, $P_L = 0.9$.

As shown in Fig. 3 although during the initial unstable phase of the system the entities are undecided choosing either the shortest or the longest path (because of the high probability of choosing the longest path), after 11 entities have entered the system, it stabilizes and all further entities which are entering the system choose the shortest path.

4.2 Simulating the Model Using eBMS

A custom developed emergent behavior simulator, called eBMS (Emergent Behavior Simulator) has been used for validating the model in (10).

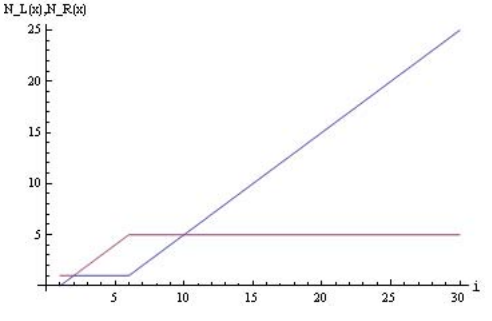


Fig. 2. The first test case validated on Mathematica

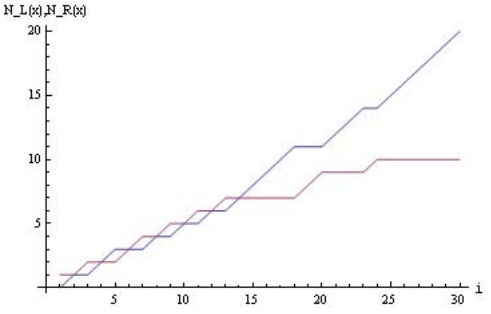


Fig. 3. The second test case validated on Mathematica

The simulator is an object oriented application written in C++ which takes as inputs a list of Entity objects each entity having the following parameter: the current position, the goal, sensor array (touch sensor, positioning sensor) and physical dimensions. The simulator uses the list of entities to simulate their behaviour according to a eBML file attached to each entity.

The application uses a simulated environment (the width is 10m, the length is 10m) and a standard entity size of 10cm. The simulated environment also has randomly placed objects.

The same cases have been simulated and the results are presented next.

Case 1. The model parameters have the same values as the ones presented in the corresponding case of formal validation As shown in Fig. 4, the simulator yields similar results to the respective case validated in Mathematica. A number of 10 entities is necessary to trigger the emergent behavior in this case. The small difference between this value and the one from Mathematica is explained by the possible collisions between the entities.

Case 2. Also in this case, the simulator yields similar results to the respective case validated in Mathematica. A number of 12 entities is necessary to trigger the emergent behavior in this case.

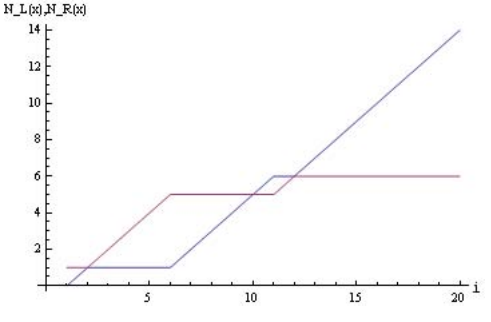


Fig. 4. The first test case validated using the custom simulator

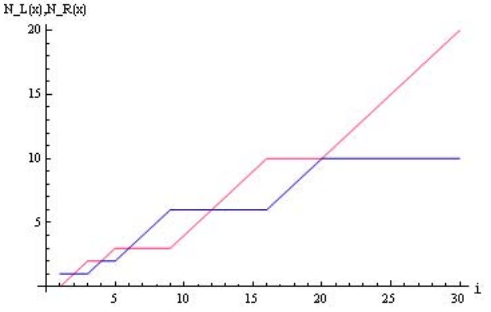


Fig. 5. The second test case validated using the custom simulator

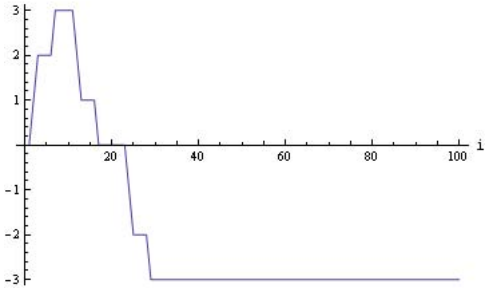


Fig. 6. Liveness of the system

4.3 Performance Evaluation

Validating the model in the simulator has presented an upper bound to the number of entities which can be used to trigger the emergent behavior. A high concentration of entities greatly affects the emergent behavior of the system and, thus, its performance, because of the interaction overhead and possible collisions in the system.

This upper bound to the number of individuals has been computed, based on the particular set of parameters of the simulator. As seen in Fig. 6 the liveness of the system drops when the number of entities increases over 30. Inside the calculated interval of 9 to 30 possible entities, increasing he number of entities further increases the overall performance of the system. When doubling the number of entities from 9 to 18 increases the performance by a maximum of 19.3%.

5 Conclusions

This paper introduces new formalisms which offer a mathematical background for the modeling of emergent behavior control patterns and demonstrates their use in modeling emergent behavior by presenting an obstacle avoidance algorithm.

These formalisms represent the mathematical background for analysing and evaluating the interactions and the degree of cooperation in robotic systems and intelligent

sensor networks. The resulting model can also be used as a basis for developing advanced control techniques on autonomous vehicles.

The set of models have been validated using both the Mathematica calculation application and the eBMS custom designed simulator. Both methods yielded similar results, thus proving the models.

The experimental results of the model applied to an obstacle avoidance algorithm have presented an interval in which the emergent behavior can be triggered with a lower bound of 9 entities and a upper bound of 30 entities.

The experimental validation of the proposed behavior control patterns endorses them as a good framework for the analysis and development of complex applications which require collaborative and distributed intelligence, perception and operation.

References

[1] Cioarga, R.D., Micea, M.V., Ciubotaru, B., Chiuciudean, D., Stanescu, D.: CORE-TX: Collective Robotic Environment - the Timisoara Experiment. In: Proc. 3-rd Romanian-Hungarian Joint Symp. on Applied Computational Intelig., SACI 2006, Timisoara, Romania, May 2006, pp. 495–506. IEEE, Los Alamitos (2006)

[2] van der Zwaan, S., Marques, C.: Ant Colony Optimization for Job Shop Scheduling. In: Proc. Third Workshop on Genetic Algorithms and Artificial Life, Lisbon, Portugal (1999)

[3] Merkle, D., Middendorf, M., Schmeck, H.: Ant Colony Optimization for Resource-Constrained Project Scheduling. In: Proceedings of the Genetic and Evolutionary Computation Conference, Las Vegas, Nevada, pp. 893–900 (2000)

[4] Iredi, S., Merkle, D., Middendorf, M.: Bi-Criterion Optimization with Multi Colony Ant Algorithms. In: Zitzler, E., Deb, K., Thiele, L., Coello Coello, C.A., Corne, D.W. (eds.) EMO 2001. LNCS, vol. 1993, p. 359. Springer, Heidelberg (2001)

[5] Dorigo, M., Di Caro, G.: The ant colony optimization meta-heuristic. In: Corne, D., Dorigo, M., Glover, F. (eds.) New Ideas in Optimization, pp. 11–32. McGraw-Hill, New York (1999)

[6] Dorigo, M., Maniezzo, V., Colorni, A.: The Ant System: Optimization by a colony of cooperating agents. IEEE Trans. Syst, Man, Cybern. 26(2), 29–41 (1996)

[7] Dorigo, M., Gambardella, L.M.: Ant colony System: A Cooperative Learning Approach to the Travelling Salesman Problem. IEEE Trans. on Evolutionary Computation 1(1) (April 1997)

[8] Maniezzo, V., Gambardella, L.M., de Luigi, F.: Ant Colony Optimization: New Optimization Techniques in Engineering, by Onwubolu, G.C., Babu, B.V., pp. 101–117. Springer, Heidelberg (2004)

[9] Van Dyke Parunak, H., Brueckner, S., Sauter, J., Matthews, R.: Global Convergence of Local Agent Behaviors. In: AAMAS 2005, Utrecht, Netherlands, July 2005, pp. 305–312 (2005)

[10] Truszkowski, W., Hinchey, M., Rash, J., Rouff, C.: NASA's swarm missions: The challenge of building autonomous software. IEEE IT Professional 6(5), 47–52 (2004)

[11] Hinchey, M., Rouff, C., Rash, J.: Requirements of an Integrated Formal Method for Intelligent Swarms. In: Proc. FMICS 2005, Lisbon, Portugal, September 2005, pp. 125–133 (2005)

[12] Cioarga, R., Micea, M., Ciubotaru, B., Chiciudean, D., Cretu, V., Groza, V.: eBML: A Formal Language for Behavior Modeling and Application Development in Robotic Collectives. In: Proc. International Workshop on Robotic and Sensors Environments ROSE 2007, pp. 1–6 (2007)

An Improved Road Network Modeling and Map Matching for Precise Vehicle Localization

Chenhao Wang, Zhencheng Hu, Naoko Hamada, and Keiichi Uchimura

Graduate School of Science and Technology, Kumamoto University, 2-39-1, Kurokami, Kumamoto, Japan, 860-8555

Abstract. This paper presents a precise and robust road model for Map Matching (MM). Map matching approach is generally used to move vehicle localization output to the nearest road segment in order to improve the accuracy of vehicle lateral localization. Most of the MM approaches adopted piece-wise line segment road network model, which will generate large bias at curve segment or turning point on the intersection. In this paper, a two-dimensional parabola road network model is employed in order to correctly represent vehicle's real state (position and orientation). Furthermore, an advanced longitudinal MM approach is also proposed here by comparing current road geometry model with visual lane detection results to improve longitudinal accuracy. Simulation and real road tests verified the proposed approaches.

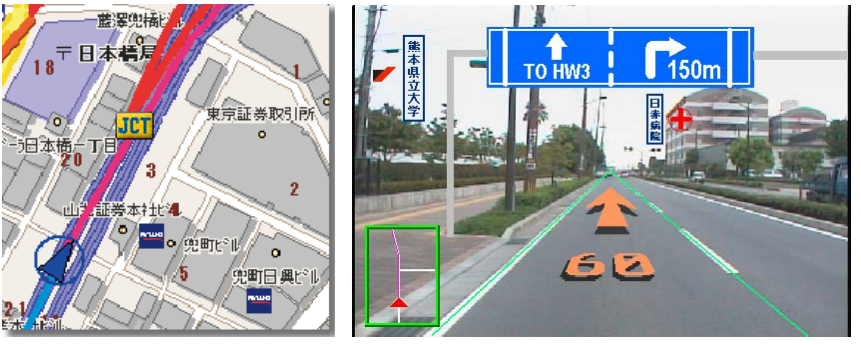
1 Introduction

GPS-based vehicle localization algorithms were well studied and widely applied in car navigation market since Pioneer®introduced the first commercial in-vehicle navigation system in 1990s. Figure 1(a) shows an example of original car navigation system output. In order to improve localization accuracy, Global Positioning System (GPS) data is always fused with Dead Reckoning (DR) systems [1] [2] [3] to provide supplement when GPS signal is not available. Other approaches like Differential GPS [4] decrease the effect of S/A (selective availability) and improve GPS performance significantly. Meanwhile, Kalman Filtering (KF) [5], Extended Kalman Filtering (EKF) [6], Particle Filtering (PF) [7] and other fusion algorithms are widely introduced in navigation system to estimate the vehicle's state robustly.

Some advanced navigation applications like VICNAS [8] requires to have more precise and faster localization results. VICNAS superimposes virtual indicators on the overlay of real scene directly (Figure 1(b)), which requires real time continuous localization results. Dead reckon method combining with GPS, 3D gyroscope and speedometer was employed to estimate vehicle localization in the previous system [9]. To further improve both lateral and longitudinal accuracy, we considered to use MM approach in our system.

Prevalent digital road maps are consisted of node points and links. Node points are generally representing road intersections or some key points on large curve segments. Links are road segments linking node points, are normally treated as straight lines. Therefore, the digial road model can be represented by a network of straight lines. However, referring to road design standard, road geometry models can be described as the combination of straight line segments, constant curvatures [14], or more complex,

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 174–184, 2009.



(a) Traditional navigation system (b) Prototype driver interface of VICNAS system

Fig. 1. Car navigation system

like parabolas [15], low-order polynomial, splines [16], or even clothoids [17]. And curvature is usually parameterized to denote each part of road network.

In conventional MM approaches, vehicle position is required to locate on the 'closest' link which assume that vehicle only runs on the road. More recent MM algorithms such as point-to-point, line-to-line [12] and point-to-curve, curve-to-curve [13], were proposed for higher precision. In these MM approaches, only position data is considered to locate the nearest link. Orientation data is only considered when vehicle approaches to the intersection where distances to different links are almost the same.

In this paper, a two-dimensional parabola road network model is introduced in the MM approach to correctly represent vehicle's real state (both position and orientation). Furthermore, an advanced longitudinal MM approach is also proposed here by comparing current road geometry model with visual lane detection results to improve longitudinal accuracy.

In the following sections, we will introduce road model in detail with comparison with previous network in Section 2.1. New road model including line and curve segment is proposed in Section 2.2. In Section 2.3 and Section 2.4, MM approach for lateral and longitudinal are introduced. Finally, we testify our approach by simulation results and outdoor experiments in Section 3.

2 A New Road Model for Map Matching

Here are some assumptions used in the proposed approach.

i. Only two-dimensional road model is considered in this paper.
ii. Road model is defined by its centreline and multi-lane road model can be easily deduced by shifting and scaling Affine transformation.
iii. Node point positions are assumed to be accurate and reliable.

According to the road design standard, a road model should combine both straight lines and curve segments (figure 2). Straight line segment is connected by two node points

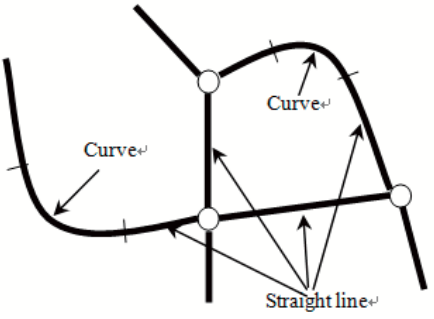


Fig. 2. General road network

with a constant orientation and zero curvature. Curve segment is more complicate with curvature variance consideration. In the practical situation curve segment is not easy to be concluded by a single geometrical model. Circular, clothoid and arbitrary polynomial are the most popular models.

2.1 Conventional MM Approaches

With the straight line road model, conventional MM methods are carried out by searching 'closest' line segment. An example is shown in figure 3 (a). According to vehicle position (v_y, v_y) and road node points (IP0, IP1), the 'closest' position (m_x, m_y) is calculated by looking for the perspective projected point on the line segment.

Road orientation is calculated by Eq.1 between two node points Figure 3(b) shows the orientation output of MM result, where you can see some significant discontinuous orientation outputs. Therefore, the conventional MM output cannot be used in the real time navigation applications like VICNAS.

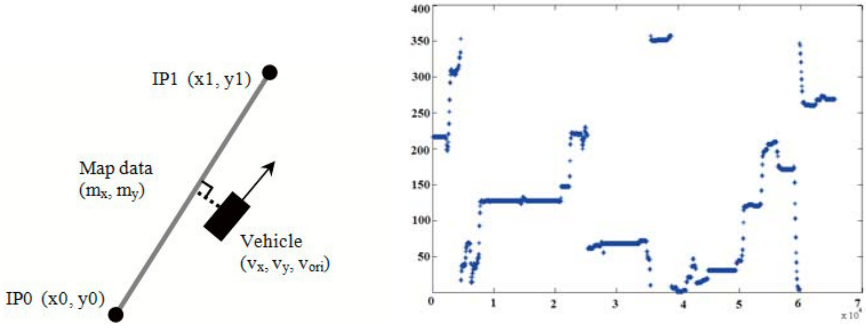
$$\boldsymbol{orientation} = \arctan \frac{y1 - y0}{x1 - x0} \tag{1}$$

2.2 Our Road Model

Our road model here is consisted of two parts, straight line and curve. To cover most of road situations, parabolas function is selected to represent road model, which is able to approximate circular arc or clothoid curve by the vertex of parabola [18]. Straight line segment can also be expressed with the zero value of curvature. The general road model is expressed as follows:

$$\boldsymbol{L(x, y)} = \boldsymbol{k\Delta L} + \frac{C_r}{2}\boldsymbol{\Delta L}^2 \tag{2}$$

where L(x,y) is position by road model and ΔL is related variance. k is linear parameter and C_r is curvature according to road model. This model could denote both straight and curve segment.



(a) Map matching on previous road model (b) Orientation outputs after map matching

Fig. 3. Conventional map matching approach

Since the road model is expressed by Eq.2, orientation function could be acquired by derivative of Eq.2.

$$orientation(\Delta \boldsymbol{L}) = \boldsymbol{k} + C_r \Delta \boldsymbol{L} \tag{3}$$

Here is an example to explain how to calculate parameters of linear and curvature segment using existing node points. All road node points are denoted with the local coordinate system (X − Y). The segment between IP0 and IP1 is a straight line and the segment between IP1 and IP2 is a curve segment. To calculate the parameters of curve model, we need to translate and rotate the current local coordinate system (X − Y) to the new coordinate system (X' − Y') which origins at IP1. The procedure of calculation is as following:

i. Calculate rotation angle θ between X − Y and X' − Y'

$$\theta = \arctan \frac{y1 - y0}{x1 - x0} \tag{4}$$

And the linear parameter could be expressed.

$$k = \tan(\theta) \tag{5}$$

ii. Rotation matrix R from X − Y to X' − Y' will be:

$$\boldsymbol{R} = \begin{bmatrix} \cos(\theta - \pi/2) & \sin(\theta - \pi/2) \\ -\sin(\theta - \pi/2) & \cos(\theta - \pi/2) \end{bmatrix} \tag{6}$$

iii. Transfer road node points in X' − Y' and $IP1'(x1', y1')$ locate at original point $(0, 0)$

$$IP2'(x2', y2') = R \times \begin{bmatrix} x2 - x1 \\ y2 - y1 \end{bmatrix} \tag{7}$$

iv. Curvature calculation by parabolas function

$$C_r = \frac{2(y2' - y1')}{x2' - x1'} \tag{8}$$

2.3 Map Matching in Lateral Direction

In section 2.2, we have shown the detailed transformation relationship between local coordinate system X − Y and related curve coordinate system X' − Y'. Because road model is set up in curve coordinate system, map matching is proposed to carry out in same coordinate system.

i. Transformation of vehicle position (v_y, v_y) to X' − Y' referring to rotation matrix T

$$\begin{bmatrix} v_x^{'} \\ v_y^{'} \end{bmatrix} = R \times \begin{bmatrix} v_x \\ v_y \end{bmatrix} \tag{9}$$

ii. Map matching in X' − Y'
$(v_x^{'}, v_y^{'})$ in X' − Y' finds the closest point $(m_x^{'}, m_y^{'})$ on new road model

iii. Transform the map position $(m_x^{'}, m_y^{'})$ to X − Y

$$\begin{bmatrix} m_x \\ m_y \end{bmatrix} = R^{-1} \times \begin{bmatrix} m_x^{'} \\ m_y^{'} \end{bmatrix} \tag{10}$$

iv. Orientation map matching
According to road model, road orientation is in continual distribution. The orientation variance $\boldsymbol{\Delta ori}$ is determined by road curvature.

$$\boldsymbol{\Delta ori} = C_r \bullet \boldsymbol{\Delta L} \tag{11}$$

Multi-lane road model can be easily deduced from this road model by shifting and scaling Affine transformation(shown in figure 4). Consequently, road network is not only consisted of road centerline, but also multiple parallel straight and curve line models.

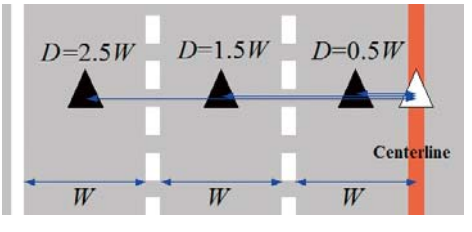
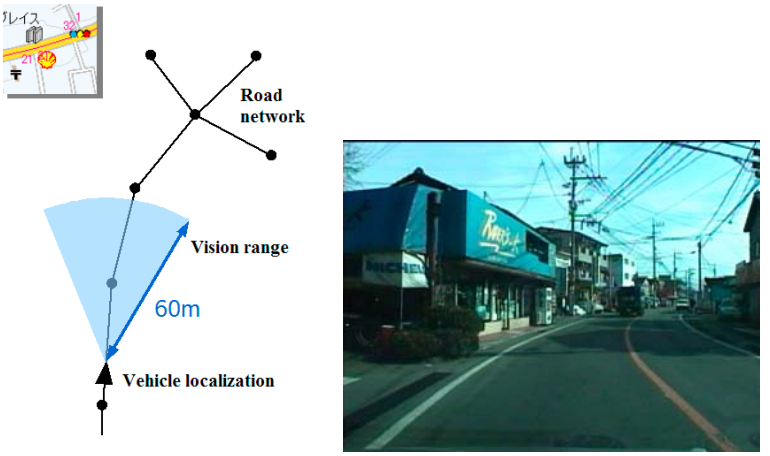


Fig. 4. Map matching on each lane

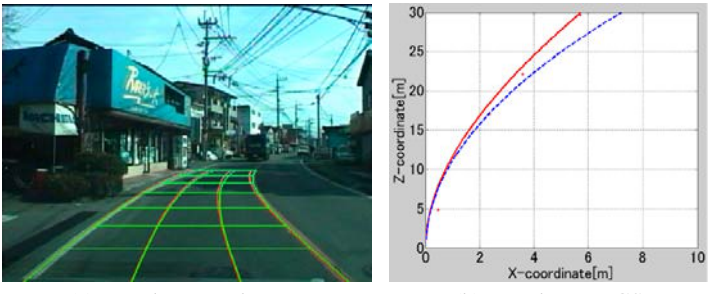
2.4 Map Matching in Longitudinal Direction

The geometry of road ahead is fused with visual lane detection result to estimate localization in longitudinal direction in our approach. Figure 5 shows vehicle localization on the road network and real scene taken by camera at same moment. Vehicle Coordinate System (VCS) and Image Coordinate System (ICS) can be simply calibrated off-line or on-line [8].



(a) Vehicle is located on road network (b) Real scene in front of vehicle

Fig. 5. Vehicle localization, road network, real scene

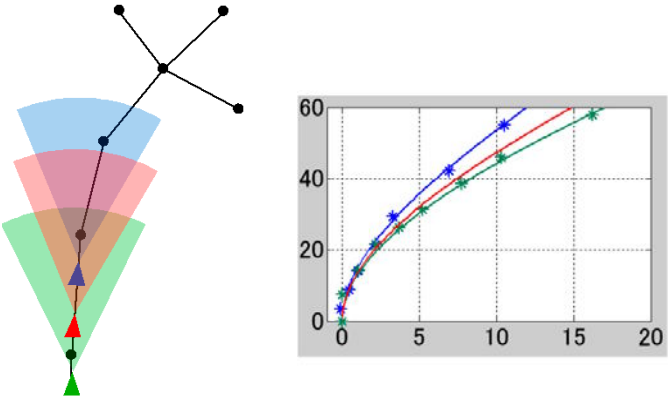


(a) Results on real scene (b) Results on VCS

Fig. 6. Comparison results between lane detection and road geometry ahead of vehicle

Firstly, road geometry could be estimated relying on vehicle localization and road node points (see Figure 5 (a)). In our approach, all node points should be transformed into VCS. After MM processing, vehicle is supposed to locate on the road centerline and vehicle central axis is parallel to the tangent direction of road centerline. Node points in front of vehicle will be approximated by the proposed road model (Section 2.2). Road geometry will be continuously estimated while vehicle is moving along the road. An example of the road geometry estimation result is expressed as red line in figure 6 (a) of ICS and figure 6 (b) of VCS.

Secondly, lane detection algorithm in VICNAS assumes in a horizontal plane and vertical curve part is ignored. Road shape parameters (road width and horizontal curvature) and camera posture (pitch and yaw) are concluded in lane detection model as vector. However, information for road geometry estimation is considered in our



(a) Vehicle localization adjustment along road network (b) Relate geometry reconstruction on VCS

Fig. 7. Road geometry reconstruction with adjustment of vehicle localization

approach. Visual detection result is shown in as green line in figure 6 (a) of ICS and road geometry estimation result is expressed as blue line in figure 6 (b) of VCS.

MM approach in longitudinal direction estimates difference of road geometry reconstruction from each information. Since vehicle localization is a key issue for geometry shape, it could be adjusted for the best matching with lane detection results. Figure 7 (a) is adjustment of vehicle position along the road network which means adjustment in longitudinal direction. And results of road geometry are shown in figure 7 (b).

3 Simulation and Experiment Results

In this paper, we choose the Super Mapple ver.6 (Shobunsha®, Japan) as digital road map database. Here, X and Y are local plane coordinates transformed from Longitude/Latitude data. And the test place (Kumamoto, Japan) is located at second district plane of Japan.

3.1 Map Matching on Curve Segment

In figure 8, 14 road node points are divided into 4 different curves and road curvature is calculated according to own curve coordinate. The positive of curvature value means left turning and the negative one means right turning. There are two 'S' shape curves in this route, from curve 1 to curve 2 and from curve 3 to curve 4.

We have discussed in section 2.2 that road model based on parabolas function could not recover the curve perfectly and road node points could not located on new road model exactly. By simulation results, figure 9 (a) shows recovery curve 2 '.' and road node points '*' in curve coordinate system. Although there are excursion errors between road model curve and node points, these errors could be ignored. And figure 9 (b) is two

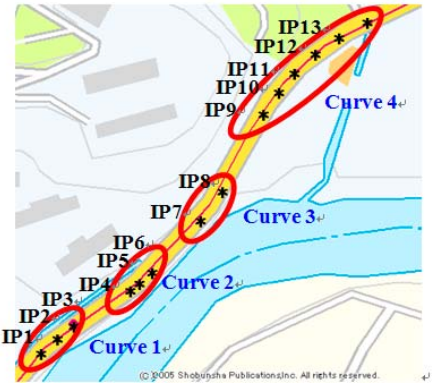
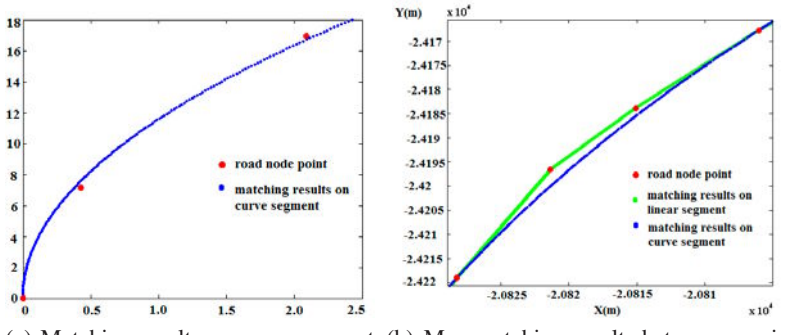


Fig. 8. Example of curve segments on digital road map



(a) Matching results on curve segment (b) Map matching results between previous model and our approach

Fig. 9. Map matching results by our approach

results individually by pervious method and our model. Red points on the graph are road node points and the green lines are supposed as matching results if using the previous road model. Blue pints are matching results on the curve based on our approach.

3.2 Extended Road Model and Matching Results

Extended road model has been mentioned in the section 2.3. Lane number and width are utilized to extend road network on the each lane. Figure 10 shows matching results on digital map directly. Comparing to traditional results on centerline (figure 10 (a)), results by extended road model provide are more accurate and reasonable.

3.3 Matching Results in Longitudinal Direction

Referring to result shown in figure 7, the best position in longitudinal direction needs back around 1.3 meters and it could match with lane detection results. Figure 11 (a)

(a) Matching results on centerline (b) Matching results on each lane center

Fig. 10. Map matching results by extend road model

(a) Successful example for correct matching (b) Bad example for failed matching

Fig. 11. Visual lane detection and road geometry on real scene

is another good example to estimate vehicle position along road network. The best position needs around 2.5 meters back. Green line is visual lane detection result and red line is recovered by current localization and digital map. Pixel error on the image less than 10 pixels is accepted in off-line processing.

However, method for longitudinal map matching is only implemented in simulations. In practical situation, two kinds of error need to be concerned. The first one is error of digital map. It depends on the type of digital map for car navigation. And the second one is error by lane detection result. It might cause big problem for vehicle localization if the result is not correct. Figure 11 (b) is bad example for MM approach in longitudinal direction.

Besides, camera posture, 3D road surface situation and so on will cause the error of matching result. That is why all tests are implemented in off-line and it is still lots of problem until application for practical situation.

4 Conclusion and Discussion

This paper concluded existing road network and its MM approach for car navigation system. To improve accuracy of vehicle localization, new type of road model with curve segment was introduced here and it would provide more accurate vehicle localization on road network as a stable reference for DR system. Besides road model was extended and MM approach could be realized on each line. MM approach in longitudinal direction was considered and algorithm was proved in this paper.

Nowadays, digital map is most reliable element in car navigation system. Despite road network and MM approach have been already employed in commercial application, more and more possible works could be brought forward for advanced application if combing with other sensors. In future, road geometry for longitudinal matching is the big challenge for our research.

References

1. Kim, J.-S.: Node Based Map Matching Algorithm for Car Navigation System. In: Proceeding of the International Symposium on Automotive Technology and Automation, pp. 121–126 (1996)
2. Degawa, H.: A New Navigation System with Multiple Information Sources. In: Proceedings of the Vehicle Navigation and Information Systems Conference, pp. 143–149 (1992)
3. Mattos, P.G.: Integrated GPS and Dead Reckoning for Low-cost Vehicle Navigation and Tracking. In: Proceedings of the Vehicle Navigation and Information Systems Conference, pp. 569–574 (1994)
4. Blackwell, E.G.: Overview of Differential GPS Methods. Global Positioning Systems, The Institute of Navigation 3, 89–100 (1986)
5. Jo, T., Haseyamai, M., Kitajima, H.: A Map Matching Method with the Innovation of the Kalman Filtering. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences E79-A, 1853–1855 (1996)
6. Kiriy, E., Buehler, M.: Three-state Extended Kalman Filter for Mobile robot localiztion. Technical report (2002)
7. Chausse, F., Laneurit, J., Chapuis, R.: Vehicle Localization on a digital map using particle filter. In: The IEEE Intelligent Vehicles Symposium, Las Vegas, Nevada, U.S.A, June 6-8 (2005)
8. Hu, Z., Uchimura, K.: Fusion of Vision, GPS and 3D Gyro Data in Solving Camera Registration Problem for Direct Visual Navigation. International Journal of ITS 4(1) (2006)
9. Wang, C., Hu, Z., Uchimura, K., Kusuhara, S.: Vehicle Localization with Global Probability Density Function for Road Navigation. In: IEEE Intelligent Vehicles Conference (2007)
10. Su, J., Zhou, D.F., Yue, C.S.: The real-time map-matching algorithm in GPS vehicle navigation. Mapping Journal 30(3), 252–256 (2001)
11. Li, J., Fu, M.Y.: Research on Route Planning and Map Matching in Vehicle GPS/Dead-Reckoning/Electronic Map Integrated Navigation System. Proc IEEE Intelligent Transportation Systems (2), 1639–1643 (2003)
12. Xiong, D.: A Three-Stage Computational Approach to Network Matching. Transportation Research C 8, 13–36 (2000)
13. White, C.E., Bernstein, D., Kornhauser, A.L.: Some map matching algorithms for personal navigation assistants. Transportation Research Part C 8(1-6), 91–108 (2000)

14. Kluge, K., Lakshmanan, S.: A deformable template approach to lane detection. In: Proc. IEEE IV, Detroit, MI (1995)
15. Redmill, K.A.: A simple vision system for lane keeping. In: Proc. IEEE ITS (1997)
16. Wang, Y., Teoh, E., Shen, D.: Lane detection and tracking using B-snake. Image Vis. Comput. 22(4), 269–280 (2004)
17. Dickmanns, E.D., Mysliwetz, B.D.: Recursive 3D road and relative ego-state recognition. IEEE trans. Pattern Anal. Mach. Intell. 14(2), 193–213 (1992)
18. Har'el, Z.: Curvature of Curve and Surfaces - A Parabolic Approach (1995)

A Vision System of Hazard Cameras for Lunar Rover BH2

Guicai Wang, Hehua Ju, and Hongpeng Feng

Beijing University of Technology, 100 Pingleyuan,
Chaoyang District, Beijing, China
wangguicai@tom.com

Abstract. According to analyze vision system of hazard cameras for lunar rover BH2 and characteristics of working environment, firstly, intrinsic and extrinsic parameters of camera are accurate calibrated and image distortion is corrected in this paper; Secondly, images are processed according to Bayer filter, image rectification, LoG filter, pyramid delaminating, photometric consistency dense matching; Finally, the paper can obtain high precision and dense disparity map. Furthermore, the paper uses three sets of experiment results on obstacle avoidance cameras vision system of lunar rover BH2 to validate the scheme. the experimental results show that the system possesses the features with less calculation amount, high reliability, robustness of illumination change and environmental noise, which can quickly and reliably achieve dense disparity map and meet the real-time avoid obstacle requirement of BH2.

Keywords: Lunar rover, stereo vision, disparity map, obstacle avoidance.

1 Introduction

Lunar exploration is key research focus of many research institutions and colleges and universities in china. Beijing University of Technology has also carried out lunar rover research works and developed the six-wheeled lunar rover BH1 and BH2. The paper mainly introduces BH2, which as shown in Fig.1 (a). BH2 has a mass of about 40 kg, is 1.1 m long, has a wheelbase of 0.85 m, and is 1.5 m high. Folded BH2 is only 0.95 m long and 0.4 m high. BH2 includes three parts: mobile structure, sensing system and control system. Sensing system perceives surrounding unknown environment and provides necessary information to autonomous navigation of BH2. The ground mobile robot can use laser, ultrasonic, infrared and radar ranging sensors and so on. However, laser demands higher energy; ultrasonic can not be used due to vacuum environment on the moon; infrared can only detect shorter range, and radar ranging sensor is heavy [1]. Meanwhile, computer vision has and will continue to play an important role in increasing autonomy of both outdoor mobile robots and planetary rovers. For example, the vision sensors have been successful used in NASA's Mars Exploration Rover (MER) [2] [3]. Matthies presents a six-wheeled binocular outdoor mobile robot which use thresholding scene points heights to detect obstacles [4]. The autonomous navigation process developed at CNES MER is based on stereo cameras perception, used to build a model of the environment and generate trajectories [5].

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 185–194, 2009.

CMU has developed a binocular stereo system for a four-wheeled lunar rover Ratler by using window selection and partial sub-sample [6] [7]. Besides Ratler, CMU has also developed a binocular stereo vision system for HMMWV [8]. Unlike Ratler, the system on HMMWV adopts a vertical baseline configuration to increase the field of view. Its stereo system provides good range information and allowed for autonomous navigation in field trials. The European Space Agency (ESA) has developed a binocular stereo vision system for Nanokhod by using automatic calibration procedure to recover the relative orientation of the cameras and pan and tilt axes [9]. These could make vision sensors as the environment sensing means of BH2 [10] [11]. BH2 has carried a total of eight cameras for four various types. There is a stereo pair of high dynamic range panoramic CCD cameras. This CCD is a 1024×1024 array of 10.6 μm pixels; the two stereo pairs of color cameras for hazard avoidance. These CCD are 782×582 array of 8.3 μm pixels; And a stereo pair of fisheye Sun senor cameras. This CCD is a 1280×1024 array of 6.0μm pixels. The panoramic cameras (PanCams) are installed on the mast. The stereo cameras separation is 300 mm. And the lens focal length is 12mm; In the lower of both ends of BH2, there are stereo pairs of hazard cameras (HazCams). One pair faces forward, the lens focal length is 12 mm. and one pair faces backward, the lens focal length is 8 mm. The stereo cameras separation is 160 mm. It is necessary that the absolute three-axes orientation of BH2 be known to enable the high-gain radio antenna dish to be accurately pointed at Earth. The direction of the Sun will be measured by a special camera called the SunCams, the lens focal length is 8 mm. The paper mainly introduces the forward-looking vision system of HazCams which uses stereo vision as shown in Fig.1 (b). The key of stereo vision is matching algorithm. On the hand, the Sunlight straight to the lunar surface without refraction and reflection light so that the lunar environment has strong light feature for without atmosphere on the moon. On the other hand, photometric consistency dense matching algorithm can have few effects of illumination and obtain dense disparity map and depth map, and which is easy to implement and meet real-time avoid obstacle requirement by using the pyramid delaminating [12]. As a result, BH2 uses photometric consistency dense matching algorithm.

Fig. 1. The lunar rover BH2 and hazard cameras

Based on above reasons, this paper introduces the algorithm process of the forward-looking vision system of HazCams is shown in Fig.2, which includes the following steps: (1) grabbing image; (2) Bayer filter; (3) image rectification using offline calibration result; (4) pyramid delaminating; (5) LoG filter; (6) ZSAD matching; (7) ordering constraint and uniqueness constraint; (8) achieving dense disparity map; (9) achieving 3D information; (10) offline calibration.

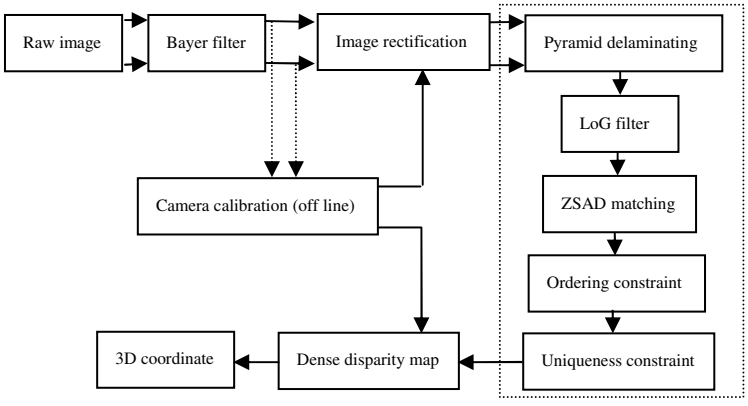


Fig. 2. The flowchart of obstacle avoidance camera vision system

The remain of this paper is structured as follows: In section 2, we introduce camera calibration, Bayer filter, image rectification, and photometric consistency dense matching, describes the proposed algorithm in this paper. In section 3, we give the results of experiment with actual test images which show the method is effective. Finally, section 4 concluded some ideas of this paper for future.

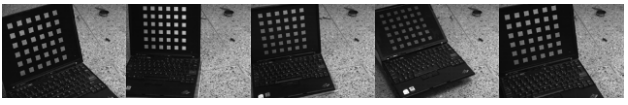
2 Algorithm Description

The key part of vision system of HazCams for lunar rover BH2 lay in camera calibration and stereo matching, camera calibration should accurate, and the matching algorithm should fast and accurate obtain disparity map as far as possible. The proposed scheme in this paper include step 4: camera calibration; Bayer filter; image rectification; and photometric consistency dense matching.

2.1 Camera Calibration

The two cameras can not be parallel placed so that can not meet the standard stereo vision system in practice, moreover, due to the lower precision of camera lens and some reasons, the process of the camera to obtain image can not strictly meet the pinhole imaging model, and the sampled images have distortion. As a result, firstly, camera calibration result is carried out; secondly, the standards stereo vision system and image distortion rectification can be achieved by using calibration results. The key of camera calibration is how to choose a calibration object? Comparing to 3D calibration object with high costs and poor machining accuracy, the paper using Zhengyou Zhang proposed 2D planar calibration pattern method [13]. It only requires camera to observe a planar pattern shown at a few (at least two) different orientations. However, Zhang's method requires planar calibration pattern to meet plane requirement as far as possible for homography transformation. Therefore, the paper uses the LCD screen of notebook as a 2D planar calibration pattern to meet calibration requirement, which display calibration points on the LCD screen. Fig.3 shows different angle calibration images.

(a) left image



(b) right image

Fig. 3. The different view calibration image

The camera calibration in this paper include step 4: introduces camera parameter; the calibration results of camera intrinsic parameter; the calibration results of camera extrinsic parameter; and analysis of calibration results.

Step1: the model of the forward-looking HazCams is JAI-CV-A70CL. the parameters of forward-looking HazCams are shown in the Table 1.

Table 1. The Camera Parameters

Camera Parameters			
Color	Resolution	Pixel Size	Baseline l_{baseline}
Bayer	782×582	8.3 μm	160 mm

Step2: the calibration results of the forward-looking HazCams intrinsic parameter are shown in Table 2, which list out one distortion factor for simplifying page length.

Table 2. The Calibration Results of Intrinsic Parameter

Camera	f_x	f_y	u_x	u_y	k_1
Left	1459.752	1450.536	385.543	251.749	-0.0045562
Right	1464.471	1455.739	371.121	268.378	-0.042636

Step3: the paper uses Euler angle expressing rigid rotation transformation: roll angle ϕ, pitch angle θ, yaw angle ψ. The rotation matrix follows (1).

$$\mathbf{R} = \begin{bmatrix} \cos\phi\cos\theta & \cos\phi\sin\theta\sin\psi - \sin\phi & \cos\phi\sin\theta\sin\psi - \sin\phi\sin\psi \\ \sin\phi\cos\theta & \sin\phi\sin\theta\sin\psi + \cos\phi\cos\psi & \sin\phi\sin\theta\cos\psi + \cos\phi\sin\psi \\ -\sin\theta & \cos\theta\sin\psi & \cos\theta\cos\psi \end{bmatrix} \tag{1}$$

However, the calibration result is rotation vector. According to Rodrigues formula (2) rotating vector is transferred into rotation matrix [14].

$$R = e^{[r]_\times} = \mathrm{I}_3 + \frac{\sin\theta}{\theta}[r]_\times + \frac{1-\cos\theta}{\theta^2} + [r]_\times^2 = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \tag{2}$$

Where $r = [\mathrm{x}, \mathrm{y}, \mathrm{z}]^{\mathrm{T}}$, $[\mathrm{r}]_\times$ is anti-symmetric matrix, I_3 is unit vector, $\theta = \sqrt{x^2 + y^2 + z^2} = \|r\|$.

The calibration results of camera extrinsic parameter are shown as follows.

Translation Vector $t = [-160.49870 \quad 2.07768 \quad 0.18791]$.

Table 3. The Calibration Results of Extrinsic Parameter

Calibration Results of Extrinsic Parameter			
Yaw angel ψ_{cal}	Pitch angel θ_{cal}	Roll angel ϕ_{cal}	Baseline $l_{\mathrm{baseline}}^{\mathrm{cal}}$
-0.0058	-0.0069	0.0248	160.49870

Step4: analysis of the calibration results. The forward-looking HazCams uses 12mm lens ($f_0 = 12mm$). Regardless of the nominal value error, the absolute error and relative error can be obtained.

$$\overline{f_{\mathrm{left}}} = \frac{f_{\mathrm{x}} \cdot \mathrm{dx} + f_{\mathrm{y}} \cdot \mathrm{dy}}{2000} = \frac{1459.752 + 1450.536}{2000} \times 8.3 = 12.0777mm ;$$

$$\overline{f_{\mathrm{right}}} = \frac{f_{\mathrm{x}} \cdot \mathrm{dx} + f_{\mathrm{y}} \cdot \mathrm{dy}}{2000} = \frac{1464.471 + 1455.739}{2000} \times 8.3 = 12.1189mm ;$$

Moreover, yaw angle ($\psi_0 = 0$), pitch angle ($\theta_0 = 0$), roll angle ($\phi_0 = 0$) in the standard stereo vision system.

Table 4. The Analysis of Camera Calibration Results

Intrinsic Parameter		Extrinsic Parameter			
Left camera	Right Camera	Yaw angel	Pitch angel	Roll angel	Baseline
Δf_{left}	$\Delta f_{\mathrm{right}}$	$\Delta\psi$	$\Delta\theta$	$\Delta\phi$	$\Delta l_{\mathrm{baseline}}$
0.0777	0.1189	0.0058	0.0069	0.0248	0.4987
$\frac{\Delta f_{\mathrm{left}}}{f_0}$	$\frac{\Delta f_{\mathrm{right}}}{f_0}$	$\frac{\Delta\psi}{\pi}$	$\frac{\Delta\theta}{\pi}$	$\frac{\Delta\phi}{\pi}$	$\frac{\Delta l_{\mathrm{baseline}}}{l_{\mathrm{baseline}}}$
0.6475%	0.9908%	0.18%	0.22%	0.79%	0.31%

The conclusion inferred from the above results is that relative error less than 1%, which suggests that using LCD screen calibration method is effective.

2.2 Bayer Filter

The forward-looking HazCams of BH2 use R, G and B three-color filter camera. Each pixel only corresponds to one of RGB three-color, and R, G and B in Fig. 4 represent the red, green and blue filters, respectively, and each element of the filter corresponds to one pixel of the image sensor. The full-colours image can be obtained by applying appropriate de-mosaicing algorithm [15].

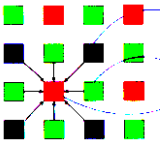


Fig. 4. The Bayer array

2.3 Image Rectification

Image rectification is also known as epipolar rectification, for corresponding epipolar parallel to horizontal scanning direction, which can eliminate of vertical disparity in stereo image, which can makes more quickly and accurately find the corresponding points, and which can improve degree of stereo matching accuracy. This paper uses the results of camera to correct image [16].

2.4 Photometric Consistency Dense Matching

Due to different physical properties and illumination condition of HazCams themselves what can cause different photometric in the images. Meanwhile, comparing to SAD pixel-based matching algorithm, ZSAD pixel-based matching algorithm has few effects of illumination, and can meet the real-time avoid obstacle requirement of BH2 by using pyramid delaminating. Thus, the paper uses ZSAD pixel-based matching algorithm to implement matching. The problem that needs to be noticed is how to set size of matching window when using ZSAD pixel-based matching algorithm? On the hand, large matching window may blur border of object, meanwhile, which could not present small object and image details in scene. On the other hand, small matching window could solve above problems problem. However, which could increase noise effect and mismatching. The computational formula as follows:

$$C_{ZSAD}(x,y,d) = \sum_{(x,y)\in W_p}^{n} \left| I_L(x,y) - I_R(x+d,y) \right| \tag{3}$$

Where left image pixel expressed by $I_L(x,y)$, right image pixel expressed by $I_R(x+d,y)$, the neighborhood window centered by (x,y) in left image by w_P, disparity of corresponding pixel can be written as d, and $d \in [d_{min}, d_{max}]$, d_{min} and d_{max} could be computed from the minimum and maximum expected distance in the pair images; $C_{ZSAD}(x,y,d)$ denotes matching cost, and the minimal $C_{ZSAD}(x,y,d)$ corresponding

disparity is ideal disparity. We could get quickly and accurate matching point by the following constraints and computational optimization technique in [17]:

Step1: Pyramid delaminating. The method is built up on the basis of multi-resolution, which is the process of coarse-to-fine. High-level pyramid image is average sampled by regions of low-level image, which is shown in Fig.5. The method can reduce the number of pixels and the disparity search range, which can only search the smaller disparity range to finish matching in high-level pyramid. For the disparity range is in direct proportion to computational complexity, which can greatly reduce matching time and the probability of mismatching in the essence. However, the pyramid layer unease too many, otherwise which can lose image detail and increase mismatching. As a result, the pyramid layer is equal to one in this paper.

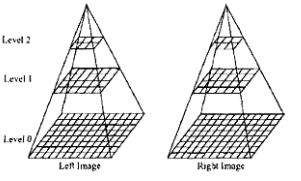


Fig. 5. The pyramid delaminating

Step2: LoG filter. Marr and Hildreth has proposed Laplace-Gaussian (Laplace of Gaussian) edge detection algorithm [18], which is LoG filter. The expression of LoG filter is shown as (4). LoG filter has the following two characteristics: (1) filtering out high frequency noise; (2) integral of LoG filter is equal to zero, which can remove the different photometric of image. In addition, Comparing to ground mobile robot, the vision system working environment of BH2 has very few obvious regular characteristics. Therefore, this paper use LoG filter to preprocess image.

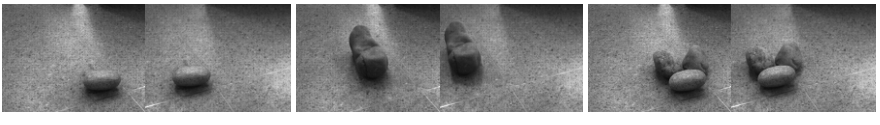
$$\nabla^2 G(x,y,\sigma)=\frac{\partial^2 G}{\partial x^2}+\frac{\partial^2 G}{\partial x^2}=\frac{1}{\pi\sigma^4}\left(\frac{x^2+y^2}{2\sigma^2}-1\right)e^{-\frac{\left(x^2+y^2\right)}{2\sigma^2}} \tag{4}$$

Step3: Ordering constraint. Stereo matching algorithm follow order constraint, if A in the left side of B for left image, then A also in the left side of B for right image, and theory of uncrossed disparity could be educed. Supposed there are $\mathrm{I}_{\mathrm{R}}(i,j)$ and $\mathrm{I}_{\mathrm{R}}(i,j+k)$ in the i-th row of right image, in which k> 0, the disparity of $\mathrm{I}_{\mathrm{R}}(i,j)$ expressed by $d_{(i,j)}$, the disparity of $\mathrm{I}_{\mathrm{R}}(i,j+k)$ expressed by $d_{(i,j+k)}$, if $\left(j+d_{(i,j)}\right)>\left(j+k+d_{(i,j+k)}\right)$, $\mathrm{I}_{\mathrm{R}}(i,j)$ and $\mathrm{I}_{\mathrm{R}}(i,j+k)$ exists crossed disparity; otherwise, does not exist.

Step4: Uniqueness constraint. Uniqueness constraint is that searching for matching point on left image, taking right image as reference, matching point with left image and reference point with right image should be consistent, which mainly detect occlusion region of stereo image pair, if $d_{\mathrm{RL}}\left(i+d_{\mathrm{LR}}(i,j),j\right)=-d_{\mathrm{LR}}(i,j)$, $\mathrm{I}_{\mathrm{R}}(i,j)$ and $\mathrm{I}_{\mathrm{R}}(i,j+k)$ is a pair of matching point; otherwise, is not a pair of matching point.

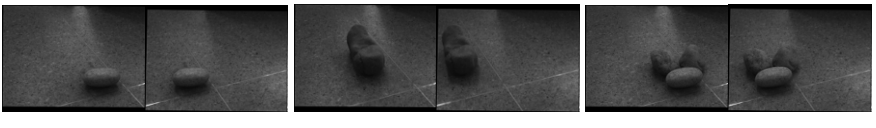
3 Algorithm Implementation and Experimental Results

Due to lack of the actual images of lunar topography and geomorphology, the paper uses simulation scene to validate vision system of HazCams. Fig.6 shows three sets of experiments results: (a) the original grabbing images (640 × 480); (b) Bayer filtered images; (c) the rectifying images (648 × 509); (d) Photometric consistency dense matching in turn is as followings: pyramid delaminating, LoG filter, ZSAD matching, ordering constraint, and uniqueness constraint in the one layer of pyramid image structure (324 × 254). ZSAD matching window size is equal to 9 × 9 in the experiment, and the disparity is in range 100 to 135. The high precision of disparity map can be obtained. The results showed that the paper proposed method well solve poor matching problems in low texture or similar texture image regions, which could accurately reconstruct outlines of the regular cobble and the irregular stone, which could exactly reconstruct locations of the obstacle. Furthermore, for validating vision system of HazCams in partial occlusion region; we placed three stones in the scene. Fig.6 shows the proposed method could exactly reconstruct locations of the obstacle with three stones. The conclusion inferred from the above results is that the proposed method is feasible and effective in this paper.



(a) Original grabbing images

(b) Bayer filtered images



(c) Rectifying images

(d) Disparity map

Fig. 6. The results of experiment

The three sets of actual experimental statistics data are shown in the Table 5: From Table 5 and Fig. 6, we could see that the paper proposed method increases has higher matching rate and less running time. The above results show that the proposed algorithm is feasible and effective.

Table 5. The Statistics Data for Experiment Result

Estimation index	Test	Test	Test Ⅲ
Mismatch%	8.7%	6.6%	9.6%
Time(second)	64.91	69.04	68.01

4 Conclusions

Firstly, this paper presents a feasible vision system of HazCams for BH2, which amplify camera calibration, Bayer filter, image rectification, pyramid delaminating, photometric consistency dense matching, and extracting disparity map. Finally, the results of experiment with actual test images show the method is effective and can adapt to obstacle detection of BH2. In addition, this paper points out the next continued research work is to apply vision system of HazCams to estimation motion of lunar rover BH2, that is, sample stereo images at different time, and the same set of 3D data perform registration for estimating the movement of BH2. The ground mobile robots use wheel encoder, inertia unit to estimate itself movement in general. However, taking into account BH2 wheels spin easily, side slip when turning; and zero drift of the inertia unit, which can produce accumulated error, which can seriously affect the results of motion estimation. Visual motion estimation increases the computational complexity [19]. However, visual motion estimation can obtain more reliable result and provide necessary information to autonomous navigation of lunar rover BH2.

References

1. Kawasaki, H., Furukawa, R.: Dense 3D reconstruction method using coplanarities and metric constraints for line laser scanning. In: 6th IEEE International Conference on 3-D Digital Imaging and Modeling (3DIM 2007), pp. 149–158. IEEE Press, Montreal (2007)
2. Matthies, L., Maimone, M., Johnson, A., et al.: Computer vision on mars. Int. J. Computer Vision 75(1), 67–92 (2007)
3. Goldberg, S.B., Maimone, M.W., Matthies, L.: Stereo vision and rove navigation software for planetary exploration. In: 2002 IEEE Aerospace Conference, vol. 5, pp. 2025–2036. IEEE Press, Montana (2002)
4. Matthies, L.: Stereo vision for planetary rovers: Stochastic modeling to near real-time implementation. Int. J. Computer Vision 8(1), 71–91 (1992)
5. Maurette, M.: Mars rover autonomous navigation. Autonomous Robots 14(2-3), 199–208 (2003)
6. Krotkov, E., Hebert, M.: Mapping and positioning for a prototype lunar rover. In: IEEE International Conference on Robotics and Automation, vol. 3, pp. 2913–2919. IEEE Press, Los Alamitos (1995)

7. Simmons, R., Krotkov, E., Chrisman, L., et al.: Experience with rover navigation for lunar-like terrains. In: Proceedings of the 1995 Conference on Intelligent Robots and Systems (IROS 1995), vol. 1, pp. 441–446. IEEE Press, Pittsburgh (1995)
8. Singh, S., Digney, B.: Autonomous cross-country navigation using stereo vision. CMU-RI-TR-99-03 (1999)
9. Vergauwen, M., Pollefeys, M., van Gool, L.: A stereo-vision system for support of planetary surface exploration. Machine Vision and Applications 14(1), 5–14 (2003)
10. Hou, J., Qi, N., Zhang, H., Liu, T.: Stereo mapping for a prototype lunar rover. In: 2006 IEEE International Conference Robotics and Biomimetics, pp. 1036–1041. IEEE Press, Kunming (2006)
11. Matthies, L., Maimone, M., Johnson, A., et al.: Mars exploration rover engineering cameras. In: SPIE Proceedings of Symposium on Remote Sensing, Toulouse (2001)
12. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Int. J. Computer Vision 47(1/2/3), 7–42 (2002)
13. Zhang, Z.: A flexible new technique for camera calibration. IEEE Trans. on Pattern Analysis and Machine Intelligence 22(11), 1330–1334 (2005)
14. Ma, S., Zhang, Z.: Computer vision-computation theory and basic algorithm, pp. 95–99. Science Press, Beijing (2003)
15. Kidono, K., Ninomiya, Y.: Visibility estimation under night-time conditions using a multiband camera. In: 2007 IEEE Intelligent Vehicles Symposium, pp. 1013–1018. IEEE Press, Istanbul (2007)
16. Fusiello, A., Trucco, E., Verri, A.: A compact algorithm for rectification of stereo pairs. Machine Vision and Applications 12(1), 16–22 (2000)
17. Chen, H.: Research on the key technology of vision-based navigation: Stereo Vision and Path Planning. PHD Thesis, Zhejiang University, Hangzhou, pp. 41–42, 56–60 (2005)
18. Marr, D., Hildreth, E.: Theory of edge detection. Proceedings of the Royal Society of London. Series B, Biological Sciences 207(1167), 187–217 (1980)
19. Cheng, Y., Mark, M., Larry, H.: Visual odometry on the mars exploration rovers. In: 2005 IEEE International Conference on Systems, Man & Cybernetics, vol. 1, pp. 903–910. IEEE Press, Hawaii (2005)

Stereovision-Based Algorithm for Obstacle Avoidance

Lazaros Nalpantidis, Ioannis Kostavelis, and Antonios Gasteratos

Robotics and Automation Lab., Production and Management Engineering Dept.,
Democritus University of Thrace,
University Campus, Kimmeria, GR-671 00 Xanthi, Greece
{lanalpa,ik3339,agaster}@pme.duth.gr
http://robotics.pme.duth.gr

Abstract. This work presents a vision-based obstacle avoidance algorithm for autonomous mobile robots. It provides an efficient solution that uses a minimum of sensors and avoids, as much as possible, computationally complex processes. The only sensor required is a stereo camera. The proposed algorithm consists of two building blocks. The first one is a stereo algorithm, able to provide reliable depth maps of the scenery in frame rates suitable for a robot to move autonomously. The second building block is a decision making algorithm that analyzes the depth maps and deduces the most appropriate direction for the robot to avoid any existing obstacles. The proposed methodology has been tested on sequences of self-captured outdoor images and its results have been evaluated. The performance of the algorithm is presented and discussed.

Keywords: Stereo vision, obstacle avoidance, autonomous robot navigation.

1 Introduction

In this work a vision-based obstacle avoidance algorithm is presented. It is intended to be used in autonomous mobile robotics. However, the development of an efficient, solely vision-based method for mobile robot navigation is still an active research topic. Towards this direction, the first step is to avoid any obstacles through vision. However, systems placed on robots have to conform to the restrictions imposed by them. Autonomous robot navigation requires almost real-time frame rates from the responsible algorithms. Furthermore, computing resources are strictly limited onboard a robot. Thus, the omission of popular obstacle detection techniques such as the v-disparity, which require Hough-transformations, would be highly appreciated. Instead, simple and efficient solutions are demanded.

In order to achieve reliable obstacle avoidance, many popular methods involve the use of an artificial stereo vision system, due to its biomimetic nature. Stereoscopic vision can be used in order to obtain the 3D position of the depicted objects from two simultaneously captured, slightly misplaced views of a scene. Mobile robots can take advantage of stereo vision systems as a reliable method to extract information about their environment [1]. Although stereo vision provides

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 195–204, 2009.

an enormous amount of information, most of the mobile robot navigation systems use complementary sensors in order to navigate safely [2]. The use of lasers, projectors and various other range finders is a commonplace. The goal of this work is to develop a real-time obstacle avoidance algorithm based only on a stereo camera, for autonomous mobile robots. The core of the presented approach can be divided into two separate and independent algorithms:

- *The stereo vision algorithm.* It retrieves information about the environment from a stereo camera and produces a depth image, i.e. disparity map, of the scenery.
- *The decision making algorithm.* It analyzes the data of the previous algorithm and decides the best direction, i.e. forward, right or left, for the robot to move in order to avoid any existing obstacles.

Both the algorithms have been modularly implemented in C++. The modularity of the system allows the easy modification, easy debugging and ensures the adaptability of the overall algorithm. Figure 1 presents the flow chart of the implemented algorithm. Besides the decision made, the implemented system stores input images and calculated disparity maps for possible later offline use.

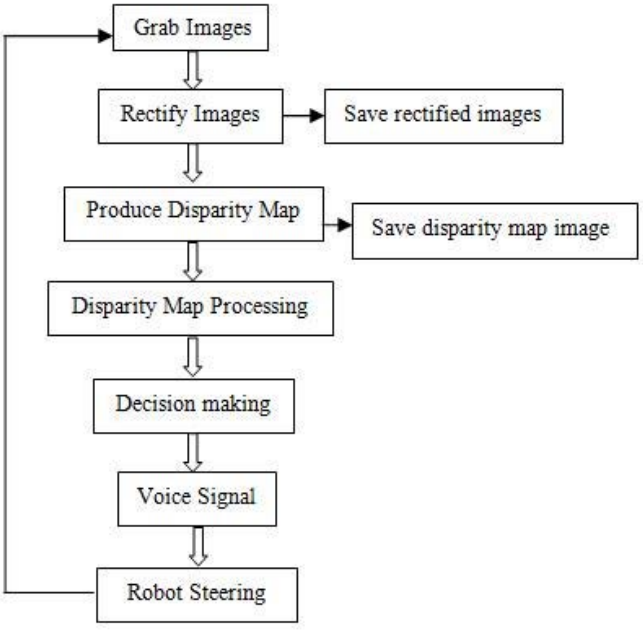


Fig. 1. Flow chart of the implemented obstacle avoidance algorithm

2 Related Work

Autonomous robots' behavior greatly depends on the accuracy of their decision making algorithms. In the case of stereo vision-based navigation, the accuracy

and the refresh rate of the computed disparity maps are the cornerstone of its success [3]. Dense local stereo correspondence methods calculate depth for almost every pixel of the scenery, talking into consideration only a small neighborhood of pixels each time [4]. On the other hand, global methods are significantly more accurate but at the same time more computationally demanding, as they account for the whole image [5]. However, since the most urgent constraint in autonomous robotics is the real-time operation, such applications usually utilize local algorithms. Muhlmann et al. in [6] describe a local method that uses the sum of absolute differences (SAD) correlation measure for RGB color images. Applying a left to right consistency check, the uniqueness constraint and a median filter, it can achieve 20 fps. Another fast Slocal AD based algorithm is presented in [7]. It is based on the uniqueness constraint and rejects previous matches as soon as better ones are detected. It achieves 39.59 fps. The algorithm reported in [8] achieves almost real-time performance. It is once more based on SAD but the correlation window size is adaptively chosen for each region of the picture. Apart from that, a left to right consistency check and a median filter are utilized.

Another possibility in order to obtain accurate results in real-time is to utilize programmable graphic processing units (GPU). In [9] a hierarchical disparity estimation algorithm is presented. On the other hand, an interesting but very computationally demanding local method is presented in [10]. It uses varying weights for the pixels in a given support window, based on their color similarity and geometric proximity. However, the execution speed of the algorithm is far from being real-time. A detailed taxonomy and presentation of dense stereo correspondence algorithms can be found in [4]. Additionally, the recent advances in the field as well as the aspect of hardware implementable stereo algorithms are covered in [11].

In the relevant literature, a wide range of sensors and various methods have been proposed in order to detect and avoid obstacles. Some interesting details about the developed sensor systems and proposed detection and avoidance algorithms can be found in [12] and [13]. The obstacle avoidance sensor systems found in literature can generally be divided into two main categories. The first category involves the use of ultrasonic sensors. They are simple to implement and can detect obstacles reliably. On the other hand, the second category involves vision-based sensor systems. This category can be further divided into the stereo vision systems (which is applied to the detection of objects in 3D) and the laser range sensors (which can be applied to the detection of obstacles both in 2D and 3D, but can barely used for real-time detection [14]). As far as the stereo vision systems are concerned, one of the most popular methods for obstacle avoidance is the estimation of the so called v-disparity image. This method requires plenty of complex calculations and is applied in order to confront the noise in low quality disparity images [15,16,17]. However, if detailed and noise-free disparity maps were available, less complicated methods could have been used instead. Considering the above as a background, the contribution of this work is the development of an algorithm for obstacle avoidance with the sole use of a stereoscopic camera. The use of only one sensor and specially

of a stereoscopic camera diminish the complexity of our system and can also be easily integrated and interact with with other stereo vision tasks such as object recognition and tracking.

3 Stereo Vision Algorithm

Contrary to most of the stereo algorithms, which directly use the camera's images, the proposed stereo algorithm uses an enhanced version of the captured images as input. The initially captured images are processed in order to extract the edges in the depicted scene. The utilized edge detecting method is the Laplacian of Gaussian (LoG), using a zero threshold. This choice produces the maximum possible edges. The LoG edge detection method smoothens the initial images with a Gaussian filter in order to suppress any possible noise. Then a Laplacian kernel is applied that marks regions of significant intensity change. Actually, the combined LoG filter, with standard deviation equal to 2, is applied at once and the zero crossings are found. The extracted edges are, afterwards, superimposed to the initial images. The steps of the aforementioned process are shown in Fig. 2. The outcome of this procedure is a new version of the original images having more striking features and textured surfaces, which facilitate the following stereo matching procedure.

The depth maps are computed using a three-stage local stereo correspondence algorithm. The utilized stereo algorithm combines low computational complexity with sophisticated data processing. Consequently, it is able to produce dense disparity maps of good quality in frame rates suitable for robotic applications. The main attribute that differentiates this algorithm from the majority of the other ones is that the matching cost aggregation step consists of a sophisticated gaussian-weighted rather than a simple summation. Furthermore, the disparity selection step is a simple winner-takes-all (WTA) choice, as the absence of any iteratively updated selection process significantly reduces the computational payload of the overall algorithm. Finally, any dedicated refinement step is also absent for speed reasons.

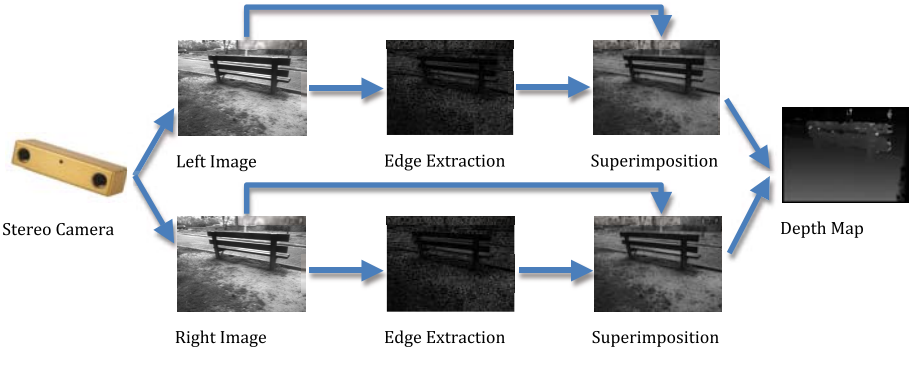


Fig. 2. Image enhancement steps of the proposed stereo algorithm

The matching cost function utilized is the truncated absolute differences (AD). AD is inherently the simplest metric of all, involving only summations and finding absolute values. The AD are truncated if they excess the 4% of the maximum intensity value. Truncation suppresses the influence of noise in the final result. This is very important for stereo algorithms that are intended to be applied to outdoor scenes. Outdoor pairs usually suffer from noise induced by a variety of reasons, e.g. lighting differences and reflections. For every pixel of the reference (left) image, AD are calculated for each of its candidate matches in the other (right) image.

The computed matching costs for every pixel and for all its potential disparity values comprise a 3D matrix, usually called as disparity space image (DSI). The DSI values for constant disparity value are aggregated inside fix-sized square windows. The dimensions of the chosen aggregation window play an important role in the quality of the final result. Generally, small dimensions preserve details but suffer from noise. On the contrast, large dimensions may not preserve fine details but significantly suppress the noise. This behavior is highly appreciated in outdoor robotic applications where noise is a major factor, as already discussed. The aggregation windows dimensions used in the proposed algorithm are 19x19 pixels. This choice is a compromise between real-time execution speed and noise cancelation. The AD aggregation step of the proposed algorithm is a weighted summation. Each pixel is assigned a weight depending on its Euclidean distance from the central pixel. A 2D Gaussian function determines the weights value for each pixel. The center of the function coincides with the central pixel. The standard deviation is equal to the one third of the distance from the central pixel to the nearest window-border. The applied weighting function can be calculated once and then be applied to all the aggregation windows without any further change. Thus, the computational load of this procedure is kept within reasonable limits.

Finally, the optimum disparity value for each pixel, i.e. the disparity map, is chosen by a simple, non-iterative WTA step. In the resulting disparity maps, smaller values indicate more distant objects, while bigger disparity values indicate objects lying closer.

4 Decision Making Algorithm

The previously calculated disparity map is used to extract useful information about the navigation of a robot. Contrary to many implementations that involve complex calculations upon the disparity map, the proposed decision making algorithm involves only simple summations and checks. This is feasible due to the absence of significant noise in the produced disparity map. The goal of the developed algorithm is to detect any existing obstacles in front of the robot and to safely avoid it, by steering the robot left, right or to moving it forward.

In order to achieve that, the developed method divides the disparity map into three windows, as in Fig. 3.

In the central window, the pixels p whose disparity value $D(p)$ is greater than a defined threshold value T are enumerated. Then, the enumeration result is

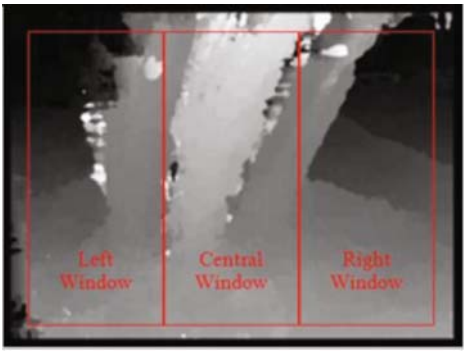


Fig. 3. Depth map's division in three windows

examined. If it is smaller than a predefined rate r of all the central windows pixels, this means that there are no obstacles detected exactly in front of the robot and in close distance, and thus the robot can move forward. On the other hand, if this enumeration's result exceeds the predefined rate, the algorithm examines the other two windows and chooses the one with the smaller average disparity value. In this way the window with the fewer obstacles will be selected. The pseudocode of the implemented simple decision making algorithm follows:

Decision Making Pseudocode

```
for all the pixels p of the central window {
  if D(p) > T {
    counter++ }
  numC++ }
if counter <  r% of numC {
  GO STRAIGHT }
else {
  for all the pixels p of the left window {
    sumL =+ D(p)
    numL++ }
  for all the pixels p of the right window {
    sumR =+ D(p)
    numR++ } }
  avL = sumL / numL
  avR = sumR / sumR
  if avL <avR {
    GO LEFT }
  else {
    GO RIGHT } }
```

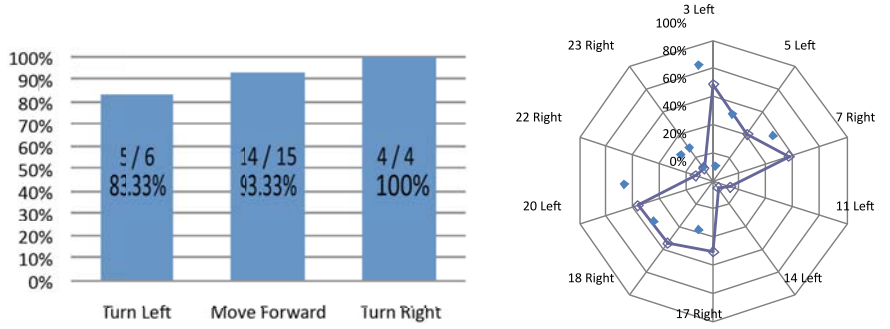
The values of the parameters T and r play an important role to the algorithm's behavior. Small values of T in conjunction with small values of r favor the

hesitancy in moving forward, ensuring obstacle avoidance but at the same time being susceptible to false alarms due to noise. On the other hand, the opposite scenario is less susceptible to false alarms but may be proven risky. The values of T and r were empirically defined as a compromise between the two aforementioned behavioral patterns. The decided values were $T = 120$ and $r = 20\%$. In the example of Fig. 3 the number of pixels in the central window whose disparity value is greater than 120 is 40160 which is more than the 20% of all the window's pixels (i.e. 8160). Consequently, the robot is not allowed to move forward and has to examine the other two possibilities. The robot will decide to turn left as the left window has smaller average disparity values than the right one.

5 Experimental Validation

The performance of the examined methodology for obstacle avoidance has been examined by applying it on a sequence of 25 stereo image pairs. The experiments took place in outdoor scenes and the obstacles were natural elements like trees, bushes, benches, pavements and other objects found in the captured scenes. Various single image pairs, as well as routes were captured. That is, the algorithm's decision concerning a test pair, dictates the next pairs' capture position. A series of such linked pairs comprise a route. All the 25 test image pairs are available to be freely downloaded from [18]. Figure 4 presents a sample route comprising of 7 image pairs. For each image pair the reference image is shown on the left, the calculated disparity map in the middle, and the direction decided by the algorithm is shown on the right. The decision for each image pair dictated the camera's next position, as easily deduced by the sequence of Fig. 4.

Fig. 4. A sample outdoor route and the algorithm's outputs



(a) Percentage of the algorithm's correct decision

(b) Percentage of certainty for the algorithm's decisions

Fig. 5. Results of the proposed algorithm

The algorithm's results, when applied on the 25 test image pairs, are summarized in Fig. 5. Figure 5(a) presents the percentage of the algorithm's correct decisions for each possible direction. A correct decision is not only one that avoids collision with an existing obstacle, but at the same time choses the most preferable direction. As shown, the algorithm is more than 83% of the times correct, for any direction. Actually the algorithm has been correct in 23 out of the 25 cases, i.e. 92% of the times.

On the other hand, the certainty of each decision is of interest as well. Due to the nature of the proposed algorithm, as presented in the previous sections, the decision about moving forward is based on some parameters' values and is totally independent from the other two possible decisions. Consequently, a measure of certainty would be meaningful only in the cases of the left or right decisions. As these two decisions are based on the same heuristic, they can be directly compared. The certainty *cert* of a direction's decision which yields an average disparity avD_1 over the other direction which yields $avD_2 > avD_1$ is calculated as:

$$cert = \frac{avD_2 - avD_1}{avD_2} \tag{1}$$

The results for the left and right decisions of the algorithm are shown in Fig. 5(b). For each decision the pair's indicating number as well as the algorithm's decision is given. The certainty ranges form 0% for no certainty at all, to 100% for absolute certainty. The bigger the area defined by the resulting points, the bigger the algorithm's overall certainty. However, big values of certainty are not always achievable. In the extreme case when both the left and the right direction are fully traversable, the certainty measure would become 0%. Despite this fact, the certainty is useful. Observing the correlation between false decisions and certainty values, a threshold could be decided, below which the algorithm should reconsider its decision.

6 Conclusion

A vision-based obstacle avoidance algorithm for autonomous mobile robots was presented. The proposed algorithm requires only one sensor, i.e. a stereo camera, and a low amount of involved computations. The algorithm's structure consists of a specially developed and optimized stereo algorithm that produces noise-free depth maps, and a computationally simple decision making algorithm. The decision making algorithm avoids complex calculations and transformations. Consider as an example, the case of the popular v-disparity implementation where Hough-transformation is needed in order to compensate for the low quality disparity maps. On the other hand, simpler than the proposed direction deciding algorithms fail to yield correct results. In this case, consider an algorithm where the three windows of Fig. 3 are treated equally and the smallest average disparity is sought. This methodology is doomed to fail in the case, among many others, where only a thin obstacle is close to the robot and other obstacles are in medium range. Such a naive algorithm would chose the direction towards the close thin obstacle, avoiding the medium ranged obstacles.

The proposed methodology has been tested on sequences of self-captured outdoor images and its results have been evaluated. Its performance has been presented and discussed. The proposed algorithm managed to avoid the obstacles successfully in the vast majority of the tested image pairs. Despite its simple calculations, both during the disparity map generation and the decision making, the algorithm exhibited promising behavior. The simple structure and the absence of heavy computational payload are characteristics highly desirable in autonomous robotics. The real-time collision-free navigation of autonomous robotic platforms is the first step towards the accomplishment of more complex activities, e.g. path planning and mapping of an area. Consequently, the proposed algorithm is suitable for autonomous robotic applications and is able to provide real-time obstacle avoidance behavior, based solely on a stereo camera.

Acknowledgments. This work was supported by the E.C. funded research project for vision and chemiresistor equipped web-connected finding robots, "View-Finder", FP6-IST-2005-045541.

References

1. Iocchi, L., Konolige, K.: A multiresolution stereo vision system for mobile robots. In: Italian AI Association Workshop on New Trends in Robotics Research (1998)
2. Siegwart, R., Nourbakhsh, I.R.: Introduction to Autonomous Mobile Robots. MIT Press, Massachusetts (2004)
3. Schreer, O.: Stereo vision-based navigation in unknown indoor environment. In: 5th European Conference on Computer Vision, vol. 1, pp. 203–217 (1998)
4. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International Journal of Computer Vision 47(1-3), 7–42 (2002)
5. Torra, P.H.S., Criminisi, A.: Dense stereo using pivoted dynamic programming. Image and Vision Computing 22(10), 795–806 (2004)

6. Muhlmann, K., Maier, D., Hesser, J., Manner, R.: Calculating dense disparity maps from color stereo images, an efficient implementation. International Journal of Computer Vision 47(1-3), 79–88 (2002)
7. Di Stefano, L., Marchionni, M., Mattoccia, S.: A fast area-based stereo matching algorithm. Image and Vision Computing 22(12), 983–1005 (2004)
8. Yoon, S., Park, S.K., Kang, S., Kwak, Y.K.: Fast correlation-based stereo matching with the reduction of systematic errors. Pattern Recognition Letters 26(14), 2221–2231 (2005)
9. Zach, C., Karner, K., Bischof, H.: Hierarchical disparity estimation with programmable 3d hardware. In: International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision, pp. 275–282 (2004)
10. Yoon, K.J., Kweon, I.S.: Adaptive support-weight approach for correspondence search. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(4), 650–656 (2006)
11. Nalpantidis, L., Sirakoulis, G.C., Gasteratos, A.: Review of stereo vision algorithms: from software to hardware. International Journal of Optomechatronics 2(4), 435–462 (2008)
12. Borenstein, J., Koren, Y.: Real-time obstacle avoidance for fast mobile robots in cluttered environments. IEEE Transactions on Systems, Man, and Cybernetics 19(5), 1179–1187 (1990)
13. Ohya, A., Kosaka, A., Kak, A.: Vision-based navigation of mobile robot with obstacle avoidance by single camera vision and ultrasonic sensing. IEEE Transactions on Robotics and Automation 14(6), 969–978 (1998)
14. Vandorpe, J., Van Brussel, H., Xu, H.: Exact dynamic map building for a mobile robot using geometrical primitives produced by a 2d range finder. In: IEEE International Conference on Robotics and Automation, Minneapolis, USA, pp. 901–908 (1996)
15. Labayrade, R., Aubert, D., Tarel, J.P.: Real time obstacle detection in stereovision on non flat road geometry through "v-disparity" representation. In: IEEE Intelligent Vehicle Symposium, Versailles, France, vol. 2, pp. 646–651 (2002)
16. Zhao, J., Katupitiya, J., Ward, J.: Global correlation based ground plane estimation using v-disparity image. In: IEEE International Conference on Robotics and Automation, Rome, Italy, pp. 529–534 (2007)
17. Soquet, N., Aubert, D., Hautiere, N.: Road segmentation supervised by an extended v-disparity algorithm for autonomous navigation. In: IEEE Intelligent Vehicles Symposium, Istanbul, Turkey, pp. 160–165 (2007)
18. Nalpantidis, L., Kostavelis, I.: Group of Robotics and Cognitive Systems (2009), `http://robotics.pme.duth.gr/reposit/stereoroutes.zip`

Improved Techniques for the Rao-Blackwellized Particle Filters SLAM

Huan Wang, Hongyun Liu, Hehua Ju, and Xiuzhi Li

Beijing University of Technololgy,
100124, Beijing, China
wanghuan_6537@163.com

Abstract. Rao-Blackwellized particle filters simultaneous localization and mapping can yield effective results but it has the tendency to become inconsistent. To ensure consistency, a methodology of an unscented Kalman filter and Markov Chain Monte Carlo resampling are incorporated. More accurate nonlinear mean and variance of the proposal distribution are obtained without the linearization procedure in extended Kalman filter. Furthermore, the particle impoverishment induced by resampling is averted after the resample move step. Thus particles are less susceptible to degeneracies. The algorithms are evaluated on accuracy and consistency using computer simulation. Experimental results illustrate the advantages of our methods over previous approaches.

Keywords: Simultaneous localization and mapping (SLAM), Rao-Blackwellized particle filters (RBPF), unscented Kalman filter (UKF), Markov Chain Monte Carlo (MCMC), consistency.

1 Introduction

Simultaneous Localization and Mapping (SLAM) is the process by which a robot can build a map of the environment and concurrently locate itself with respect to the map. Historically the earliest probabilistic SLAM algorithm was introduced by Smith et al.[1]. A fully correlated posterior estimate based on extended Kalman filter (EKF) has the weakness of computational complexity, nonlinearity and data association. In large-scale environments, it is difficult to avoid inconsistency [2].

Rao-Blackwellized particle filter (RBPF) was introduced by Murphy [3] as an efficient solution to the SLAM problem. Montemerlo et al.[4] subsequently extended this frame. The approach uses a particle filter in which each particle carries an individual map of features. The posterior over robot pose states are represented by particles, and each individual feature is estimated using a separate low-dimensional EKF. RBPF SLAM is also called FastSLAM due to its computational advantages over the plain EKF SLAM. An alternative approach is based on expectation maximization algorithm to address the correspondence problem. Thrun has employed it in occupancy grid maps learning with forward sensor models [5] .

The conventional RBPF SLAM has some drawbacks that lead to filter divergence. First, the classical extended Kalman filter is employed in RBPF SLAM to compute

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 205–214, 2009.

the proposal distribution. Due to linearization errors, large errors in the true posterior mean and covariance is introduced. Second, at the particle resampling step, a few highly weighted samples are copied many times. Since resampling is repeated frequently, the pose history and map estimate associated with each particle tend to become identical. Due to the above two reasons the filter insufficiently encodes uncertainty in the estimation. It is revealed in [6] that RBPF SLAM can produce accurate results but will always underestimate its uncertainty in the long-term. Several researches on consistency improvement can be found in [7, 8].

In this paper we describe a new sampling strategy for RBPF SLAM which improves the consistency of the estimation and the diversity of the samples. The approach uses the unscented Kalman filter (UKF) to compute the proposal distribution, which averts the linearization error. A Markov Chain Monte Carlo (MCMC) move step is incorporated to combat the particle depletion after resampling.

The remaining of the article is structured as follows. In the next section, we review the original particle filtering SLAM briefly and analyze the consistency issue. In Section 3, we explain the UKF and MCMC resampling detailing the combined idea of the two algorithm. Section 4 provides an illustrative evaluation example for comparison with the previous approaches.

2 Problem Statement

2.1 RBPF SLAM

From a probabilistic perspective, SLAM involves estimating the posterior over the entire path along with the map concurrently.

$$p(x_{1:t}, M \mid z_{1:t}, u_{1:t}) \ . \tag{1}$$

Here $x_t = [x_{v,t}, y_{v,t}, \phi_{v,t}]$ is the pose at time t, M is the map, $z_{1:t} = [z_1, ..., z_t]^T$ and $u_{1:t} = [u_1, ..., u_t]$ are the measurements and controls respectively.

It is feasible to reduce the sample-space by applying the Rao-Blackwellization, whereby a joint state is factored into a vehicle component and a conditional map component.

$$p(x_{1:t}, \mathrm{M} \mid z_{1:t}, u_{1:t}, c_{1:t}) = p(x_{1:t} \mid z_{1:t}, u_{1:t}, c_{1:t}) \prod_{n=1}^{N} p(m_n \mid x_{1:t}, z_{1:t}, c_{1:t}) \, . \tag{2}$$

Given the robot path, landmarks are conditionally independent, and each is estimated by a small constant-size extended Kalman filter (EKF). $c_{1:t}$ are the correspondences of measurements and landmarks in the map.

The basic idea is to employ sequential importance sampling with resampling (SISR). The current generation of samples is drawn from the last generation, by sampling from a proposal distribution. i.e.:

$$x_t^{(i)} \sim \pi(x_t \mid x_{t-1}^{(i)}, u_t, z_t) \, . \tag{3}$$

Next, FastSLAM updates the posterior over the feature estimates using EKF. According to the measurement likelihood [8, 9], an individual importance weight $w_t^{(i)}$ is assigned to each particle.

$$w_t^{(i)} \propto w_{t-1}^{(i)} \frac{p(z_t \mid x_t^{(i)}) p(x_t^{(i)} \mid x_{t-1}^{(i)}, u_t)}{q(x_t^{(i)} \mid x_{1:t-1}^{(i)}, z_{1:t}, u_{1:t})} \quad . \tag{4}$$

In a final step, particles are resampled according to their weights in order to give more presence to highly-weighted samples.

2.2 Consistency Analysis

Consistency means unbiased estimate and estimate variance can be acquired. For an available ground truth x_t and an estimated mean and covariance $\{\hat{x}_t, P_t\}$, we can use the normalized estimation error squared (NEES) to carry out the consistency test.

$$\varepsilon_{\mathrm{t}} \quad x_t - \hat{x}_t \ ^{\top} P_t^{-1} (x_t - \hat{x}_t) \quad . \tag{5}$$

For the 3-dimensional vehicle pose, a consistent filter over 50 Monte Carlo runs should have an NEES less than 3.72 with 95% probability. Details about measuring consistency of a filter via the average NEES can be found in [6].

In the implementation of the algorithm, EKF is used in the prediction step, which can be viewed as providing "first-order" approximations to the optimal terms. The linearization, however, can introduce large errors in the true posterior mean and covariance of the transformed random variable, which can lead to particle degeneration. In this case, resampling is in need.

Whenever resampling is performed, for each particle replaced, an entire pose history and map hypothesis is lost for ever. This depletes the number of samples representing past poses and consequently erodes the statistics of the landmark estimates conditioned on these past poses. Such phenomena become worse as time goes on. Since particle filtering is always susceptible to particle impoverishment (i.e. the number of distinct particles decrease) the covariance P of samples will decrease. Thus the NEES will increase.

In a nutshell, filter inconsistency can be induced by particle degeneracies. On one hand, linear error in proposal distribution produce particles with skewed weights, on the other, samples tend to become degenerate after repeated resampling.

3 Improvement on RBPF SLAM

Particle degeneracy can have negative effect on estimation consistency, and the loss of sample diversity induced by repetitive resampling is also an essential reason for inconsistency. First, improvement is made on proposal distribution to produce tighter mean and covariance estimates. The unscented Kalman filter is utilized. Furthermore, a MCMC move step is incorporated after resampling to recover the diversity of samples.

3.1 Improved Proposal Distribution

As described in Section 2, one needs to draw particles from a proposal distribution π in the prediction step. The optimal proposal distribution takes not only the control command but also the measurement z_t into account. According to Doucet [9], the distribution can be produced by the EKF. Due to the high nonlinear nature of the vehicle and sensor models, the effects of linearization required in this approach will yield inaccurate proposal distribution.

One deterministic approach for the nonlinear propagation of means and covariances is the UKF algorithm. The state distribution in UKF is specified using a minimal set of carefully chosen samples (sigma points $\chi^{[i]}$ and weights) $\{\chi^{[i]}\} = [\hat{\chi}\ \hat{\chi} \pm (\sqrt{(n+\lambda)P})_i]$. The nonlinear model is applied to each point, and the weighted statistics of the transformed points capture the posterior mean and covariance accurately to the 3rd order for any nonlinearity [10]. Its efficient implementation lies on the elimination of the derivation and evaluation of Jacobian matrix. How to choose the sigma points and their weights can be found in [11]. The computational complexity of the UKF is the same order as that of the EKF [12] .

In the application to the SLAM problem, the sigma points are first passed through the motion model to predict a prior mean and covariance. When observation is made the UKF can be further used to update the first and second order statistics of the distribution. More accurate estimation of the proposal distribution is achieved. Thus, degeneracies in the weights of particles are much less likely to occur.

3.2 MCMC Methods

The MCMC methods can be used to increase the diversity of particles. Assume that particles are distributed according to a posterior $p(x^*_{0:t} \mid z_{1:t})$, then applying a Markov chain transition kernel with invariant distribution, still results in a set of samples distributed according to the posterior of interest $p(x^*_{0:t} \mid z_{1:t})$. The transition satisfies: $\int K(x_{0:t} \mid x^*_{0:t}) p(x^*_{0:t} \mid z_{1:t}) = p(x_{0:t} \mid z_{1:t})$. However, the new particles might have been moved to more interesting areas of the state-space. It is essential to construct the kernel. The standard methods include Gibbs sampler and Metropolis Hastings (MH) algorithms [13, 14].

We use the MH algorithms on each particle after resampling as follows:

For $i = 1$ to N

1) Sample the proposal candidate $\tilde{x}_t^{(i)} \sim p(x_t \mid x_{t-1})$;
2) Draw u from the interval [0, 1] according to uniform distribution;
3) If $u < \min\{1, p(z_t \mid \tilde{x}_t^{(i)}) / p(z_t \mid x_t^{*(i)})\}$, then accept the move: $x_t^{(i)} = \tilde{x}_t^{(i)}$, else $x_t^{(i)} = x_t^{*(i)}$.

3.3 Unscented MCMCPF SLAM

The Unscented MCMC particle filter (UMCMCPF) is done by performing the following steps:

We assume that, at time $t-1$, the joint state is represented by $\{w_{t-1}^{(i)}, x_{0:t-1}^{(i)}, p(M|x_{0:t-1}^{(i)}, z_{0:t-1})\}_{i}^{N}$.

1) Generate augmented mean and covariance sigma points

$$\hat{x}_{t-1}^{(i)a} = (x_{t-1}^{(i)T} \ v_t \ r_t)^T \ . \tag{6}$$

$$P_{t-1}^{(i)a} = \begin{pmatrix} P_{t-1}^{(i)} & 0 & 0 \\ 0 & Q_t & 0 \\ 0 & 0 & R_t \end{pmatrix} . \tag{7}$$

where v_t is the process noise, which has the covariance of Q_t r_t is the measurement noise with covariance of R_t .

2) Generate sigma points

$$\chi_{t-1}^{\alpha(i)} = [\hat{x}_{t-1}^{\alpha(i)} \ \hat{x}_{t-1}^{\alpha(i)} + \sqrt{(L+\lambda)P_{t-1}^{\alpha(i)}} \ \hat{x}_{t-1}^{\alpha(i)} - \sqrt{(L+\lambda)P_{t-1}^{\alpha(i)}}] \ . \tag{8}$$

3) Pass sigma points through motion model and compute Gaussian statistics

$$\chi_{t|t-1}^{x(i)} = f(\chi_{t-1,j}^{x(i)}, \chi_{t-1,j}^{v(i)}, u_t) \ . \tag{9}$$

$$\hat{x}_{t|t-1}^{(i)} = \sum_{j=0}^{2L} w_j^{(m)} \chi_{j,t|t-1}^{x(i)} \ . \tag{10}$$

$$P_{t|t-1}^{(i)} = \sum_{j=0}^{2L} w_j^{(c)} (\chi_{t|t-1,j}^{x(i)} - \hat{x}_{t|t-1}^{(i)})(\chi_{t|t-1,j}^{x(i)} - \hat{x}_{t|t-1}^{(i)})^{\mathrm{T}} \ . \tag{11}$$

4) Predict observations at sigma points and compute Gaussian statistics

$$Z_{t|t-1}^{(i)} = h(\chi_{t|t-1}^{(i)x}, \chi_{t|t-1}^{(i)n}) . \tag{12}$$

$$\hat{z}_{t|t-1}^{(i)} = \sum_{j=0}^{2L} w_j^{(m)} Z_{t|t-1,j}^{(i)} \ . \tag{13}$$

5) Update mean and covariance

$$P_{\hat{z}_k \hat{z}_k} = \sum_{j=0}^{2L} w_j^{(c)} (Z_{t|t-1,j}^{(i)x} - \hat{z}_{t|t-1}^{(i)})(Z_{t|t-1,j}^{(i)} - \hat{z}_{t|t-1}^{(i)})^T \ . \tag{14}$$

$$P_{x_k z_k} = \sum_{j=0}^{2L} w_j^{(c)} (\chi_{t|t-1,j}^{(i)} - \hat{x}_{t|t-1}^{(i)})(\chi_{t|t-1,j}^{(i)} - \hat{x}_{t|t-1}^{(i)})^T \ . \tag{15}$$

$$K_t = P_{x_t z_t} P_{\hat{z}_t \hat{z}_t}^{-1} \ . \tag{16}$$

$$\hat{x}_t^{(i)} = \hat{x}_{t|t-1}^{(i)} + K_t (z_t - \hat{z}_{t|t-1}^{(i)}) \ . \tag{17}$$

$$\hat{P}_t^{(i)} = P_{t|t-1}^{(i)} - K_t P_{\hat{z}_t \hat{z}_t} K_t^T \quad . \tag{18}$$

6) Sample from the proposal

$$x_t^{(i)} \sim \sum_{j=1}^{N} w_{t-1}^{(j)} q(x_t^{(i)} \mid z_t, x_{t-1}^{(j)}) \quad . \tag{19}$$

where $q(x_t^{(i)} \mid z_t, x_{t-1}^{(i)}) \sim N(\hat{x}_t^{(i)}, P_t^{(i)})$.

7) Measurement update

For each observed feature, update the mean and covariance according to the drawn pose and the observation.

8) Importance weight: Calculate the importance weight for the new particle.

9) Resampling

Our algorithm follows the approach proposed by Doucet et al. [9] to calculate the effective number of particles N_{eff} to evaluate how well the current particle set represents the true posterior. This quantity is computed as

$$N_{eff} = \frac{1}{\sum_{i=1}^{N} (w_t^{(i)})^2} \quad . \tag{20}$$

The worse the degeneracy the lower the N_{eff} will be. This adaptive resampling reduces the frequency at which resampling takes place. Stratified resampling is implemented when N_{eff} drops below the threshold of $75\%N$.

10) MCMC move

MH sampling is implemented and an equal weight of $1/N$ is assigned to each particles.

11) State estimation

$$x_t = \sum_{i=1}^{N} x_t^{(i)} w_t^{(i)} \quad . \tag{21}$$

4 Simulation Experiments

4.1 Motion Model and Measurement Model

The robot motion model is depicted in Figure 1. The robot state is represented by 3 variables as $X_v = [x_v, y_v, \phi_v]$. The motion model determines the location state of the robot and the state of landmarks. The control set provides an estimate of the velocities $u = (v, \alpha)$, including a linear velocity and a rotational velocity with white Gaussian noise. The state of the robot can be predicted via the odometry motion model:

$$X_{v,t} = \mathrm{f}(X_{\mathrm{v,t\ 1}}, \mathrm{u_t}) = \begin{bmatrix} x_{v,t\ 1} \\ y_{v,t\ 1} \\ \phi_{v,t\ 1} \end{bmatrix} + \begin{bmatrix} \mathrm{v_{t\text{-}1}} \cdot \cos(\phi_{vt\ 1} + \alpha_{\mathrm{t}\ 1}) \\ \mathrm{v_{t\text{-}1}} \cdot \sin(\phi_{vt\ 1} + \alpha_{\mathrm{t}\ 1}) \\ \dfrac{\mathrm{v_{t\text{-}1}} \cdot \sin(\alpha_{t\ 1})}{\mathrm{L}} \end{bmatrix} \cdot \Delta \mathrm{t} + \begin{bmatrix} w_x \\ w_y \\ w_\phi \end{bmatrix} . \tag{22}$$

Here Δt is the fixed duration to carry out the control, L is the wheelbase of the vehicle and α is the steering angle.

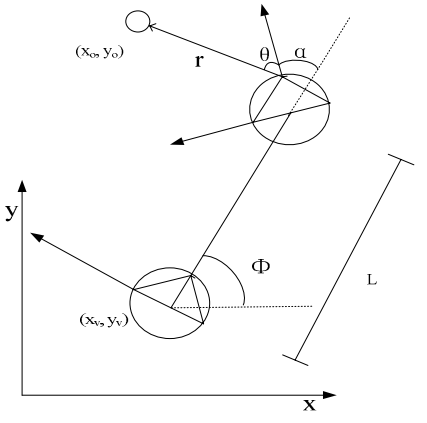


Fig. 1. Vehicle motion model

Under the hypothesis that the environment is static, the feature will stay unchanged, the state of which is as follows:

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} x_{t\ 1} \\ y_{t\ 1} \end{bmatrix} . \tag{23}$$

The robot commonly observes the range r and bearing θ to nearby landmarks using a laser scanner. Assuming that the current landmark position stored in the map is written as (x_o, y_o), the vehicle pose is described as (x_v, y_v, ϕ_v), and the measurement noise in range and bearing is represented by $\varepsilon = (\varepsilon_r, \varepsilon_\theta)^T$. The measurement function can be written as the following matrix function:

$$z = \begin{bmatrix} r \\ \theta \end{bmatrix} = h(x_o) + \varepsilon_t = \begin{bmatrix} \sqrt{(x_o - x_v)^2 + (y_o - y_v)^2} \\ \tan^{-1}\left(\dfrac{y_o - y_v}{x_0 - x_v}\right) - \phi_v \end{bmatrix} + \begin{bmatrix} \varepsilon_r \\ \varepsilon_\theta \end{bmatrix} . \tag{24}$$

4.2 Simulation Results

In the simulation experiments, the vehicle velocity is $v = 3m/s$, the control noise is $\sigma_v = 0.3m/s$, $\sigma_\gamma = 3°$ .The observation model used $\sigma_r = 0.1m$, $\sigma_\theta = 1°$. Controls

are updated at 40 Hz and observation scans are obtained every 5 Hz. Experiments were performed in the simulated 250m×200m environment consisting of 35 2-D point features. Additionally, known data association is considered.

Simulations were performed with 100 particles. Each run involved one loop trajectory. All results were obtained by averaging 50 Monte Carlo trials under the same conditions. Due to the comparison and a second time sampling in the MCMC move step, the computational complexity is increased.

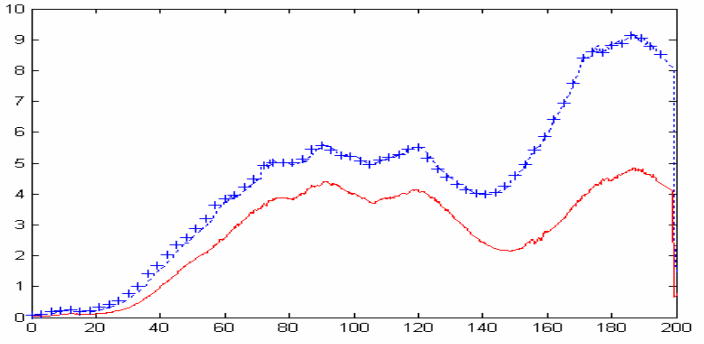


Fig. 2. Root-mean-square of location error. This figure shows the root-mean-square along the x-axis component of the vehicle position estimate for the RBPF SLAM (dotted line with cross) versus that of the UMCMCPF SLAM (solid line).

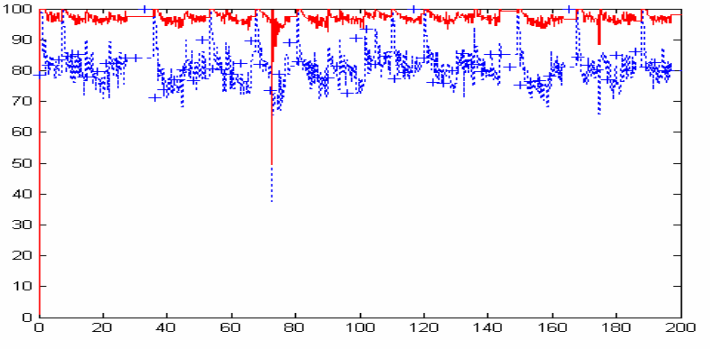


Fig. 3. Number of distinct particles. This figure depicts the number of distinct particles of the RBPF SLAM (dotted line with cross) and that of the UMCMCPF SLAM (solid line).

By comparing the root-mean-square between the traditional RBPF SLAM2.0, RBPF SLAM2.0 incorporated with unscented MCMC, Figure 2 shows that the location error of UMCMCPF SLAM is smaller than the traditional method over time. An estimate of the rate of loss of particle diversity is obtained by the number of distinct particles. It can be appreciated from Figure 3 that UMCMCPF SLAM exhibits significant improvement over RBPF SLAM. The NEES results in Figure 4 indicate that UMCMCPF SLAM stays consistent for a long term, whereas RBPF SLAM becomes optimistic after 70 seconds shown.

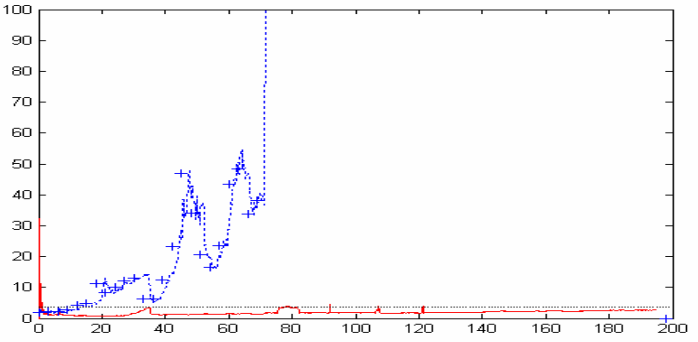


Fig. 4. Comparison of the NEES of the RBPF SLAM (dotted line with cross) and that of the UMCMCPF SLAM (solid line). The horizontal dashed line indicates the upper bound of the 95% probability concentration region for a consistent filter.

5 Conclusions

Upon analysis we have concluded that the inconsistency of RBPF SLAM is in large part due to particle degeneracies and depletion. By applying the UKF to the vehicle states we obtain more accurate proposal mean and variance estimate, leading to more evenly weighted particles. So there is less need for resampling. The second method, an MCMC resampling, improves the diversity of the trajectory samples. In the presented approach, it is possible to maintain consistency in the long-term. Results of simulation experiments show that our approach leads to reductions in estimation error and improved consistency over previous approaches. It is preferable in performance and implementation.

Acknowledgments. This work was supported by the National High Technology Research and Development Program of China (863 Program) under granted number 2006AA12Z305-1.

References

1. Smith, R., Self, M., Cheeseman, P.: A stochastic map for uncertain spatial relationships. In: Fourth International Symposium of Robotics Research, pp. 467–474. MIT Press, Cambridge (1987)
2. Castellanos, J., Neira, J., Tardos, J.: Limits to the consistency of EKF-based SLAM. In: 5th IFAC Symposium on Intelligent Autonomous Vehicles (2004)
3. Murphy, K.P.: Bayesian map learning in dynamic environments. J. Advances in Neural Information Processing Systems 12, 1015–1021 (2000)
4. Montemerlo, M., Thrun, S., Koller, D., Wegbreit, B.: FastSLAM: A factored solution to the simultaneous localization and mapping problem. In: 18th National Conference on Artificial Intelligence (AAAI 2002), pp. 593–598. American Association for Artificial Intelligence, Edmonton (2002)
5. Thrun, S.: Learning occupancy grid maps with forward sensor models 15, 111–127 (2003)

6. Bailey, T., Nieto, J., Nebot, E.: Consistency of the FastSLAM algorithm. In: International Conference on Robotics and Automation (ICRA), vol. 1-10, pp. 424–429. IEEE Press, New York (2006)
7. Jianhui, G., Chunxia, Z., Jianfeng, L.: Consistency of Simultaneous Localization and Map Buiding with Rao-Blackwellised Particle Filter. J. Sys. Sim. 20, 6401–6403 (2008)
8. Beevers, K.R., Huang, W.H.: Fixed-lag sampling strategies for particle filtering SLAM. In: International Conference on Robotics and Automation, pp. 2433–2438. IEEE Press, New York (2007)
9. Doucet, A., De Freitas, N., Gordan, N.: Sequential Monte-Carlo Methods in Practice. Springer-Verlag, New York Inc. (2001)
10. van der Merwe, R., Wan, E.: Gaussian mixture sigma-point particle filters for sequential probabilistic inference in dynamic state-space models. In: International Conference on Acoustics, Speech, and Signal, pp. 701–704. IEEE Press, New York (2003)
11. Julier, S., Uhlmann, J., Durrant-Whyte, H.F.: A New Method for the Nonlinear transformation of means and covariances in filters and estimators. J. IEEE T. Automat. Contr. 45, 477–482 (2000)
12. Wan, E., van der Merwe, R.: The unscented Kalman filter for nonlinear estimation. In: Proceedings of Symposium on Adaptive Systems for Signal Processing Communications and Control, pp. 153–158. IEEE, Piscataway (2000)
13. Gilks, W.R., Berzuini, C.: Following a moving target - Monte Carlo Inference for Dynamic Bayesian models. J. Roy Stat. Soc. Ser. B-Stat. Meth. 63, 127–146 (2001)
14. Andrieu, C., De Freitas, N., Doucet, A., Jordan, M.I.: An introduction to MCMC for machine learning. J. Mach. Learn. 50, 5–43 (2003)

A Deceleration Assistance Control for Collision Avoidance Based on Driver's Perceptual Risk

Takahiro Wada, Shoji Hiraoka, and Shun'ichi Doi

Faculty of Engineering, Kagawa University,
2217-20 Hayashi-cho Takamatsu, Kagawa, Japan
{wada,sdoi}@eng.kagawa-u.ac.jp

Abstract. It is important to judge necessity of deceleration assistance as earlier as possible and initiate the assistance naturally in order to reduce rear-end crashes. On the other hand, we have derived a mathematical model of driver's perceptual risk of proximity in car following situation and successfully derived driver deceleration model to describe deceleration patterns and brake initiation timing of expert driver. In this research, an deceleration assistance control for collision avoidance will be proposed based on the formulated braking behavior models of expert drivers to realize smooth, secure brake assistance. It will be shown that the proposed control method can generate smooth profile for various conditions. In addition, experimental results using a driving simulator will show validity of the proposed system based on subjective evaluation.

Keywords: Collision avoidance, deceleration assistance control, automotive safety, perceptual risk.

1 Introduction

To reduce and mitigate crashes in road traffic, driver assistance systems such as warning system and pre-crash safety system have been developed. Let us consider a scenario in which a driver follows a car in the same lane as shown in Fig.1. Rear-end crash occurs when the driver does not notice the approach of the preceding car due to driver's errors etc. In the view point of preventive safety, deceleration assistance control is effective when collision risk is high and it is difficult for the driver to avoid it. On the other hand, driver can feel anxiety or nuisance against the system if the initiation timing of automatic brake and/or deceleration profile is not appropriate and it may make the system inefficient. Thus, in order to realize an acceptable and efficient system, it is important to know characteristics of comfortable deceleration behavior and apply them to deceleration assistance system.

Recently, several research studies have been conducted on driver's deceleration behaviors including collision risk for application to driver assistance system. Kondoh et al. investigate the risk perception and showed that it can be represented by TTC (Time-to-collision) and THW(Time-Headway) [1]. Isaji et al.[2] and Wada et al.[3] have proposed a performance index of approach and alienation, K_{dB} as a model of driver's perceptual risk of a preceding vehicle based on area change of preceding

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 215–225, 2009.

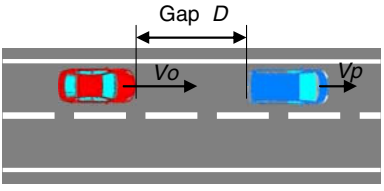


Fig. 1. Car following situation

vehicle on driver's retina and its another version, K_{dB_c}. These indices have been applied to modeling of braking behaviors of expert drivers [4]. Kitajima et al. have surveyed such evaluation indices concerning rear-end collision risk [5].

On the other hand, there are researches concerning design and evaluation of collision avoidance system and ACC(Adaptive Cruise Control) system. For example, Goodrich et al. [6] characterized the behavior in a phase plane of TTC vs. THW. Bareket et al. have evaluated efficacy of ACC based on Gipps model that is a car-following model in traffic engineering [7]. Hiraoka et al. derived car-following model for realizing comfortable ACC system by applying concept of minimum jerk model to longitudinal vehicle behavior [8]. Suzuki et al. proposed a method to estimate driver status in car following situation and its application to driver assistance system [9]. It is important to introduce driver's perceptual risk described before into its design of such driver assistance systems in order to realize comfortable and secure system. To implement this concept, we have proposed a deceleration control method of automobile based on the perceptual risk [10][11].

A deceleration assistance control method will be proposed for preventing rear-end crash based on an expert driver's deceleration model derived from driver's perceptual risk. Initiation timing of brake assistance will be determined by driver's brake initiation timing model. Final target status of two vehicles, say, convergence distance by the braking system will be determined based on driver's risk model. Finally, deceleration profile connecting the brake initiation timing and final target status will be determined by driver' deceleration pattern model. Validity of the proposed control method will be shown by the experiments using a driving simulator.

2 Deceleration Behavior Model of Expert Drivers Based on Perceptual Risk

2.1 Index of Perceptual Risk of Proximity

In car following situation as shown in Fig. 1, the driver evaluates the risk against approach of the preceding car appropriately and realizes safe driving by operating pedals and a steering wheel based on the perceived results. So far, we have hypothesized that drivers detect the approach of the preceding car and recognize the risk by its area changes on the retina and determine the operation of deceleration based on it and a perceptual risk index K_{dB} has been derived as eq.(1) [2],[3].

$$K_{dB} = \begin{cases} 10\log_{10}(|4\times10^{7}\times\frac{V_r}{D^3}|)\mathrm{sgn}(-Vr) & (|4\times10^{7}\times V_r / D^3| \geq 1) \\ 0 & (|4\times10^{7}\times V_r / D^3| < 1) \end{cases} \quad (1)$$

where D denotes gap between two vehicles. Relative velocity V_r is defined as eq.(2) using velocity of the preceding vehicle V_p, and velocity of the following vehicle V_o.

$$V_r(t) = V_p(t) - V_o(t) = \frac{d}{dt}D(t) \quad (2)$$

We call this variable K_{dB} "performance index for approach and alienation" at the moment of the driver's operation such as deceleration and acceleration. Index K_{dB} is increased when the preceding car is approaching relatively to the following car as similar as increase of the driver's visual input. Index K_{dB} is increased when the driver does not react to this regardless of cause of risky conditions such as low arousal level, inattention or other reasons depending on driver's status. It has been shown that K_{dB} can discriminate between braking behaviors of normal safe driving and those in crash accidents that are extracted from micro data of crashes [2], [3]. In addition, it has been also shown that KdB can be a trigger to transit following mode to deceleration mode in the case that driver's following behavior is modeled by mode transition model based on hybrid dynamical system [9].

In addition, another perceptual risk index K_{dB_c} has been proposed as eq.(3) by introducing effect of changes of perceptual risk by preceding vehicle's velocity into account and it has been shown that it can formulate brake initiation timing [4].

$$K_{dB\_c}(a) = \begin{cases} 10\log_{10}(|4\times10^{7}\ \frac{-V_r + aV_p}{D^3}|)\mathrm{sgn}(-V_r + aV_p) & (|4\times10^{7}\times\frac{-V_r + aV_p}{D^3}| \geq 1) \\ 0 & (\text{else}) \end{cases} \quad (3)$$

2.2 Model of Expert Driver's Deceleration Pattern

Model of Deceleration Profile

Tsuru et al.[11] and Wada et al.[10] showed that deceleration pattern of expert drivers can be characterized using risk index K_{dB} as follows based on the results of the experiments with real cars.

P1) Index K_{dB} is changed with the same slope $dK_{dB}/dD=dK_{dB}(t_{bi})/dD$ as in phase I or constant slope phase in Fig. 2.

P2) Constant deceleration is held after peak deceleration until V_r=0. as in phase II or constant deceleration phase in Fig. 2.

Deceleration Model of Constant slope phase

Expert driver's deceleration behavior was modeled by constant slope feature(P1) and the peak hold feature (P2). The peak hold feature is difficult to be installed in the automatic braking system due to lack of robustness against situation changes. Thus, let us focus on P1 for applying deceleration profile generation method.

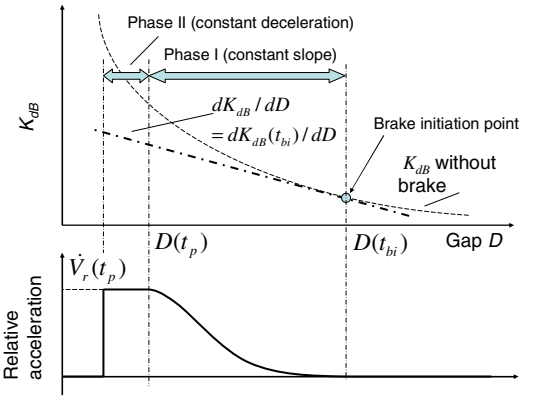


Fig. 2. Schematic image of expert driver's deceleration model

Suppose that we deal with only approaching condition because we are considering deceleration assist system. And assume that $|4\times10^7\times V_r/D^3|\geq 1$ is satisfied. In this case, deceleration profile model and its integration form, say, relative velocity profile can be given eqs.(4) and (5).

$$\dot{V}_r(t)=\left(\frac{3}{D(t)}-\frac{3}{D(t_{bi})}+\frac{\dot{V}_r(t_{bi})}{V_r^{\,2}(t_{bi})}\right)V_r^{\,2}(t) \tag{4}$$

$$V_r(t)=\frac{V_r(t_{bi})}{D^3(t_{bi})}D^3(t)\exp\left\{\left(\frac{\dot{V}_r(t_{bi})}{V_r^{\,2}(t_{bi})}-\frac{3}{D(t_{bi})}\right)(D(t)-D(t_{bi}))\right\} \tag{5}$$

Constant relative velocity situation

For the sake of simplicity, we deal with the situation approaching to a preceding car driving with a constant velocity. In this situation, relative deceleration is zero until braking behavior of the following car's driver. Based on the assumption, substituting $\dot{v}_r(t_{bi})=0$ into eqs.(4) and (5) leads to eqs.(6) and (7), respectively.

$$\dot{V}_r(t)=\left(\frac{3}{D(t)}-\frac{3}{D(t_{bi})}\right)V_r^{\,2}(t) \tag{6}$$

$$V_r(t)=V_r(t_{bi})\,d^3(t)\exp\{3(1-d(t))\} \tag{7}$$

where $d(t)=D(t)/D(t_{bi})$. The derived deceleration profile results in collision with $V_r = 0$ under the given assumptions as long as the calculated deceleration can be generated, that is, the state is uniquely converged to its equilibrium point $[V_r, D]^{\mathrm{T}} = [0, 0]^{\mathrm{T}}$. Very smooth deceleration profile can be obtained with only simple calculation of eq.(6) [10].

2.3 Model of Expert Driver's Deceleration Timing

We have analyzed expert drivers' braking timing with experiments with real cars and it has been shown that brake judgment line eq.(8) can describe timing of expert driver based on the index KdB_c.

$$\phi(V_r, V_p, D) = K_{dB\_c}(a) - b\log_{10} D - c = 0 \tag{8}$$

The coefficients *a, b,* and *c* are determined so that the approximated error of the equation is minimized in terms of least squares. For the experimental results with test drivers, a=0.2, b=-22.66, c=74.71 were obtained.

It has already been shown that the judgment line of brake initiation can discriminate between normal safe driving and micro data of crashes very well. Probability that the plots for the normal driving were located in the upper area than the line is 0.00694. On the other hand, for the crash data, probability that the crash data was located lower than the line is 0 [4].

By solving eq.(8) by *D*, the gap when brake initiation timing is obtained as eq.(9)

$$D = \left[4\times 10^{7-\frac{c+\Delta d}{10}} \times (-v_r + a\times V_p) \right]^{\frac{10}{30+b}} \tag{9}$$

Fig. 3 illustrates relationship between the gap at brake initiation and *Vr* for each *Vp* by eq.(9). As seen from this, Larger gap can be obtained in larger *Vr* and larger *Vp*.

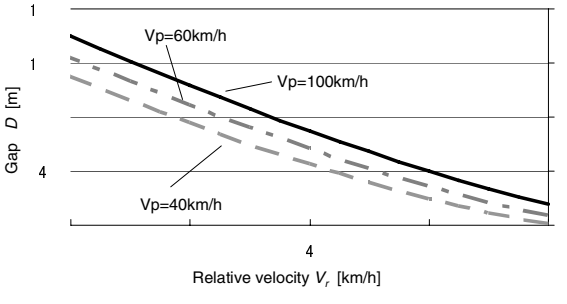


Fig. 3. Relationship between gap at brake initiation and relative velocity

3 Deceleration Control Method for Collision Avoidance Based on Perceptual Risk

The proposed system aims that the vehicle equipped with the system starts to decelerate automatically and avoid collision if the driver does not decelerate or decelerates insufficiently even in high risk situation against the preceding vehicle by driver's failure etc.

3.1 Determination of Brake Initiation Timing

Brake judgment model obtained in the previous section is an averaged result of driver's brake initiation. Thus, drivers might rely on the system excessively and do not act any braking because the system starts to brake automatically just when the driver will start brake if the model is employed as brake initiation algorithm without any change and it works perfectly. In addition, there are individual differences in brake initiation timing. For example, aggressive drivers start to brake in more risky situations. Thus, it is important to take individual differences of brake initiation into account in brake initiation judgment to avoid discomfort.

Therefore, brake assist control is initiated when the state enter the dangerous status defined by eq.(10) by adding an offset of the line Δc to eq.(8) as follows:

$$\Omega_{dangerous} = \{[V_r, V_p, D] \mid \phi(V_r, V_p, D) \geq \Delta c\} \tag{10}$$

where Δc is determined by taking individual difference into account. Let us consider the effect of the offset Δc . By soloving $\phi = \Delta c$ by D yields the eq.(11). This shows Δc =1 decreases the gap by a factor of 0.731.

$$D = \left[4 \times 10^{7 - \frac{c + \Delta c}{10}} \times (-v_r + a \times V_p) \right]^{\frac{10}{30+b}} = 10^{-\frac{\Delta c}{30+b}} D \mid_{\Delta d = 0} = (0.731)^{\Delta c} D \mid_{\Delta d = 0} \tag{11}$$

3.2 Determination of Final Target Status [12]

Final converged status of two vehicles after collision avoidance by deceleration assist is important because it affects driver's recovering behavior after the assist and peripheral traffic flow especially for the system that works during relatively earlier stage. In this paper, the final target status will be determined based on driver's perceptual risk.

Function ϕ (Vr, Vp, D) defined in eq.(8) can be understood as driver's perceptual risk of collision. Eq.(9) with assumption of $-V_r + aV_p > 0$ yields eq.(12).

$$\phi(Vr, Vp, D) = 10 \log_{10}(4 \times 10^7 \times \frac{-V_r + aV_p}{D^3}) - b \log_{10} D - c = 0 \tag{12}$$

Here, let us consider a method to determine the target converged status by specifying driver's perceptual risk at the final status. Namely, the driver's risk at the converged status is specified as eq.(13) by taking safe margin Δd into account.

$$\Omega_{conv} = \{[V_r, V_p, D] \mid \phi(V_r, V_p, D) = \Delta d\} \tag{13}$$

Now, also assume that Vr =0 at the converged status. Therefore, converged gap satisfying eq.(13) is calculated as eq.(14) given Vr=0 and current Vp.

$$D = \left[4 \times 10^{7 - \frac{c + \Delta d}{10}} \times aV_p \right]^{\frac{10}{30+b}} \tag{14}$$

Namely, converged gap is determined by specifying preceding car's velocity *Vp*. As seen from equation, D_{conv}=0 at V_p=0. So, eq.(15) is derived by adding gap offset ΔD>0.

$$D_{conv} = \left[4 \times 10^{7-\frac{c+\Delta d}{10}} \times a V_p \right]^{\frac{10}{30+b}} + \Delta D \tag{15}$$

By setting target converged gap as eq.(15), converged status can be determined as $[V_r, V_p, D]^T = [0, V_p, D_{conv}]^T$, $\phi(Vr, Vp, D) \le \Delta d$ is realized.

3.3 Generation of Deceleration Pattern

Let us consider a way to generate deceleration pattern based on the formulated expert driver's deceleration profile P1, say constant slope feature as eq.(6). Namely, velocity control based on eq.(7) is employed. It should be noted that the profile calculated by eq.(7) has a equilibrium point $[V_r\ ,\ D]^T=[0,\ 0]^T$. Thus, we need to have a way to change the equilibrium to $[V_r\ , D]^T = [0, D_{conv}]^T$. Let us define variable δ as eq.(16).

$$\delta(t) = \frac{D(t) - D_{conv}}{D_{tbi} - D_{conv}} = \frac{\overline{D}(t)}{\overline{D}_{tbi}} \tag{16}$$

Replacing $d(t)$ by this $\delta(t)$ yields a new desired velocity profile eq.(17).

$$Vr(t) = V_r(t_{bi})\delta^3(t)\exp\{3(1-\delta(t))\} \tag{17}$$

As seen from the equation, the following car always decelerates in approaching condition Vr<0 and the equilibrium point is represented as $[\mathrm{Vr}, \delta]^T=[0,\ 0]^T$.

3.4 Procedure of Deceleration Assistance Control

Consequently, the following control method has been obtained for deceleration assistance control for collision avoidance:

1) $\phi(V_r, V_p, D)$ is calculated in real-time from measured $V_r(t)$, $D(t)$ and $V_p(t)$. Brake control starts when the status is judged as dangerous status by eq.(10)
2) Desired relative velocity can be determined by the profile model eq.(17).
3) The brake control terminates if $V_r >= 0$.

Acceleration command from the given velocity profile is generated by the following simple method as an example:

$$G = -k_p(V_r{}^d(D) - V_r(t)), \tag{18}$$

where k_p is feedback gain.

3.5 Simulation Results of Proposed Deceleration Assistance Control

In order to show the effectiveness of the proposed deceleration assistance control, some numerical simulations are performed with the sequences given in Section 3.4. Fig. 4-(a) illustrates the behavior of the proposed control method in approaching to the preceding car driving at 40km/h with a constant relative velocity of Vr=-40km/h. As seen in the figure, it is found that smooth velocity profiles are realized and they avoid the collision. As same as the results of the method proposed in literature [10], deceleration is relatively increased rapidly just after the brake initiation, and then smoother profile realizes the target velocity. Fig. 4-(b) illustrates the sudden deceleration situation in order to see the effectiveness of the robustness of the proposed method. In the initial state, the following vehicle and the preceding vehicle drive at 40km/h with gap D=25m. At t=2s, the preceding vehicle starts to decelerate in -2m/s^2 then stopped. From this result, it is found that the proposed deceleration method can realize smooth deceleration and avoid collision even in such a situation including deceleration of the preceding car.

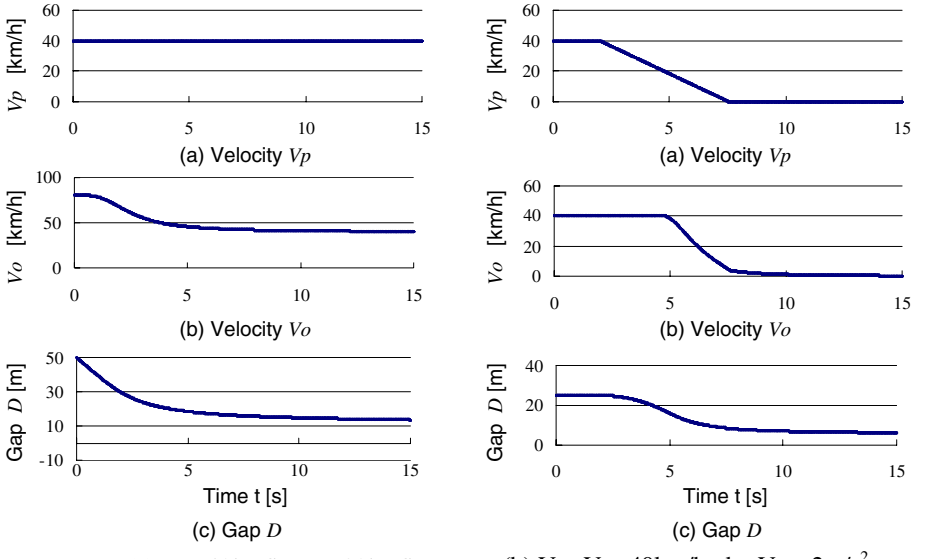


Fig. 4. Simulation results

4 Experiments Using Driving Simulator

A fixed-base driving simulator developed by Kagawa University was utilized for the experiments. A 100inch screen and two 80inch screens are located in front of the cockpit. Distance from driver's eyes to the main screen is 2.3m in depth. Visual angle

by the three screens is about 140deg. Vehicle motion is calculated by vehicle dynamics simulation software CarSim based on driver's operation

Experimental scenario was car following situation in a lane as shown in Fig. 1. A following car with constant velocity approached to the preceding car that drives in a constant velocity. Following car always drives faster than the preceding car. Participants sit on the driver's seat and take driving posture and step on the gas pedal even though the operation is not reflected to the vehicle motion and the car drives at constant velocity automatically. After that, the following car started to decelerate automatically based on the proposed control method. As the experimental conditions, the preceding car velocity was Vp=40, 60, 80km/h and there are three conditions for relative velocity as Vr=-20, -40, -60km/h for each Vp condition. Order of the experimental conditions was randomized.

Participants were observed deceleration of the following vehicle by the proposed control method and asked to evaluate subjectively the converged status, say, the converged gap of two vehicles after termination of the deceleration control based on the five levels of 1) near, 2)slightly near, 3) good, 4)slightly far, 5) far after each trial. Participants were five male students of Kagawa University of 21 to 24 yrs. Parameters in the control method were set as $\Delta c = \Delta d = 0$. The offset for the converged gap was set as ΔD=5m based on the rehearsal experimental results.

Fig. 5 shows subjective evaluation results of all participants. For almost all conditions, the participants evaluated as "good". There was some participants who evaluated "slightly near" in Vr=-40, -60km/h conditions. This implies that deceleration profile to the converged status affects driver's perceptual risk even for the same converged status. In addition, distance feeling of the driving simulator can be another reason of that. In addition, all participants have positive comments about the total deceleration patterns from free comments. This shows that the proposed method realize smooth deceleration control for overall.

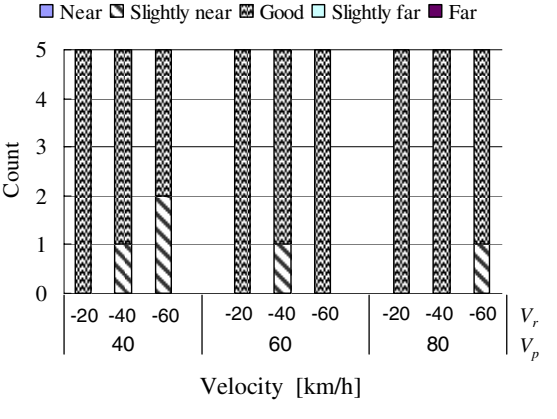


Fig. 5. Experimental Results

5 Conclusion

A new deceleration assistance control to avoid rear-end collision was proposed based on driver's perceptual risk that works also in relatively low emergency level. The simulation results showed that smooth profile can be generated appropriately with the proposed method. In addition, almost good results were obtained about converged status by the subjective evaluation in the simulator experiments. However, it is found that some participants evaluated the same target status differently in the case of the different deceleration profiles. This implies that the deceleration profile affects the driver's perceptual risk largely.

Investigation of robustness of the proposed method in the relatively risky conditions will be important future work. Namely, the it is important to show the affinity of the system with the conventional driver assistance system by investigating the covered range of the collision risk by the system. Furthermore, it is also one of the most important future studies a method to avoid driver's over-trust including human machine interface.

References

1. Kondoh, T., Yamamura, T., Kitazaki, S., Kuge, N., Boer, E.R.: Identification of Visual Cues and Quantification of Drivers' Perception of Proximity Risk to the Lead Vehicle in Car-Following Situations. In: Proc. of Driving Simulation Conference Asia/Pacific (2006)
2. Isaji, K., Tsuru, N., Wada, T., Imai, K., Doi, S., Kaneko, H.: Emergency Evaluation by Approaching in Longitudinal Direction (First Report) – Proposal of Performance Index for Approach and Alienation based on the Driver Operations. Transactions of Society of Automotive Engineers of Japan 38(2), 25–30 (2007) (in Japanese)
3. Wada, T., Doi, S., Imai, K., Tsuru, N., Isaji, K., Kaneko, H.: Analysis of Drivers' Behaviors in Car Following Based on A Performance Index for Approach and Alienation. SAE 2007 Transactions, Journal of Passenger Cars – Mechanical Systems, SAE paper 2007-01-0440, 306–320 (2008)
4. Wada, T., Doi, S., Imai, K., Tsuru, N., Isaji, K., Kaneko, H.: Analysis of Braking Behaviors in Car Following Based on A Performance Index for Approach and Alienation. In: Proceedings of IEEE Intelligent Vehicles Symposium, pp. 547–552 (2007)
5. Kitajima, S., Marumo, Y., Hiraoka, T., Itoh, M.: Theoretical and Empirical Comparison of Evaluation Indices Concerning Rear-end Collision Risk. In: Proceedings of International Symposium on Advanced Vehicle Control, vol. 2, pp. 602–607 (2008)
6. Goodrich, M.A., Boer, E.R., Inoue, H.A.: Characterization of Dynamic Human Braking Behavior with Implications for ACC Design. In: Proc. IEEE Conference on Intelligent Transportation Systems, pp. 964–969 (1999)
7. Bareket, Z., Fancher, P.S., Peng, H., Lee, K., Assaf, C.A.: Methodology for Assessing Adaptive Cruise Control. IEEE Transactions on Intelligent Transportation Systems 4(3), 123–131 (2003)
8. Hiraoka, T., Kunimatsu, T., Nishihara, O., Kumamoto, H.: Modeling of Driver Following Behavior based on Minimum-jerk Theory. In: Proceedings. of 12th World Congress on Intelligent Transport Systems, CD-ROM (2005)
9. Suzuki, T., et al.: Modeling and Analysis of Vehicle Following Task based on Mode Segmentation. In: Proceedings of 9th International Symposium on Advanced Vehicle Control, pp. 614–619 (2008)

10. Wada, T., Doi, S., Nishiyama, A., Tsuru, N., Isaji, K., Kaneko, H.: Analysis of Braking Behavior in Car Following and Its Application to Driver Assistance System. In: Proc. of International Symposium on Advanced Vehicle Control, vol. 2, pp. 577–583 (2008)
11. Tsuru, N., Isaji, K., Doi, S., Wada, T., Imai, K., Kaneko, H.: Emergency Evaluation by Approaching in Longitudinal Direction (Third report). In: Proc. of Annual Conference of JSAE, Paper No.20065721 (2006) (in Japanese)
12. Wada, T., Doi, S., Hiraoka, S.: A Deceleration Control Method of Automobile for Collision Avoidance based on Driver's Perceptual Risk. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (in Press, 2009)

Toward Preventive ACC Systems against Crashes Due to Another Vehicle's Cut-in

Makoto Itoh

Department of Risk Engineering, University of Tsukuba, 1-1-1 Tennodai,
Tsukuba 305-8573, Japan
itoh@risk.tsukuba.ac.jp

Abstract. This paper discusses design of a driver support system for managing risks of crashes due to another vehicle's cut-in. In study 1, we investigate to what extent changing of the target can be automated when a vehicle is cutting in. Experimental results suggest that the autonomous target change may be acceptable for drivers. In study 2, we discuss about reducing driver annoyance with information given from the system if "providing information only" design is employed. The idea is that such information is inhibited when a driver is aware of the situation. A prototype method to detect driver's "preparedness" to hit the brake is shown.

Keywords: Human-computer collaboration, human-machine interface, human-centered design, automobile, driver monitoring.

1 Introduction

When road traffic is congested, it is often observed that a vehicle cuts in very closely to another vehicle. According to an investigation of behavior of real-world truck drivers [4], changing lanes is very often even for large trucks with a speed governor. It is a vital issue to develop methods for reducing risks of vehicle-to-vehicle crashes due to a vehicle's cutting in.

What kind of support is useful for reducing the risks of the crashes? Systems informing the existence of a vehicle in the "dead-zone" have already been put into market. In addition to them, Inagaki, Itoh and Nagai [2] discussed protective functions which prevent maneuver of changing lanes when a crash may occur if the maneuver is done. Note here that such information systems or protective functions are for drivers who intend to change lanes. On the other hand, it is also important to develop a "defensive" support for a driver whose vehicle is cut in by another vehicle. This paper discusses about the development of a "defensive" support when an Adaptive Cruise Control (ACC) system is active.

Conventional ACC systems have no ability to avoid a hazard before the cutting-in vehicle completes to change lanes [3]. However, today's image recognition techniques (e.g., see, [7]) may contribute to develop novel "proactive" supports for avoiding the hazard when a vehicle is going to cut in.

What kind of support is useful for avoidance of the hazard? This paper reports results of two studies on this issue. In study 1, we discuss possibility of system's

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 226–234, 2009.

autonomous changing of the target to follow when the system detects that a vehicle has an intention to cut in. Experimental results suggest that the autonomous target change may be acceptable by drivers. "Providing information only" was the second best. In study 2, we discuss evaluation of driver situation awareness in a non-intrusive manner. Estimation of driver situation awareness may be useful for reducing driver annoyance caused by the provided information if "providing information only" design should be employed. The idea is that such information is inhibited when the systems regards that a driver is aware of the situation. We develop a method to detect driver's "preparedness" to hit the brake when a vehicle is cutting in. Experimental results suggest that the proposed method is effective to detect the preparedness.

2 Study 1

Suppose the host vehicle (H) is running on a cruising lane. In the vehicle H, an Adaptive Cruise Control system is activated to maintain the headway distance to the forward vehicle. The traffic is so congested that the speed of H is around 30 - 50 km/h, i.e., the ACC is doing "low speed following." Now a vehicle named A running on the next lane is going to cut in just front of H (host vehicle) as shown in Fig. 1. Conventional ACC systems recognize the vehicle A only if the lane change of the vehicle A is almost completed. Therefore, the host vehicle H does not decelerate to widen the distance to the vehicle A at the exact time point shown in Fig. 1. However, according to today's computer vision techniques, we can assume that the host vehicle H has an ability to detect that the vehicle A has intent to change lanes.

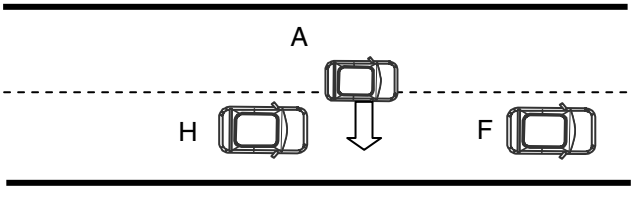


Fig. 1. A Situation where a Vehicle is Cutting in

In order to reduce the risk of a collision with the vehicle A, it might be necessary for a driver of the host vehicle to decelerate its speed before completion of the vehicle A's lane change.

As for the way of supporting the deceleration for widening the distance to the vehicle A, at least three types are distinguished in terms of levels of automation (LOA): L4, L6, and L7 (see, Table 1). At L4, the system informs that a vehicle is going to cut in and suggests slight deceleration to the driver for widening the distance to the vehicle cutting in. The driver should hit the brake by him/herself when necessary. At L6, the system informs that a vehicle is going to cut in and proposes to change the target (and to reduce its speed for maintaining a proper clearance to the new target (vehicle A)). If the driver does not veto within several seconds, the target change will be done automatically (We can assume that vetoing is done by just pushing a button). At L7, the system changes the target automatically and informs about it to the driver afterwards.

Table 1. Scales of Levels of Automation [9]

1.	The computer offers no assistance; human must do it all.
2.	The computer offers a complete set of action alternatives, and
3.	narrows the selection down to a few, or
4.	suggests one, and
5.	executes that suggestion if the human approves, or
6.	allows the human a restricted time to veto before automatic execution, or
7.	executes automatically, then necessarily informs humans,
8.	informs him after execution only if he asks,
9.	informs him after execution if it, the computer, decides to.
10.	The computer decides everything and acts autonomously, ignoring the human.

We conducted an experiment in order to investigate which LOA is the best among the three. Twelve graduate/undergraduate students participated in this experiment. LOA was a within-subject factor. The order of experiencing LOAs was randomized for each participant. Each participant received 10 scenarios for each LOA, including events #1, #2, and #3 that are potentially disadvantageous for L4, L6, and L7, respectively.

We measured participants' subjective feelings on trust in each support type. Lee & Moray [5] distinguished four dimensions of trust. Among the four, the following three dimensions are important: (1) 'performance' that rests on the "expectation of consistent, stable, and desirable performance or behavior," (2) 'process' that depends on "an understanding of the underlying qualities or characteristics that govern behavior," and (3) 'purpose' that rests on the "underlying motives or intents." The participants were requested to evaluate the following questions in 11-point scales: (Q1) "To what extent do you think this system is consistent, stable and desirable?" (0: not at all, 5: I am not sure, 10: completely), (Q2) "Is the logic of this system understandable?" (0: not at all, 5: I am not sure, 10: completely), (Q3) "Is the purpose of this system acceptable?" (0: not at all, 5: I am not sure, 10: completely), and (Q4) "In total, to what extent do you trust this system?" (0: not at all, 5: I am not sure, 10: completely).

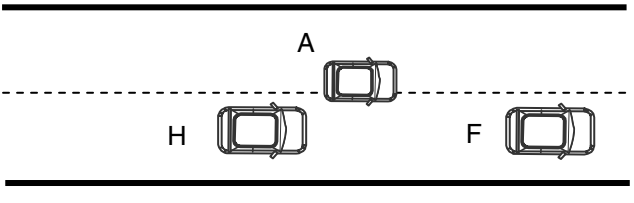


Fig. 2. Event #1 (The vehicle A remains on the border for a while)

Figure 5 shows results of event #1. An ANOVA on the subjective feeling on 'purpose' (Q3) showed that the main effect of LOA was nearly significant ((F(2,22)=3.39, p=0.052). Tukey's HSD test suggested that the difference between L6 and L7 was significant. Similar tendency can be found in the subjective feelings on total trust (Q4). There was not significant main effect for other subjective feelings. Why these results

were obtained? Table 2 may give an answer. Table 2 shows what each participant does when the vehicle A remains on the border. It can be claimed that L6 has a disadvantage that drivers may be confused trying to make a decision on canceling automatic change of the target to follow if the drivers interact with the support system at L6.

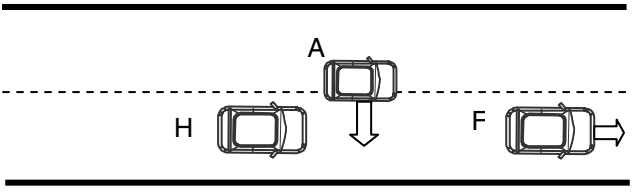


Fig. 3. Event #2 (The vehicle F begins to accelerate during the vehicle A's changing lanes.)

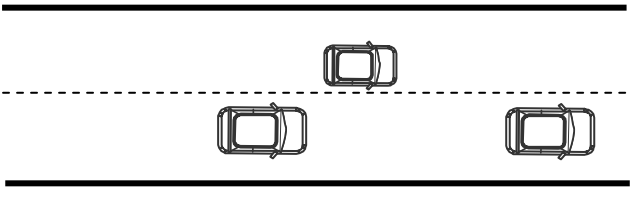
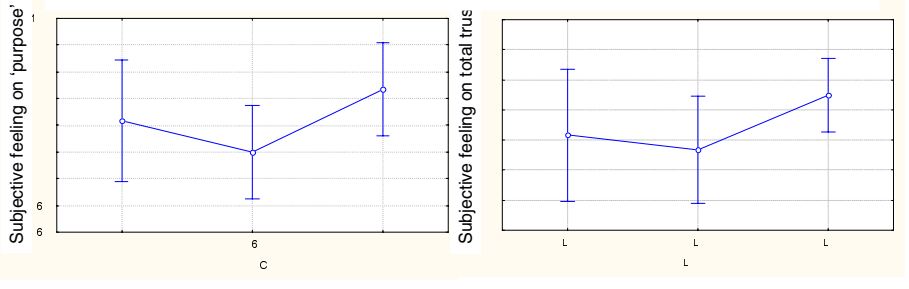


Fig. 4. Event #3 (The vehicle A is on the right hand side of the passing lane. However, the vehicle A will not change lanes.)



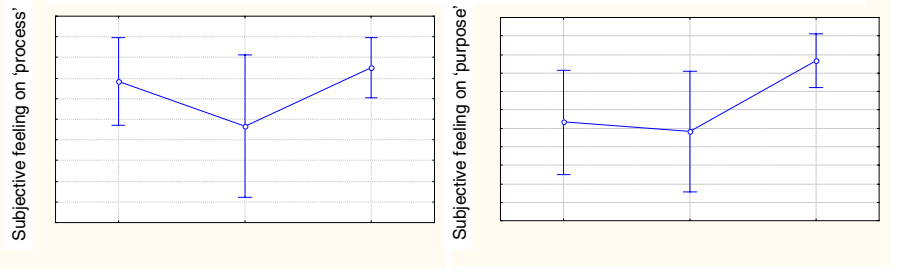
(a) Q3 (b) Q4

Fig. 5. Subjective Feelings (Event #1)

Table 2. Actions in Event #1

	L4	L6	L7
Hit the brake	2	1	0
Prepared but not to hit the brake	1	0	0
Pushed the cancel button	-	6	-
Prepared but not to push the cancel button	-	2	-
None	9	3	12

Figure 6 shows results of event #2. An ANOVA on the subjective feeling on 'process' (Q2) showed that the main effect of LOA was statistically significant ((F(2,22)=3.73, p=0.04). Tukey's HSD test showed that the difference between L6 and L7 was significant. An ANOVA on the subjective feeling on 'purpose' (Q3) showed that the main effect of LOA was statistically significant (F(2,22)=3.91, p=0.035). Tukey's HSD test showed that the difference between L6 and L7 was significant and that the difference between L4 and L7 was nearly significant. Similar tendency can be found in the subjective feelings on total trust (Q4) ((F(2,22)=2.70, p=0.089). There was no significant main effect on performance (Q1). The above results were almost as expected.



(a) Q2 (b) Q3

Fig. 6. Subjective Feelings (Event #2)

On the subjective feelings in event #3, there were no significant main effects of LOA. An example is shown in Figure 7. It is interesting that trust in the system with L7 is not low even if the system changes the target to follow from the vehicle F to the vehicle A which does not cut in.

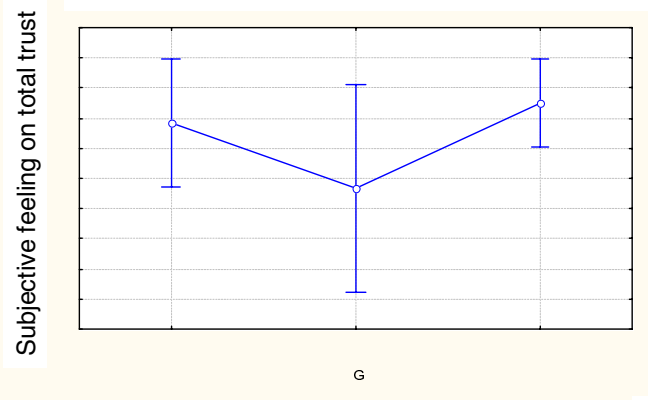


Fig. 7. Subjective Feeling on Total Trust (Event #3)

The above results suggest that L7 is the most preferable among the three LOAs. The results are interesting in a sense that automation surprises [8] were not observed even when the LOA was 7. On the other hand, level of trust in the system at L7 was higher than expected.

It can be also claimed that L4 would be better than L6. Each subjective rating for L4 was almost the same level as or higher than that for L6 in any event. One possible reason for this observation is that drivers may dislike making a decision under such "busy" situation. If designers are still insisted that LOA must be lower than 7 (because of the philosophy of "human-centered automation" [1, 10], L4 should be employed rather than L6.

3 Study 2

If L4 is employed for the risk management in the "car-cutting in" situations, we need to investigate how we can reduce driver's level of annoyance which may be caused from providing information to a driver. If a driver has been aware of the existence of a vehicle which is going to cut in, the information providing is useless. One way to reduce the driver's annoyance level is to managing the system's actions by inferring driver's awareness of the traffic situations. That is, the information is inhibited when a driver is aware of the hazard.

When a vehicle running on the next lane is going to cut in just front of the host vehicle, the host driver may have to prepare to hit the brake by covering his or her leg over the brake pedal. Such "preparedness" is important when the driver uses an ACC system because he or she may hesitate to hit the brake in order to avoid unnecessary disengagement of the ACC system (In general, an ACC system is disengaged when a driver hit the brake).

If a system is able to understand whether or not the host vehicle's driver is aware that the other vehicle is going to cut in by analyzing the driver's "preparedness", message for informing the risk of crashes with the other vehicle can be provided only when it is really necessary (i.e., the host vehicle's driver is not aware of it). In this sense, methods for evaluating driver's "preparedness" for hitting the brake are very important.

In this study, we collected data of drivers' postures with a pressure distribution sensor on the seat cushion in a driving simulator. Ten persons (5 females and 5 males) have participated in this data collection. Five postures are distinguished: (a) bending both knees, (b) covering brake pedal, (c) covering gas pedal, (d) bending left knee, and (e) sprawling (Figure 8). Figure 9 depicts the pressure distribution sensor. Figure 10 gives an example of pressure distribution on the seat cushion.

We have developed a method of detecting the drivers' "preparedness" by analyzing the pressure distribution on the seat cushion. In this study we applied the Higher-order Local Auto-Correlation (HLAC) feature extraction method [6] and a linear discriminant analysis method for detecting the "preparedness."

(a) Bending both knees (b) Covering brake pedal (c) Covering gas pedal

(d) Bending left knee (e) Sprawling

Fig. 8. Five Classes on Positions of Driver's Legs

100 pictures were recorded for each posture of each participant in order to develop a detection system. After recording the data for the learning, the test data were also recorded. For the test data, each participant took posture (a) 40 times, posture (b) 10 times, posture (c) 10 times, posture (d) 10 times, and posture (e) 10 times. The order was randomized. Taking each posture lasts for approximately five seconds. Thirty-six pictures were recorded during each taking of a posture. The result of the evaluation of the proposed method is shown in Table 3. "Rcg" in Table 3 refers to the correct recognition rate. If we are interested in distinguishing Category (b or c) from Category (a, d, e), the correct detection rate was 89.4%. This binary categorization is meaningful because it is hardly that he or she wants to accelerate in the context when a driver is using an ACC system. Thus, it can be claimed that the method is effective for drivers' understanding of a traffic situation.

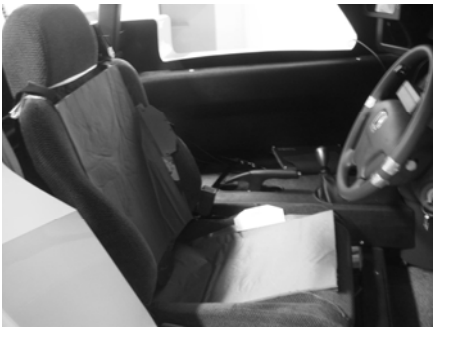


Fig. 9. Pressure Distribution Sensor on the Driving Seat

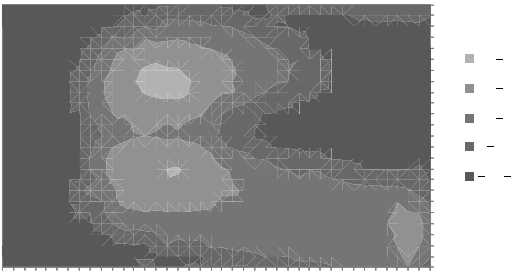


Fig. 10. Pressure Distribution on the Seat Cushion (posture b)

Table 3. Linear Discriminant Analysis on Preparedness to Hit the Brake

				c			c
Actual							0.893
							0.842
	c	2		2			0.514
					2 2		0.646
							0.886

4 Concluding Remarks

This paper discussed what kind of support is useful for avoiding the hazard of a vehicle-to-vehicle crash when another vehicle is cutting in front of the host vehicle. Experimental results in Study 1 suggest that the autonomous target change may be acceptable by drivers. Providing information only was the second best. In Study 2, we discussed on information management for reducing driver annoyance with the information given from the system when a driver is aware of the hazard, if "providing information only" design should be employed. In order to realize such information management framework, it is necessary to develop a method to detect driver's "preparedness" to hit the brake when a vehicle is cutting in. Experimental results suggest that the proposed method to detect driver's preparedness is potentially effective to detect the preparedness.

Acknowledgments

The authors thank Sahori Ito for her assistance in data collection in Study 1 and Shuwei Chi for his assistance in data collection in Study 2.

References

1. Billings, C.E.: Aviation Automation, LEA (1997)
2. Inagaki, T., Itoh, M., Nagai, Y.: Driver Support Functions under Resource-limited Situations. Journal of Mechanical systems for Transportation and Logistics 1(2), 213–222 (2008)
3. ISO 15622: Transport Information and Control Systems – Adaptive Cruise Control Systems – Performance Requirements and Test Procedures, International Organization for Standardization (2002)
4. Itoh, H., Yoshimura, K.: Understanding Human Factors in Long-distance Vehicle Operation. In: Proc. IFAC-HMS, CD-ROM, 6 pages (2007)
5. Lee, J., Moray, N.: Trust, Control Strategies and Allocation of Function in Human Machine Systems. Ergonomics 35(10), 1243–1270 (1992)
6. Otsu, N., Kurita, T.: A New Scheme for Practical Flexible and Intelligent Vision Systems. In: Proc. IAPR Workshop on Computer vision (MVA 1988), pp. 431–435 (1988)
7. Nishida, K., Adachi, E., Kurita, T.: The Recognition of a Safe-overtaking Driving Environment. In: Proc. IFAC-HMS, CD-ROM, 6 pages (2007)
8. Sarter, N.B., Woods, D.D., Billings, C.E.: Automation Surprises. In: Salvendy, G. (ed.) Handbook of Human Factors and Ergonomics, 2nd edn., pp. 1926–1943. Wiley, Chichester (1997)
9. Sheridan, T.B.: Telerobotics, Automation, and Human Supervisory Control. MIT Press, Cambridge (1992)
10. Woods, D.D.: The Effects of Automation on Human's Role: Experience from Non-Aviation Industries. In: Norman, S.D., Orlady, H.W. (eds.) Flight Deck Automation: Promises and Realities, NASA CP-10036, pp. 61–85 (1989)

How Do Cognitive Distraction Affect Driver Intent of Changing Lanes?

Huiping Zhou[1], Makoto Itoh[2], and Toshiyuki Inagaki[2]

[1] Doctoral Program in Risk Engineering, Graduate School of Systems and Information Engineering,
University of Tsukuba, Tsukuba, Japan
zhouhp@css.risk.tsukuba.ac.jp
[2] Department of Risk Engineering, Faculty of Systems and Information Engineering,
University of Tsukuba, Tsukuba, Japan
{itoh,inagaki}@risk.tsukuba.ac.jp

Abstract. This paper revealed effects of cognitive distraction on driver intent of changing lanes. We conducted an experiment to collect data of driver's eye-movement on checking traffic conditions and driving behavior in lane changes in two conditions: driving only and performing a secondary task during driving. The result indicated that there were two types of changes in terms of the intent: (1) driver intent emerges lately and (2) the intent emerges without enough checking on traffic conditions. The investigation also showed that a driver made a decision of changing lanes in a relatively short time period due to driver distraction, which might increase risk in driving. Those findings implied necessity for an intent detection method adaptive to driver psychological state.

Keywords: Lane change, driver intent, driver distraction, preventive safety, adaptive automation.

1 Introduction

Driver distraction is one of major contributing factors to traffic accidents. For example, National Highway Traffic Safety Administration (NHTSA) has shown that approximately 25% of crashes were attributed to driver distraction [1].

It is reported that changing traffic lanes is one of important scenarios that are related with distraction-related crashes [2]. Several attempts have been done for reduction of the distraction-related crashes in lane changes. Inagaki, Itoh, & Nagai [3] developed a framework for providing effective support to prevent such an accident. The support system will be activated when the system regards that the driver is going to change lanes but he or she is not aware of an approaching vehicle on the next lane. In order to implement such a system, it is necessary to develop a method for inferring driver intent to change lanes.

Several methods have already been proposed and developed for distinguishing intended lane departure for changing lanes from unintended departure [4-6]. Salvucci

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 235–244, 2009.

and Siedlecki [7] had developed a computational framework for inferring persons' thoughts by mapping observable actions, e.g., surrounding environment and steering-based features. Based on the framework proposed in [7], Salvucci [8] showed that 90 percent of lane changes were correctly detected just before a vehicle passed a central line. Note that the above studies [4-8] attempted to detect the driver intent just *after* initiating lane changes maneuvers. Zhou et al. [9], on the other hand, showed that driver intent could be detected via analyzing driver's eye-movement to look at the side-view mirror *before* the initiations of lane changes. Their study distinguished four levels of the intent in terms of the eye-movement to look at the side-view mirror. The result revealed that driver intent's level increased as the host vehicle was approaching to a slower lead vehicle and that 80 percent of lane changes were detected at a relatively higher level before the initiations of lane changes. However, there were some cases in which driver intent was detected at a relatively lower level. Interviews after the experiment suggested that driver distraction occurred in those cases. Their next studies [10, 11] investigated driver's eye-movement in distracted driving. The results revealed that the number of driver's looking at the side-view mirror decreased and that the time length of each looking was relatively longer.

The purpose of this paper is to clarify the relationship among distracting cognitive activity, creation of intent to change lanes, and the driver's behavior based on the intent.

2 Experiment

2.1 Apparatus

A motion-base driving simulator (Fig. 1) was used in the experiment. It can simulate driving on an expressway with two lanes. Maximum speed of the host vehicle was limited at 90km/h by a speed governor. An eye-tracking system having three cameras was used to indentify where a driver was looking at.

Fig. 1. Motion-base driving simulator

2.2 Participants

Twelve females and eight males participated in the experiment whose ages ranged from 20 to 55 (Mean = 34.5, standard deviation (*SD*) = 8.43). Each participant held a valid driver's license. The participants were recruited from a local society but the participation was voluntary.

2.3 Driving Task and Scenario

All participants were requested to drive safely on the cruising lane at 90 km/h when there was no slower lead vehicle. The participants were allowed to pass a slower lead vehicle, which drove at 80 km/h. To obtain data for natural lane changes, no instruction was given to participants, e.g., when and how to change lanes.

Each scenario consisted of a "cruising phase" and a "preparation phase of changing lanes". A cruising phase was defined as the time period when the host vehicle had to follow the lead vehicle. During a cruising phase, the lead vehicle's speed was at 90 km/h and the headway-distance from the host vehicle to the lead vehicle was 150 m. Each cruising phase lasted for approximately 30 s. After that, the lead vehicle began to decelerate its speed to 80 km/h. Sometime later, the host vehicle would begin to change lanes to pass the lead vehicle. The time period from the beginning of the lead vehicle's deceleration to the initiation of changing lanes are defined as a preparation phase.

2.4 Secondary Task

A secondary cognitive task of "story talking" was given to each participant in order to simulate cognitive distraction. More concretely, four keywords (e.g., "Coffee, battery, paper, and water") were given once verbally during a cruising phase. Each of participants is asked to remember the four keywords to make up a story including the four keywords. He or she was instructed to talk the story from the start of a preparation phase. Each story talking must last at least 50 s.

2.5 Experimental Design

A single-factor within-subjects design was used in this experiment. Two conditions are distinguished. The first one was Driving-Only. In this condition, participants are asked to drive without performing the secondary task. The other one was Talking-during-Driving. Drivers are asked to perform the "story-talking" task during driving.

2.6 Measures

Checking behavior. Driver's eye-movement of looking at the side-view mirror was thought as one of important performances to checking traffic conditions, which was defined as *checking behavior*. Three types of frequency were distinguished on the checking behavior. We recorded the number of the checking behaviors in the last 12-second time period, denoted as $f_{12s}(t)$. If the increase of $f_{12s}(t)$ is delayed, it can be regarded as the delay in creating the intent to change lanes.

Secondly, a frequency, f_j, was calculated as:

$$f_j = \frac{\sum_i n_{ij}}{N_{L.C.}} \tag{1}$$

where, n_{ij} was the cumulative number of checking behaviors in 10-second time period *j*, for participant #*i*, and $N_{L.C.}$ was the total number of lane changes. The frequency was manipulated from the starting time moment of a preparation phase, T_0, and then the time period *j* was denoted as $[10(j-1), 10j)$. If the increase of f_j is at a late stage, it can be thought that a driver begins to be aware of a slower lead vehicle lately. If the value is low during the same period, it can also be regarded as the insufficiency of checking traffic conditions. A relative frequency F_j' for f_j was calculated from the initiation time moment, T_0', and then the time period *j* was denoted as $(-10(j+1), -10j]$. If the sharp increase of F_j' occurs before initiation of changing lanes, it can be considered that a driver creates the intent within a short time length and at a relative risk situation.

Driving behaviors. Headway distance (HW), time to collision (TTC) and time headway (THW) were also measured in this experiment. Those data were investigated at some key time points, such as T_0' and time points when the value of $f_{12s}(t)$ shifted from 0 to 1, from 1 to 2, and from 3 to 4. Let $T_{0\to1}$, $T_{1\to2}$ and $T_{3\to4}$ denote the three time points of the value shifts respectively.

Subjective ratings. To investigate mental workload dealing with the secondary task, a questionnaire was assigned to each participant during each trial. The questionnaire consisted of three items: time load, mental effort load and psychological stress load. The three items were based on the Subjective Workload Assessment Technique (SWAT) [13]. To obtain a detailed subjective rating to the mental workload, ten-level scales from 0 (not at all) to 10 (Extremely high) were used in the experiment instead of the original three-level scales.

A driving style questionnaire (DSQ) [14] sheet and a workload sensitivity questionnaire (WSQ) [15] were used to discuss driver style.

2.7 Procedure

Data were collected under the Driving-Only condition in the *first day* and under the Talking-during-Driving condition in the *second day*.

In the *first day*, participants were requested to read a written instruction of the experiment and to fill out the DSQ [14] sheet and the WSQ [15] sheet. After that, participants were given practice drives for approximately 10 min. All participants were instructed to complete sixteen scenarios in each of which vehicles on the passing lane drove differently. In the *second day*, participants received drives with performing the secondary task. Before the experimental drives, the participants had given the

explanation about the secondary task. All participants drove in the same sixteen scenarios under Talking-during-Driving.

2.8 Hypotheses

When the host vehicle approaches to the slower lead vehicle but a driver is distracted as a result of performing a cognitive activity, there are four possibilities as follows:

1. a driver fails to intend to change lanes,
2. he or she intends to change lanes, but

   2.1 the intent is changeable, i.e., it is easily that a driver changes his or her mind to initiate lane changes because of dynamic traffic environment,
   2-2. the intent is created lately, i.e., the timing of his or her determining to change lanes is delayed, or that traffic environment is at a relatively dangerous stage,
   2-3. the intent is created under insufficient checking on traffic conditions, i.e., the awareness on traffic conditions is not enough yet when he or she determines to change lanes.

Note that it is also possible that some cases involving ***Hypothesis-2-2*** and ***Hypothesis-2-3*** happen in the same trial.

3 Results

3.1 Initiation of Lane Changes

Note that data collection was unsuccessful for three of twenty participants (Two females, 35 and 30 years old; one male, 55 years old). This might reduce sample of data.

For all 272 lane changes under the Talking-during-Driving condition, there was only one trial failed to initiate changing lanes. An interview showed that the participant forgot to change lanes due to performing the secondary task. This result suggested that it was rarely that a distracted driver did not intend to change lanes when a host vehicle was closing to a slower forward vehicle. According to this result, ***Hypothesis-1*** may be rejected.

Values of *HW*, *TTC* and *THW* were measured at initiation time point, T_0', from the host vehicle to a lead vehicle and from a closet rear vehicle (on the passing lane) to the host vehicle. ANOVAs on those values showed no statistical differences (See, Fig. 2a-2c). We recorded cases in which lane-changing behavior initiated until after a last rear vehicle (on the passing lane) passed the host vehicle. Ratio of the cases to the whole trials was calculated for each participant. An ANOVA on the ratio showed no statistical difference between the two conditions (See, Fig. 2d). The result indicated that driver intent scarcely changed at initiations when a driver was distracted. Accordingly, ***Hypothesis-2-1*** may be rejected.

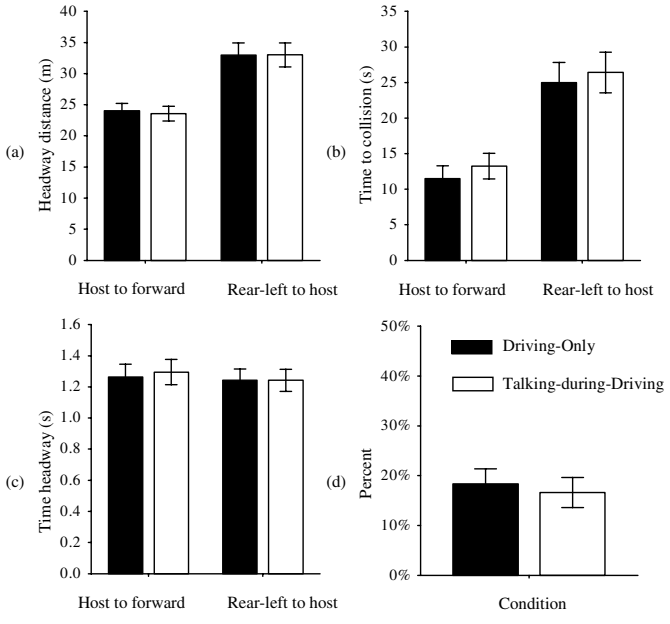


Fig. 2. (a-c). Driving behavior at initiations of changing lanes: *HW*, *TTC* and *THW*. (d). Ratio of cases, in which lane-changing behavior initiated until after a last rear vehicle (on the passing lane) passed the host vehicle, for each participant.

3.2 Checking Behaviors

Fig. 3a depicted a frequency of f_j' overall preparation phase under the two conditions. A two-way ANOVA showed a main effect of driving conditions, $F(1, 288) = 6.867$, $p < .01$ (Driving-Only Mean = 1.21, $SD = 1.15$; Talking-during-Driving Mean = 0.96, $SD = 0.99$). No statistical interaction was shown between driving condition and time order.

A *sharp* increase of the frequency, f_j', was indicated at the time period within approximately 20 s to initiations. This suggested that a driver created the intent to change lanes in a short time period just before initiating lane changes. The frequency of f_j' in Fig. 3a also showed that approximately 20-second delay happened in the crease of the frequency, f_j', under Talking-during-Driving conditions in comparison with Driving-Only. The result showed driver distraction might lead a delay of creating the intent. An accumulative frequency for f_j between the two conditions was illustrated in Fig. 4. A delayed tendency was showed from the start of trials. The results on the frequencies of f_j and f_j' showed that ***Hypothesis-2-2*** may be accepted from a viewpoint of time-order.

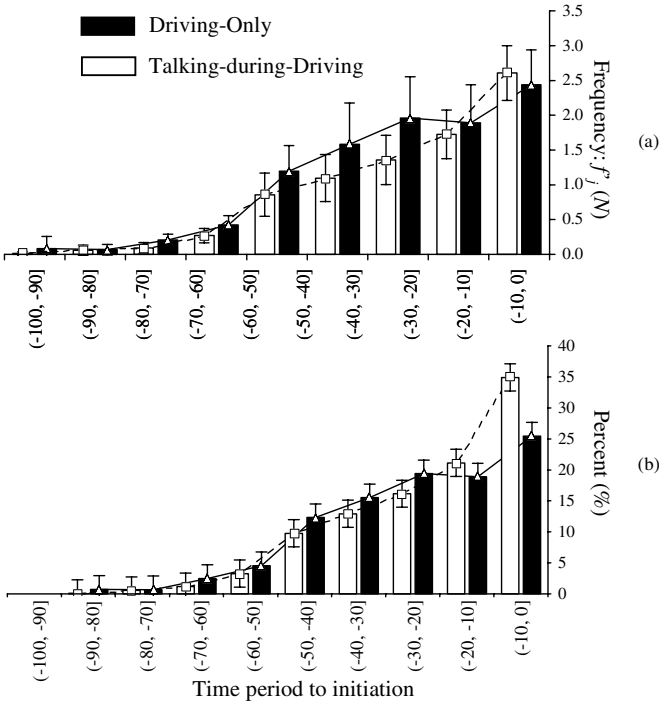


Fig. 3. Frequency, f'_j, and relative frequency, F'_j, of checking behavior to the initiations of changing lanes

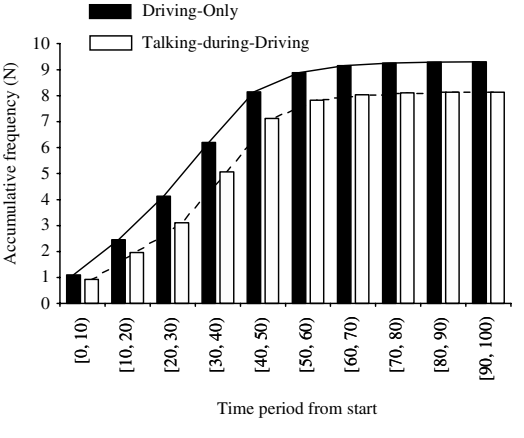


Fig. 4. The accumulative relative frequency of checking behavior from start of lane changing preparations

Statistic analysis on the frequency, f'_j, at each time period showed that the checking behavior was significantly less under conditions of Talking-during-Driving than Driving-Only from 20 to 70 s prior to initiations of changing lanes (See, Fig. 3a). The insufficient checking behavior could be seen from the starts, but also during the whole process (See, Fig. 4). A relative frequency of F'_j was illustrated in Fig. 3b. Significant interaction was shown between driving condition and time order, $F(8, 288) = 6.188$, $p = .00$. In particularly, approximately 44 percent of the checking behavior was operated during a process until before 20 s of the initiations under Talking-during-Driving conditions, but 56 percent under Driving-Only conditions. The above results also showed that checking traffic condition might be insufficient at the previous stage. These findings showed that ***Hypothesis-2-3*** may be also accepted.

3.3 Driving Behavior

Values of THW from the host vehicle to the forward vehicle were measured at time points of $T_{0\to1}$, $T_{1\to2}$ and $T_{3\to4}$, T'_0. Means were calculated for each participant's trails. An ANOVA on Driving-Only versus Talking-during-Driving showed a main effect on the driving conditions, $F(1, 128) = 7.050$, $p < .009$ (see, Fig. 5). Highly significant difference was indicated at $T_{3\to4}$ ($F(1, 16) = 8.652$, $p < .01$), and a near significant difference at $T_{1\to2}$ ($F(1, 16) = 3.896$, $p = .066$). The result revealed a remarkable delay from $T_{1\to2}$ to $T_{3\to4}$. This indicated that ***Hypothesis-2-2*** may be also accepted in consideration of traffic environment.

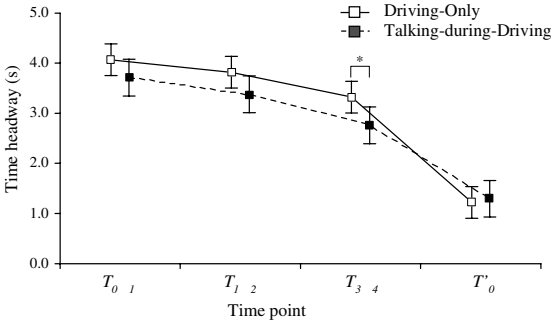


Fig. 5. Values of THW at time moments of $T_{0\to1}$, $T_{1\to2}$ and $T_{3\to4}$, T'_0 in the two conditions

3.4 Subjective Rating

Mean scores were calculated for the subjective ratings to three dimensions: time load, mental effort and psychological stress for each participant. An ANOVA was manipulated to the mean scores on the two conditions. Significant differences of the two conditions were shown on the three dimensions ($F(1, 32) = 4.758$, $p = .037$, see, Fig. 6). That is, higher mental workload was observed under conditions of Talking-during-Driving. This indicated that participants felt into a distracted situation in cases of performing the talking-story task in driving.

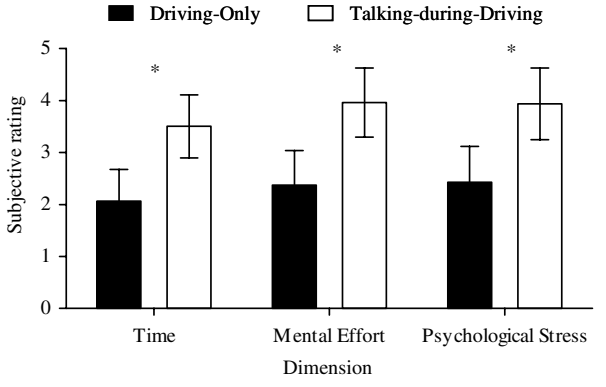


Fig. 6. Scores for three dimensions: time load, mental effort load and psychological stress load based on the Subjective Workload Assessment Technique (SWAT)

4 Discussion and Conclusion

This paper showed that two tendencies of lane-changing intent might occur in distractive situation. One was that a driver tended to make a decision to change lanes at a late stage. The other was that he or she might make the decision to change lanes with insufficient checking of the traffic environment. Additionally, the sharp increasing of checking behavior just before initiations suggested that a driver tried to make a decision in a short time period because he or she lost a lot of time at a previous stage due to driver distraction.

The above findings showed that a different method to detect driver intent should be used to a distracted driving. This is important to realize a safety support system, which adapts its function to driver situation, for avoiding a crash during a lane change. Further research is necessary to investigate on how to develop a detection method which can adapt to driver situation.

References

1. Wang, J., Knipling, R., Goodman, M.: The Role of Driver Inattention in Crashes: New Statistics From the 1995 Crashworthiness Data System. In: 40th Annual Proceedings of the Association for the Advancement of Automotive Medicine, October 7-9, pp. 377–392 (1996)
2. Eby, D.W., Kostyniuk, L.P.: Distracted-driving Scenarios: A Synthesis of Literature. In: 2001 Crashworthiness Data System (CDS) Data, and Expert Feedback. University of Michigan Transportation Research Institute, Ann Arbor (2004)
3. Inagaki, T., Itoh, M., Nagai, Y.: Support by Warning or by Action, Which is Appropriate under Mismatches between Driver Intent and Traffic Conditions? IEICE Trans. Fundamentals E90-A(11), 264–277 (2007)
4. Pentland, A., Liu, A.: Modeling and Prediction of Human Behavior. Neural Computation 11, 229–242 (1999)

5. Kuge, N., Yamamura, T., Shimoyama, O., Liu, A.: A Driver Behavior Recognition Method Based on a Drive Model Framework. In: Proc. the society of automotive engineers world congress 2000 (2000)
6. Oliver, N., Pentland, A.: Graphical Models for Driver Behavior Recognition in a Smartcar. In: Proc. the IEEE intelligent vehicles symposium 2002 (2002)
7. Salvucci, D.D., Siedlecki, T.: Toward a Unified Framework for Tracking Cognitive Processes. In: Proc. 25th Annual Conference of the Cognitive Science Society, pp. 1023–1028. Erlbaum, Mahwah (2003)
8. Salvucci, D.: Inferring Driver's Intent: A Case Study in Lane-change Detection. In: Proc. HFES 48th Annual Meeting, pp. 2228–2231 (2004)
9. Zhou, H., Itoh, M., Inagaki, T.: Eye Movement-Based Inference of Truck Driver's Intent of Changing Lanes. SICE Journal of Control, Measurement and System Integration (in press)
10. Zhou, H., Itoh, M., Inagaki, T.: Influence of Cognitively Distracting Activity on Driver's Eye Movement during Preparation of Changing Lanes. In: Proc. SICE Annual Conference 2008, pp. 866–871 (2008)
11. Zhou, H., Itoh, M., Inagaki, T.: Effects of Cognitive Distraction on Checking Traffic Conditions for Changing Lanes. In: Proc. HFES 53rd Annual Meeting (to appear)
12. Akamatsu, M., Sakaguchi, Y., Okuwa, M.: Modeling of Driving Behavior when Approaching an Intersection based on Measured Behavioral Data on an Actual Road. In: Proc. HFES Annual Meeting 2007, pp. 1895–1899 (2007)
13. Reid, G.B., Nygren, T.E.: The Subjective Workload Assessment Technique: A Scaling Procedure For Measuring Mental Workload. In: Hancock, P.A., Meshkati, N. (eds.) Human Mental Workload, pp. 185–218. Elsevier, Amsterdam (1988)
14. Ishibashi, M., Okuwa, M., Doi, S., Akamatsu, M.: Indices for Characterizing Driving Style and their Relevance to Car Following Behavior. In: Proc. SICE Annual Conference 2007, pp. 1132–1137 (2007)
15. Ishibashi, M., Okuwa, M., Doi, S., Akamatsu, M.: Characterizing Indices of Driver's Workload Sensitivity and tieir Relevance to Route Choice Preferences. Transactions of Society of Automotive Engineers of Japan 39(5), 169–174 (2009)

Effects of Missed Alarms on Driver's Response to a Collision Warning System According to Alarm Timings

Genya Abe[1], Makoto Itoh[2], and Tomohiro Yamamura[3]

[1] Japan Automobile Research Institute, 2530 Karima Tsukuba 305-0822, Japan
[2] University of Tsukuba, 1-1-1 Tennoudai Tsukuba 305-8573, Japan
[3] Nissan Motor Co., Ltd. 1-1 Morinosato-Aoyama Atsugi 243-0123, Japan
agenya@jari.or.jp, itoh@risk.tsukuba.ac.jp,
t-yamamura@mail.nissan.co.jp

Abstract. By using a driving simulator, two kinds of alarm timing were compared to investigate how drivers respond to missed alarms regarding different alarm timings: (1) an alarm was given based on ordinary braking behaviour for the individual; alarm timing T, (2) an alarm was given by using an particular alarm trigger logic (Stopping Distance Algorithm) as a common timing for all drivers; alarm timing S. Alarm timing S was earlier than alarm timing T in this study. The results showed that compared to alarm timing T, alarm timing S induced earlier braking behaviour independent of degree to which an imminent collision was critical. However, effects of a missed alarm on braking behaviour may be mitigated by alarm timing T, compared to alarm timing S. Moreover, it is possible that effect of missed alarm on driver's trust may vary according to alarm timings and the number of experience of missed alarms.

Keywords: Forward collision warning systems, driver trust, missed alarms braking behaviour.

1 Introduction

The reduction of traffic accidents is an important research goal, and rear-end collisions are one of the most common types of the accidents. The Forward Collision Warning Systems (FCWS) may be of great potential benefit to drivers who do not pay sufficient attention to driving, because these systems may reduce the number of traffic accidents [1], [2]. However, rear-end collisions tend to occur in time-critical situations; therefore, the expected results depend on when the alarm is triggered [3]. When determining the alarm timing, the Stopping Distance Algorithm (SDA) is often used [4]. SDA is so termed because a warning distance is defined based on the difference between the stopping distances of the lead and following vehicles. It is known that the braking response time may be improved when several parameters of SDA are manipulated to provide drivers with early alarms, compared the response times in non-assisted driving conditions [5].

However, it is not obvious that early alarms achieve safe driving. Drivers need to repeatedly implement situational recognition, decision-making, and action implementation

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 245–254, 2009.

to ensure safe driving. If drivers hear alarms before they recognise risky situations, then decide to implement actions to avoiding critical situations, then drivers decisions might be replaced by alarms. Moreover, if drivers are excessively adapted to such early alarms, they might assume that it should not be necessary to immediately take appropriate actions to avoid collisions as long as a alarm is not triggered. As a result, there is the possibility that the driver's braking behaviour might be delayed if "missed alarms" which should have been triggered but were not (including alarms which drivers did not notice for some reason also) occur. However, it is known that the degree to which braking behaviour is impaired as a result of missed alarms may vary depending on the alarm timing [6].

There also may be individual differences in the timing of braking behaviour before imminent collision situations [7]. Thus, if the driving characteristics for an individual would be considered for determining an alarm timing, which is called adaptive alarm timing, then it would be possible to minimize differences in the timing between the driver's collision avoidance behaviour and alarm presentations. Consequently, it is possible to prevent a driver's excessive reliance on alarms, thus resulting in drivers' robust behaviour toward missed alarms.

In the present study, two experiments were conducted. In the first experiment (pilot study), the variation of braking behaviours of individual drivers in response to imminent collision situations in which a lead vehicle suddenly decelerated was explored. In the second experiment, two issues were investigated: the differences in alarm effectiveness and the differences of braking behaviour due to missed alarms with difference timings. Here, two kinds of alarm timings were considered: adaptive alarm timing, based on the ordinary braking behaviour of the individual, and non-adaptive alarm timing, which uses trigger logic (Stopping Distance Algorithm) as a common timing for all drivers. Both alarm timings were determined based on the pilot study.

2 Experiment I

2.1 Method

Apparatus. This experiment was conducted with a driving simulator owned by the Japan Automobile Research Institute. The simulator has six-degrees-of-freedom motion and uses complex computer graphics to provide a highly realistic driving environment. The simulated horizontal forward field of view was 50 degrees and the vertical field of view was 35 degrees.

Participants and experimental tasks. Twenty-four participants (Mean age = 25.5 years, SD = 8.0 years) took part in the experiment. All of the participants were licensed drivers. Each participant was given two tasks in the experiment. One was to maintain the speed of the vehicle at a target speed of 80 km/h, and the other task was to follow a lead vehicle while avoiding a rear-end collision.

Experimental design. The participants were divided into two groups according to time headway conditions and each participant experienced two different lead vehicle decelerations, namely 5.88 m/s^2 (high time criticality) and 3.92 m/s^2 (low time criticality). Twelve participants were assigned in to each group. Two values, 2.0 s and 1.4 s, were considered in the experiment as the time headway conditions while following a vehicle. For one lead vehicle deceleration, nine potential collision events, in which the lead

vehicle decelerated at a constant rate with illuminated brake lights but with irregular time intervals, could occur. The order of the experimental conditions was controlled to reduce the potential effects of the experimental conditions on driver behaviour.

Procedure. All participants gave informed consent and were instructed regarding the task requirements. Each participant was allowed a 10-min practice session to familiarize themselves with the simulator and experience the lead vehicle decelerations that would be involved in the experiment so that they could learn how imminent collision situations would occur. The drivers were directed to use the brakes instead of changing lanes to avoid a collision with the lead vehicle. The potential collision events were repeatedly presented at irregular time intervals to prevent drivers from predicting their occurrence. Each participant had a 3-min break after finishing nine events for one deceleration condition.

Dependent variables. Two dependent variables describing the collision avoidance behaviour of the driver were recorded:

- Accelerator release time: This is the time period between the braking event and the release of the accelerator by the following vehicle.
- Braking response time: This is the time period between the braking event of the lead vehicle and application of the brakes by the following vehicle.

2.2 Results

Main purpose of this experiment was to obtain data of ordinary braking behaviour for individual drivers for determining alarm timings. Here it was briefly summerised how braking behaviour for the individual differ according to driving conditions. Specifically, Figs. 1 and 2 show that the mean values of the accelerator release time and braking response time for all drivers for the 1.4 s time headway and 3.92 m/s^2 deceleration of the lead vehicle and for the 2.0 s time headway and 5.88 m/s^2 deceleration of the lead vehicle, respectively. The braking behaviour represented by both variables varies according to individual drivers. Also, it seems that there were differences in braking behaviour depending on experimental conditions. A similar tendency was observed for both variables regardless of the driving conditions.

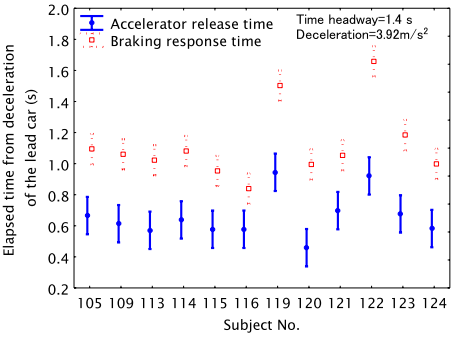


Fig. 1. Individual differences in braking behaviour (time headway=1.4 s, Deceleration=3.92 m/s^2

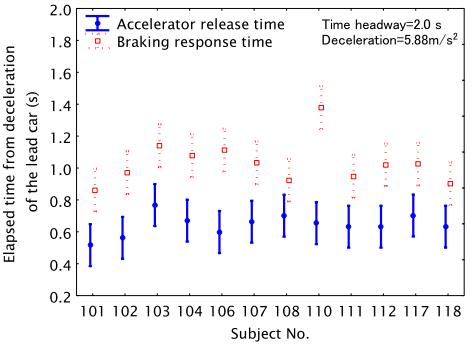


Fig. 2. Individual differences in braking behaviour (time headway=2.0 s, Deceleration=5.88 m/s^2

3 Experiment II

The purpose of this experiment was to investigate two issues: how differences in alarm timings influence the driver's response to alarms, and how missed alarms influence the braking behaviour in response to alarm timings.

3.1 Method

Apparatus. All materials used in this experiment were the same as those used in the pilot study. In this experiment, FCWS was introduced with a simple auditory beep.

Participants and experimental tasks. In this experiment, the participants were the same as those in the experiment I. All drivers were required to perform the same tasks as in the experiment I, except that a forward collision warning system was available.

Experimental design. The participants were divided into two groups in the same manner as in the pilot study, namely, according to 1.4 s time headway or 2.0 s time headway. That is, each participant experienced the same time headway condition as in the pilot study. The participants who were assigned to the 1.4 s time headway group experienced a deceleration of the lead vehicle of 5.88 m/s^2 at a driving speed of 80 km/h. The participants who were assigned to the 2.0 s time headway group experienced a deceleration of the lead vehicle of 3.92 m/s^2 at a driving speed of 80 km/h.

Each participant of each group was subjected to 14 potential collision events in which the lead vehicle decelerated at 5.88 m/s^2 or 3.92 m/s^2. An alarm was correctly given for each potential collision event, except for the fifth event and the 13th event (Fig. 3). In these two events, missed alarms occurred; that is, alarms were not triggered even when a potential collision event occurred.

As for alarm timings, each participant experienced adaptive alarm timing or non-adaptive alarm timing. Moreover, half of the participants who were assigned into each group experienced adaptive alarm timing and the remaining participants experienced non-adaptive alarm timing. For each group, the order of the alarm timings was counterbalanced according to the number of participants in order to reduce the effects of the experimental conditions on driver behaviour.

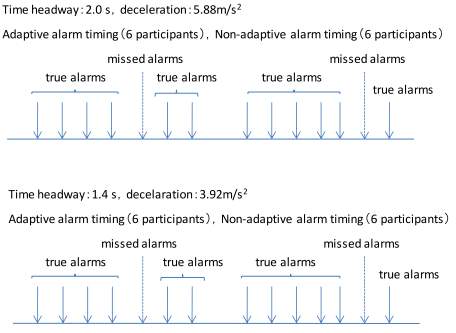


Fig. 3. Experimental conditions for trial by trial

The setting of alarm timings

Adaptive alarm timing. Adaptive alarm timing was determined for each participant based on the data obtained in the pilot study:

$$\text{Adaptive alarm timing} = (\text{MART} + \text{MBRT})/2$$

MART and MBRT are the median values of the accelerator release time and the braking response time, respectively, for nine trials for each driving condition obtained in the pilot study (Fig. 4). Table 1 constitutes a summary of the values of adaptive alarm timing which were calculated by using the adaptive alarm timings for all drivers according to the driving conditions.

Non-adaptive alarm timing. The non-adaptive alarm was issued at the same time for all participants based on the Stopping Distance Algorithm, which is one of the preventative alarm trigger logics for FCWS. The SDA is so termed because a warning distance is defined based on the difference between the stopping distances of the leading and following vehicles. This algorithm has three parameters: reaction time to the alarm (*RT*), deceleration of the leading vehicle (D_l) and deceleration of the following vehicle (D_f). *RT* is the assumed reaction time of the driver of the following vehicle. The warning distance (D_W) is found on the basis of these parameters along with the velocities of the leading (V_l) and following (V_f) vehicle, as follows.

$$D_w = V_f RT + {V_f}^2 \Big/ 2D_f - {V_l}^2 \Big/ 2D_l \tag{1}$$

An alarm is trigged when the current headway distance is less than the warning distance (D_W). In this experiment, the values of the three parameters were determined by considering that alarms would be triggered at the timing of the minimum accelerator release time for all participants in response to the driving conditions. Table 2 contains a summary of the values of each parameter and the average alarm timings in response to the driving conditions.

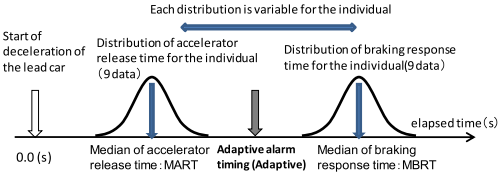


Fig. 4. The concept of adaptive alarm timing based on braking behaviour

Table 1. Summary of values of adaptive alarm timing in response to driving conditions

me e w y) e e e o		A m m)	
		M M	A e e
	5 88		8
		5	88
	5 88	5 4	5
4		8	8

Table 2. Summary of values of alarm parameters and non-adaptive alarm timing in response to driving conditions

h y) [2] 2 2 g

Procedure. All participants were instructed regarding the task requirements. Each participant was allowed a 10-min practice session in order to familiarize themselves with the simulator and to experience the lead vehicle decelerations. The participants were directed to use only the brakes instead of changing lanes to avoid a collision with the lead vehicle. Moreover, they were instructed that FCWS provided alarms when potential collision events were happening, although FCWS did not always work correctly, that is, there was a possibility that alarms would not be triggered due to system failures. In the practice session, the participants did not experience missed alarms.

The potential collision events repeatedly occurred at irregular time intervals to prevent drivers from predicting their occurrence.

Dependent variables. The following variables were collected.

- Braking response time: This is the time period between the braking event of the lead vehicle and application of the brakes by the following vehicle.
- Trust in the system: The subjective estimation of the driver's trust in the systems was obtained using an 11-point rating scale, in which 0 indicated "none at all" and 10 indicated "completely." All subjective measurements were recorded by using the following question; How much did you trust the system?, immediately after the braking event, in which the lead car suddenly decreased its speed.

3.2 Results

Braking response time. Fig. 5 shows the mean values of braking response times for each potential collision event in response to the alarm timings. As can be seen in this figure, the non-adaptive alarm timing induced a swift response of braking, compared to that of adaptive alarm timing when true alarms occur. The result suggests that non-adaptive alarms are presented earlier than adaptive alarms, resulting in timely implementation of the brakes.

With respect to the trial in which a missed alarm happened for the first time (trial 5), independent of the type of alarm timing, the braking response time was dramatically delayed compared to the braking response times obtained in other trials where true alarms occurred. Some interesting phenomena appeared for the trial in which there was a second missed alarm (trial 13). Specifically, for non-adaptive alarm timing, a missed alarm still induced a longer response time for braking, compared to the braking response times for the true alarm conditions. However, for adaptive alarm timing, the braking response time for the second missed alarm trial was almost the same as the braking response time for the true alarms conditions, indicating that drivers managed to apply the brakes in the same manner as they did for true alarms, even if a missed alarm happened. A two-way ANOVA of trials and alarm timings with repeated measures on the factor of trials showed a significant interaction between factors, $F(13, 273)=2.17$, $p<0.05$ and a Tukey's HSD post-hoc test showed that there was a significant difference in braking response time for non-adaptive alarm timing between trial 12 and trial 13 ($p<0.01$) but there was not a significant difference in braking response time for adaptive alarm timing between trial 12 and trial 13.

These results suggest that, by experiencing a missed alarm once, drivers become cautious about malfunctions, including missed alarms. In other words, there is a possibility that awareness features such as alarms must always be triggered when imminent collision situations happen, is diminished. As a results, drivers who experienced adaptive alarm timings exhibited robust behaviour toward subsequent missed alarms. This was not the case with drivers who experienced non-adaptive alarm timing. This can be explained from the braking response times which were obtained in the trials, in which true alarms occurred for non-adaptive alarm timing. That is, true alarms for non-adaptive alarm timing have the potential to lead to early braking response time since a delayed response of braking may become obvious when missed alarms happen compared to the response of adaptive alarm timing in the same situation.

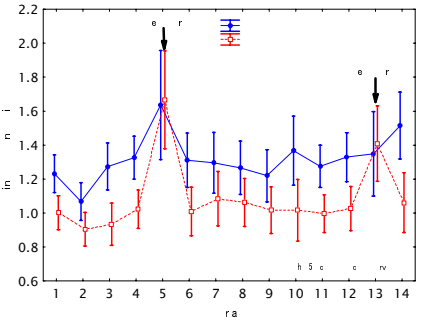


Fig. 5. Braking response time for each trial in response to alarm timings

Effects of missed alarms on driver`s trust according to alarm timings. Fig. 6 shows the mean values of trust ratings for each trial in response to the alarm timings. As can be seen in this figure, trust ratings (an 11-point rating scale) for both alarm timings for the first four trials were relatively high (above 5), indicating that adaptive and non-adaptive alarm timings are reasonably trustworthy as collision warnings. For the 5th trial, in which drivers experienced a missed alarm for the first time, trust ratings dramatically decreased for both alarm timing conditions. Next, seven trials between the 6th and 12th trials were considered to assess how trust ratings changed after experiencing the first missed alarm. It can be said that the values of trust ratings for adaptive and non-adaptive alarm timings reached at the similar levels obtained before the first missed alarm by the 12th trial.

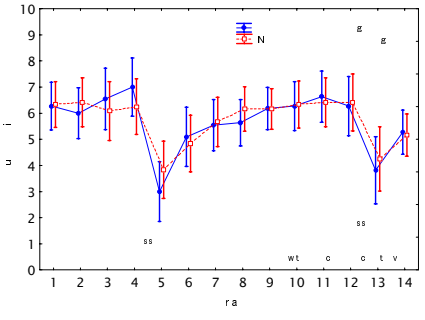


Fig. 6. Trust ratings for each trial in response to alarm timings

One interesting result was obtained from trial 13, in which a second missed alarm occurred. That is, there were differences in the impacts of a missed alarm on the driver's trust in the system according to the alarm timings. Without a doubt, the second missed alarm decreased the trust ratings for both alarm timings. However, it seems that for adaptive alarm timing, the degree to which trust ratings decreased due to the second missed alarm improved compared to the degree to which the trust ratings decreased was due to the first missed alarm. In contrast, for non-adaptive alarm timing, decreased trust ratings for the first and second missed alarms seemed to be at similar levels. In order to confirm the impact of missed alarms on driver's trust in the FCWS in response to alarm timings, a variable relevant to the rates of decreased trust for missed alarms was introduced, as follows.

Rate of decreased trust for missed alarms = (trust rating for immediately before a missed alarm - trust rating for a missed alarm) / (trust rating for immediately before a missed alarm)

Table 3 contains a summary of the rates of deceased trust for missed alarms in response to alarm timings for the first missed alarm (trial 5) and the second missed alarm (trial 13). For non-adaptive alarm timing, there were no significant differences in trust ratings between the first and the second missed alarm. For the adaptive alarm timing, however, compared to decreased trust caused by the first missed alarm, the second missed alarm produced gentle reduction of trust ratings, $F(1,11) = 6.11$, $p < 0.03$.

The results indicate that there is a possibility that deceased trust caused by missed alarms may vary in response to alarm timings.

Table 3. Effect of alarm timing on decreased trust for missed alarms

	o	
	h	h o
v	8	8))
No v	)	)

3.3 Discussion

The present study focused on the alarm timing for FCWS and investigated driver response to two kinds of alarm timing. One was individually determined based on braking behaviour for the individual (adaptive alarm timing) and the other was for one common trigger logic (SDA) for all drivers (non-adaptive alarm timing). The impact of missed alarms on braking behaviour on the response to alarm timings was also investigated.

When drivers experienced a missed alarm for the first time, compared to the presentation of true alarms, the response time to the brakes was impaired by the missed alarm for both alarm timings. With respect to trust ratings, moreover, the first missed alarm resulted in dramatically decreased trust ratings independent of alarm timings. These results suggest that the driver's trust in the system may be impaired when the braking response time is delayed in imminent collision situations, when an alarm is expected.

As for a missed alarm happening for the second time, drivers who experienced adaptive alarm timing managed to apply the brakes in the same manner as when true alarms were presented. From this result, it can be said that adaptive alarm timing has the potential to diminish the impairing influences of missed alarms on braking behaviour earlier than this would occur with non-adaptive alarm timing. Consequently, it is possible for adaptive alarm timing to restrain the decreased trust ratings for missed alarms when the impact of missed alarms on driver behaviour is relatively small.

FCWS should provide drivers with sufficient time to avoid imminent collisions. However, it is necessary to consider how alarm timing should be determined in order to minimize the effects of missed alarms on braking behaviour. For this reason, it is reasonable to conclude that adaptive alarm timing may be useful.

4 Concluding Remarks

The results support the following conclusions.

1. Both adaptive alarm timing and non-adaptive alarm timing maintain reasonable driver's trust in FCWS in the case of driving conditions used in the present experiments.
2. Adaptive alarm timing mitigates the impacts of missed alarms on driver behaviour compared to the impact for non-adaptive alarm timing.

3. It is likely that the effects of missed alarms on driver's trust in FCWS vary in response to not only differences in alarm timing but also to the number of experiences of missed alarms.

Acknowledgments. This work was supported by Grant-in-Aid for Scientific Research(A) 18201031.

References

1. Alm, H., Nilsson, L.: Incident warning systems and traffic safety. A comparison between the PORTICO and MELYSSA test site systems. Transportation Human Factors 2, 77–93 (2000)
2. Ben-Yaacov, A., Maltz, M., Shinar, D.: Effects of an in-vehicle collision avoidance warning system on short- and long-term driving performance. Human Factors 44, 335–342 (2002)
3. Janssen, W.H., Alm, H., Michon, J.A., Smiley, A.: Driver support. In: Michon, J.A. (ed.) Generic Intelligent Diver Support, pp. 53–66. Taylor and Francis, London (1993)
4. ISO/TC 204/WG 14.: Road Vehicles-Forward Vehicle Collision Warning Systems-Performance Requirements and tests Procedures, SO/DIS 15623 (1994)
5. Lee, J.D., McGehee, D.V., Brown, T.L., Reyes, M.L.: Collision warning timing, driver distraction, and driver response to imminent rear-end collisions in a high-fidelity driving simulator. Human Factors 44, 314–334 (2002)
6. Abe, G., Richardson, J.: The influence of alarm timing on driver response to collision warning systems following system failure. Behaviour and Information Technology 25, 443–452 (2006)
7. Abe, G., Itoh, M.: Exploring appropriate alarm timing for a driver-adaptive forward collision warning system. In: De Ward, D., Flemisch, F.O., Lorenz, B., Oberheid, H., Brookhuis, K.A. (eds.) Human Factors for assistance and automation, pp. 103–115. Shaker Publishing, Maastricht (2008)

SLAM Estimation in Dynamic Outdoor Environments: A Review

Zheyuan Lu, Zhencheng Hu, and Keiichi Uchimura

Graduate School of Science and Technology, Kumamoto University, 2-39-1, Kurokami, Kumamoto, Japan, 860-8555

Abstract. This paper gives a review of the literature on Simultaneous Localization and Mapping (SLAM). SLAM has been intensively researched in recent years in the field of robotics and intelligent vehicles, many approaches have been proposed including occupancy grid mapping method (Bayesian, Dempster-Shafer and Fuzzy Logic), Localization estimation method (edge or point features based direct scan matching techniques, probabilistic likelihood, particle filter). In this paper, we classify SLAM approaches into three main categories: visual SLAM, Lidar SLAM and sensor fusion SLAM, while visual and lidar can also contain many types and levels, such as monocular camera, stereovision, laser scanner, radar and fusion of these sensors. A number of promising approaches and recent developments in this literature have been reviewed in this paper. To give a better understanding of performance difference, an implementation of Lidar SLAM is presented with comparative analysis result.

1 Introduction

The simultaneous localization and mapping (SLAM) problem for a mobile robot is to build a consistent map of the environment and at the same time determine its location within this map [(Hartley 04)]. The solution to the SLAM problem has been seen as the fundamental in making a robot truly autonomous [(Bosse 02)]. One of the common assumptions used in SLAM is that the unknown environment is assumed to be static containing only rigid, stationary objects. Non-rigid or moving objects are processed as outliers and filtered out.

In the robotics literature, SLAM has been seen as the prime tool to solve the so-called DAMTO (detection and tracking of moving objects) problem concurrently. While SLAM provides the vehicle with a map of static parts of the environment as well as its location in the map, DATMO allows the vehicle being aware of dynamic entities around, tracking them and predicting their future behaviors. It is believed that if we are able to accomplish SLAM reliably in real time, we can detect every critical situation to warn the driver in advance and this will certainly improve driving safety and can prevent traffic accidents.

Basically, SLAM approaches have been proposed including the process of occupancy grid mapping, and the process of localization estimation. Occupancy grid(OG) mapping, which is also called Map-learning, is the process of memorizing the data acquired by the robot during exploration in a suitable representation. Bayesian,

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 255–267, 2009.

Dempster-Shafer and Fuzzy Logic are the typical methods for occupancy grid mapping. Localization is the process of deriving the current position of the robot within the map, such as point and edge features based direct scan matching techniques (ICP, RANSAC), probabilistic likelihood, particle filter and Extended Kalman Filter (EKF).

Sensor selection also takes a critical role in SLAM process. In this paper SLAM approaches are classified into three main categories: visual SLAM, Lidar SLAM and SLAM by sensor fusion. While visual and Lidar can also contain many types and levels, such as monocular camera, stereovision, laser scanner, radar and fusion of these sensors. This paper reviews a number of promising approaches and provides an overview of recent developments in this domain. The emphasis of this paper is to discuss the various methods with different sensor(s) data to estimate SLAM and global localization, and provide a comprehensive performance analysis among the common SLAM approaches, like computation speed, accuracy and cost.

The reminder of this paper is organized as follows. Section 2 presents a general review on occupancy grid mapping approaches and Section 3 describes the method of localization. Practical analysis of different SLAMs is shown in Section 4. Section 5 concludes the paper.

2 SLAM (1) – Occupancy Grid Mapping

In the past two decades, occupancy grid maps have become a dominant paradigm for environment modeling in mobile robotics. Occupancy grid maps are spatial representations of robot environments. Once acquired, they enable various key functions necessary for mobile robot navigation, such as localization, path planning, collision avoidance.

2.1 Techniques for Building Occupancy Grid Map

To build a map, is to take a number of sensor readings, a sensor reading being a discrete-time sample, and integrate them into a map. This is not as straight forward as it might sound as there are many reasons why robotic mapping is a hard problem.

- **Noisy Sensors.** For example, laser and stereovision are sensitive to differences in lighting, some surfaces does not reflect sound well enough to be sensed by sonar.
- **Sensor integration.** Generated maps that are inconsistent with the data. This problem is due to the fact that existing algorithms decompose the high-dimensional mapping problem into many one-dimensional estimation problems-one for each gird cell-which are then tackled independently [(Thrun 03)].
- **High computational complexity.** Time complexity is also a concern as robotic mapping algorithms are supposed to work in real time.

The first OG map algorithm was introduced by [(Moravec 85)], its basic idea is simple: Represent the prat of the world you want to map with a grid. When you observe an obstacle, mark the cells covered by that obstacle as occupied. This OG algorithm was implemented and anumber of experiments were conducted to investigate how it would perform given different types of sensor noise [(Deans 05)] [(Groecke 07)] [(Tardif 08)].

Nowadays many efforts in mobile robotics are directed to develop some kind of "uncertainty calculi" techniques for recovering spatial information from obtained sensor data. In the literature, three different uncertainty calculi techniques for building OGs of an unknown environment based on sensor information are discussed. These techniques are based on Bayesian theory (probabilistic approach) [(Moravec 01)] [(Elfes 92)], Dempster Shafer theory of evidence (evidence theoretic approach) [(Ribo 01)][(Gambino 96)], and fuzzy set theory (possibility approach) [(Oriolo 99)] [(Ribo 01)] [(Gambino 96)].The probabilistic approach is the most widely found in mobile robotic literature [(Moravec 85)] [(Murray 00)] [(Elfes 92)] [(Thrun 03)] [(Fox 99)]. The Bayesian method rules the greatest part of the work related to the probabilistic sensor fusion in building OGs. This attraction stems from the property of the Bayes' updating rule which facilitates recursive and incremental schemes [(Ribo 01)]. However, in order to avoid huge calculation processes, one must assume that the cell states are independent. It has been observed that this assumption may induce large errors in the presence of even a slight degree of dependence between the random variables, this is exactly the case for map building, since the occupied cells are not evenly distributed, but concentrated in clusters (obstacles). Moreover, the prior probabilities needed to initialize the field are typically estimated with the maximum entropy assumption, namely by regarding emptiness and occupancy as equiprobable [(Oriolo 99)]. As a consequence, the convergence of the Bayesian updating procedure towards and acceptable characterization of the OG requires a large number of measures.

The articles [(Thrun 03)] [(Fox 99)] describe algorithms for acquiring occupancy grid maps with mobile robots, which rely on the probabilistic approach. These algorithms employ the expectation maximization (EM) algorithm for searching maps that maximize the likelihood of the sensor measurements. The approach presented in [(Thrun 03)] relies on a statistical formulation of the mapping problem using forward models. Experimental results are presented, which are obtained using a RWI B21 robot equipped with 24 sonar sensors. The disadvantages of this approach are an apparent increased sensitivity to changes in the environment, and a need to go through the data multiple times, which prohibits its real-time application. Moreover, in [(Duckett 00)], it is pointed out that the EM-based techniques suffer from a high computational complexity. Besides, EM is not guaranteed to converge to a global optimum.

2.2 Comparative Analysis of Building OGs

The well known techniques of OGs building were experimented and comparisons were performed in [(Ribo 01)] [(Gambino 96)]. It was shown that the fuzzy logic approach has the best performance, especially in the case of unstructured environments. The experimental results indicated that the method based on fuzzy logic is more robust with respect to the occurrence of false reflections in the measuring process [(Oriolo 99)]. Also it was shown that stochastic techniques based on Bayesian updating are very sensitive to the occurrence of outliers in the measuring process. In [(Oriolo 99)], the authors integrated topological and grid-based representations for the purpose of constructing globally consistent metric models of large, real world environments. Using their approach, a sonar-equipped mobile robot is able to construct detailed models of large environments [(Duckett 00)]. The approach includes an off-line algorithm for constructing a global

grid map. The algorithm is based on the fuzzy logic approach for OGs building, given in [(Oriolo 99)].

In [(Ribo 01)]'s work, The building algorithm of the probabilistic approach is the fastest and has minimum memory consumption. The processing time for the evidence theoretic and possibilistic approaches are approximately 1.5 times higher than the probabilistic approach. Concerning the stored OGs, the possibilistic approach is the one with the highest memory requirement. This could be a drawback and might require compact representations, e.g., quad trees to store the processed OGs.

To sum up, within a general framework of navigation in mobile robotics, the possibilistic approach may produce the most suitable OGs thanks to its robustness with respect to outliers. The probabilistic and the evidence theoretic approaches produce good results in certain cases, but their performances with respect to outliers are very poor. However, a possible criticism to the possibilistic approach is the conservative behavior of its building process. It can infer the loss of information in certain parts of the explored environment and induce the emergence of some artifacts in the OG.

3 SLAM (2) – Localization Estimation

Localization estimation approaches can also be divided into three major categories: Visual SLAM, Lidar SLAM and SLAM by sensor fusion.

3.1 Visual SLAM

Monocular Camera. SLAM has been most often performed with other sensors than regular cameras, however in the last years successful results have been obtained using single cameras alone. Successful results with only a single camera have been obtained in the last decade using both perspective and omnidirectional cameras. Some of these works will be reviewed in this Section.

Practical real-time monocular SLAM was first demonstrated by [(Davison 03)], who uses the Extended Kalman Filter (EKF), a mainstay of SLAM literature. He resolves the problem of real-time operation by careful maintenance of the map to ensure that it is sparse but sufficient, and by using the map uncertainty to guide feature matching. More recently, [(Pupilli 05)] have demonstrated real-time camera tracking using a particle filter, which provides a good robustness, but theirs is predominantly a tracking system; its mapping ability is currently rudimentary, which restricts its range of applications. [(Eade 06)] have developed a system based on the FastSLAM algorithm, which combines particle filtering for localisation with Kalman filtering for mapping. FastSLAM has the advantage that it scales better with the number of features, but the absence of an explicit full covariance matrix can make loop-closing more difficult.

Straight lines are common in man-made environments and are arguably better features to track than points. Described just by a step change in intensity (which does mean that they lack discrimination), they are trvially stable under a wide range of viewing angles, and a number of measurements can be made along their length to localise them accurately. As a result, many camera-tracking systems have used line features. In perhaps the earliest work in visual SLAM using lines, [(Bosse 04)] use a single omnidirectional camera to detect and track parallel lines, both with reasonable results.

Point and line features are complementary in a camera localisation system: point features provide good discrimination, but are view-dependent, while line features are robust to viewing changes, but are more fragile. This idea has been studied recently by [(Rosten 05)], who bootstrap line-tracking (using a prior accurately-known three dimensional model) with detected point features. They shows that monocular tracking with the fusion of point features and line features, and model-building with both types of feature, can be performed within a standard SLAM framework.

Stereovision. Stereovision has been employed for map building in decades, e.g., active stereo approach with spot lighting [(Scaramuzza 09)], 3D mapping from stereo range data with planar modeling assumption [(Moravec 01)], and 3D SLAM based on feature point matching [(Iocchi 00)] [(Elinas 06)] [(Garcia 04)]. The most popular approach in recent years is the feature-point based one, in which the camera motion is estimated with feature-point matching between consecutive frames, and 3D point clouds are generated based on the estimated camera motion. As mentioned, however, the SLAM process is unstable in non-textured environments, where sufficient corner-like features cannot be extracted. Since many man-made environments are non-textured, the importance of alternative feature forms such as lines is indicated in [(Se 01)] [(Herath 07)]. For this reason, in [(Tomono 09)], Tomono computes 3D points from the edge points detected in a stereo image pair, and then estimates the camera motion by matching the next stereo image with the 3D points. The proposed method estimates camera poses and builds detailed 3D maps robustly by aligning edge points between frames using the (Iterative Closest Points) ICP algorithm.

Lines have also been used for some time in SLAM systems.In perhaps the earliest work in visual SLAM using lines, [(Ayache 88)] used a stereo pair of calibrated cameras to directly extract the three-dimensional location of line segments and filtered these within an EKF SLAM framework. More recently, [(Dailey 05)] described the application of ”FastSLAM” to the problem of estimating a map from observations of 3D line segments using a trinocular stereo camera rig.

3.2 Lidar SLAM

We classify the Lidar SLAM into two main categories: scan matching approach and probabilistic approach (maximum likelihood estimation approach). Scan matching approaches also can be categorized based on their association method such as Feature-to-Feature, Point-to-Feature and Point-to-Point.

Feature-to-Feature matching approaches should have the shortest run-time, since by these approaches hundreds of range points are reduced to dozens of features. For most indoor applications, line segments [(Gutmann 00)], corners [(Lingemann 04)] and other simple geometrical features are rich and easy to detect. [(Zhao 01)] picked a site that is similar to indoor environments and employed feature-to-feature approaches to construct an urban map successfully. [(Guivant 00)] used intensity (reflectance) of laser signal and geometrical primitives to define and detect features. Their approaches are still limited to some specific environments or conditions.

Point-to-Feature approaches, such as one of the earliest by [(Cox 91)], the points of a scan are matched to features such as lines. The line features can be part of a predefined

map. Features can be more abstract as in [(Biber 03)], where features are Gaussian distributions with their mean and variance calculated from scan points falling into cells of a grid. Basically, Feature-to-Feature approaches try to use less information to represent the raw data in order to speed up algorithms. If features cannot be detected robustly and contain some uncertainties, the whole performance of the approaches will decrease. On the contrary, Point-to-Point based approaches do not have these disadvantages; instead, they use all the raw data.

Examples of Point-to-Point matching approaches are the following: iterative closest point (ICP), iterative matching range point (IMRP) and the popular iterative dual correspondence (IDC). ICP algorithm is one of the most successful and popular algorithms. The basic idea of ICP is that using a closest-point rule to initial guess of their relative pose, and then solving the Point-to-Point least-squares problem to compute their relative pose. Finally the relative pose is updated and the whole process iterates until the result is satisfying. Since ICP introduced by [(Chen 1991)], many variants have been proposed on the basic ICP concept. In the [(Besl 92)] proposed ICP, where for each point of the current scan, the point with the smallest Euclidean distance in the reference scan been selected. IMPR was proposed by [(Lu 94)], where corresponding points are selected by choosing a point which has the matching range from the center of the reference scan's coordinate system. IDC also proposed by [(Lu 94)], combines ICP and IMRP by using the ICP to calculate translation and IMPR to calculate rotation. The mentioned point to point methods can find the correct pose of the current scan in one step provided the correct associations are chosen. Since the correct associations are unknown, several iterations are performed. Matching may not always converge to the correct pose, since they can get stuck in a local minima. In [(Diosi 05)], Diosi presented a novel method for 2D laser scan matching called Polar Scan Matching (PSM). This method avoids searching for point associations by simply matching points with the same bearing. This association rule enables the construction of an algorithm faster than the iterative closest point (ICP).

Another Lider SLAM estimation method has been done for example by minimizing an energy function [(Lu 95)], using a combination of maximum likelihood with posterior estimation [(Thrun 00)] [(Vu 08)], using local registration and global correlation [(Gutmann 00)] and using FastSLAM [(Hahnel 03)]. A Kalman filter implementation can be found in [(Bosse 04)].

Based on correspondence establishment, we categorized various SLAM estimation methods showed in Table 1.

4 Implementation of Lidar SLAMs

To give a better understanding of different SLAM approaches' performance, in this section, we present three kinds of experiment method for SLAM estimation: (1) Probability edge matching method with traditional OG Maps. (2) Maximum measurement probability method with Bayesian updated OG Maps. (3) RANSAC method with Bayesian updated grid map. The experiments use a Regal laser scanner at a 0.1 degree bearing resolution in outdoor environments.

Table 1. Comparison of SLAM Estimation Method. F-T-F=Feature to Feature, P-T-F=Point to Feature, P-T-P=Point to Point, EKF=Extended Kalman Filter.

Sensor	System	Approach	Comments
Monocular	[(Bosse 02)]	F-T-F	vanishing points and 3D lines to recover mapping on a 946 meter path
	[(Dailey 05)]	EKF	resolved the problem of real-time operation
	[(Rosten 05)]	F-T-F	bootstrap line-tracking with detected point features
	[(Pupilli 05)]	F-T-F	EKF based, first practical real-time monocular SLAM
	[(Eade 06)]	Kalman filtering	combined particle filtering for localisation with Kalman filtering for mapping
Stereovision	[(Moravec 85)]	F-T-F	3D mapping from stereo range data with planar modeling assumption
	[(Ayache 88)]	EKF	used stereovision to directly extract the 3D location of line segments and filtered these within an EKF SLAM
	[(Garcia 04)]	F-T-F	3D SLAM based on feature point matching
	[(Lhuillier 05)]	P-T-P	used 5-point RANSAC and bundle adjustment to recover 3D map
	[(Dailey 05)]	F-T-F	estimating a map from observation of 3D line segments using a trinocular stereovision
	[(Jeong 06)]	F-T-F	build 3D maps robustly by aligning edge points between frames using the ICP algorithm
Lidar	[(Herath 07)]	F-T-F	build 3D maps by line features
	[(Cox 91)]	P-T-F	scan points matched to features like lines
	[(Lu 94)]	P-T-P	based on matching points with tangent directions in two scans to compute the relative pose of two scans.
	[(Thrun 00)]	Maximum likelihood	used combination of maximum likelihood with posterior estimation
	[(Gutmann 00)]	F-T-F	using local registration and global correlation used line segments for feature matching
	[(Guivant 00)]	F-T-F	used intensity of laser signal and geometrical primitives to define and detect features
	[(Zhao 01)]	F-T-F	employed feature-to-feature approaches to construct an urban map successfully
	[(Biber 03)]	P-T-F	features are Gaussian distributions with their mean and variance calculated
	[(Lingemann 04)]	F-T-F	used corners for feature
	[(Bosse 04)]	Kalman filter	used Kalman filter to estimation SLAM
	[(Tomono 04)]	F-T-F	utilizes Euclidean invariant features to match an input scan with reference scans without an initial alignment
	[(Vu 08)]	Maximum likelihood	estimate of the vehicle pose with the measurement likelihood that is nearly unimproved

4.1 Probability Edge Matching with Traditional Grid Map

The method represents the environment by means of a two dimensional evenly-spaced grid. Each grid cell estimates the probability of the corresponding region being occupied or free space area of the environment. The traditional grid map algorithm uses the Bayes' rule to model the sensor measurement and combine the probabilities from multiple measurements. Experimental researches have demonstrated that an alternative simple counting method may provide similar results with less computational burden, allowing safe real time operation.

In our test, after building the probability OG maps above, a simply clustering filter is adopted to remove outliers and find meaningful line segments. We detect edge features using the clusters of occupied data in occupancy gird map, small clusters were removed. When edge features have been established in each occupancy grid map, a pattern matching technology is used to estimate translation T and rotation R between two consecutive frames, this method is quite similar to [(Gutmann 00)] method. The total computational time needed for such a single scan matching is about 140-170 ms on a low-end PC.

4.2 Estimation Maximum Likelihood with Bayesian Updated Grid Map

In the occupancy grid representation, we apply Bayesian Update scheme (refer for detail section 2) that provides an elegant recursive formula to update the grid map.

In our test, we used an approach of matching problem as a maximum likelihood problem to estimate SLAM (similar to [(Vu 08)] approach). Given an underlying vehicle dynamics constraint, the current scan's position is corrected by comparing with the local grid map constructed from all observations in the past instead of only with one previous scan. By this way, we can reduce the ambiguity and weak constraint especially in outdoor environment and when the vehicle moves at high speeds. The resulting pose will be the pose at which the measurement probability achieves a maximum value. Because of the inherent discretization of the grid, the approach turns out to work very well. In practice, with a grid map resolution of 8 cm, it is to generate enough pose samples to obtain a good estimate of the vehicle pose with the measurement likelihood that is nearly unimproved even with more samples. The total computational time needed for such a single scan matching is about 70-110 ms on a low-end PC.

4.3 RANSAC Method with Bayesian Updated Grid Map

In this test, building OG map and clustering process is same with the first SLAM method. After the clustering process, the start point and the end point of each cluster are extracted. The correspondences between features identified in two successive frames are found by comparing the region (in our practice, the region that is the angle is not exceed 5 degree and distance is not exceed 2 meters). For a feature in the landmark set of the first frame, we assume that the features in the region of the second frame are its corresponding point. The resulting set of point correspondences is used as input for the RANSAC algorithm. The total computational time needed for such a single scan matching is about 30-80 ms on a low-end PC.

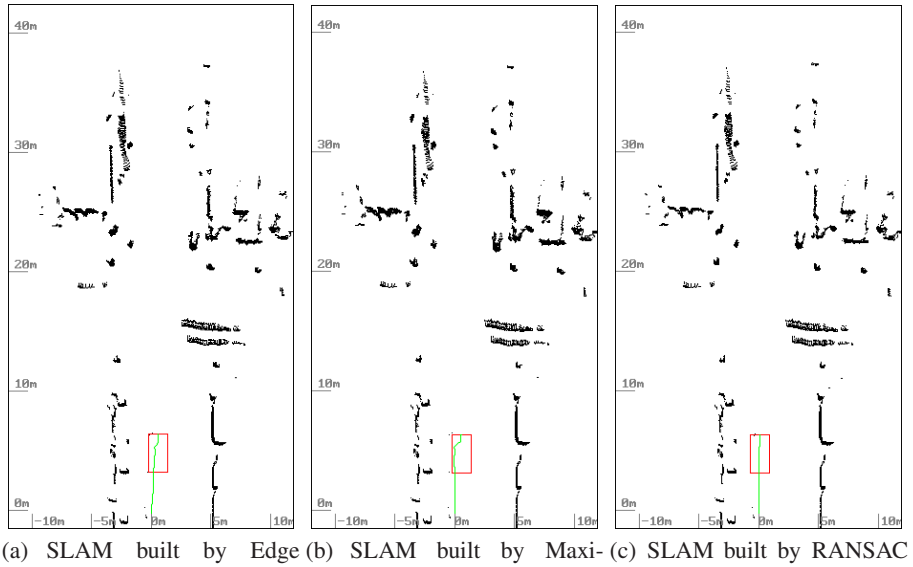
An overview of the algorithm is shown in follow: Repeat the following steps until stop condition satisfied:

– Choose a random subset of two set of corresponding points and calculate and by least-squares algorithm.
– Compute the consensus set by applying the calculated rotation and translation to all points of and then computing the distance to the corresponding points of set, add points whose distance is below a certain threshold.
– Save the consensus set if its size is bigger than a certain number.

Stop condition: iteration number is over N, or, the consensus set size is larger than a pre-defined threshold. If at least one consensus set is found, use the one with the most elements and calculate from these points the rotation and translation . If no consensus set is found, the algorithm outputs nothing.

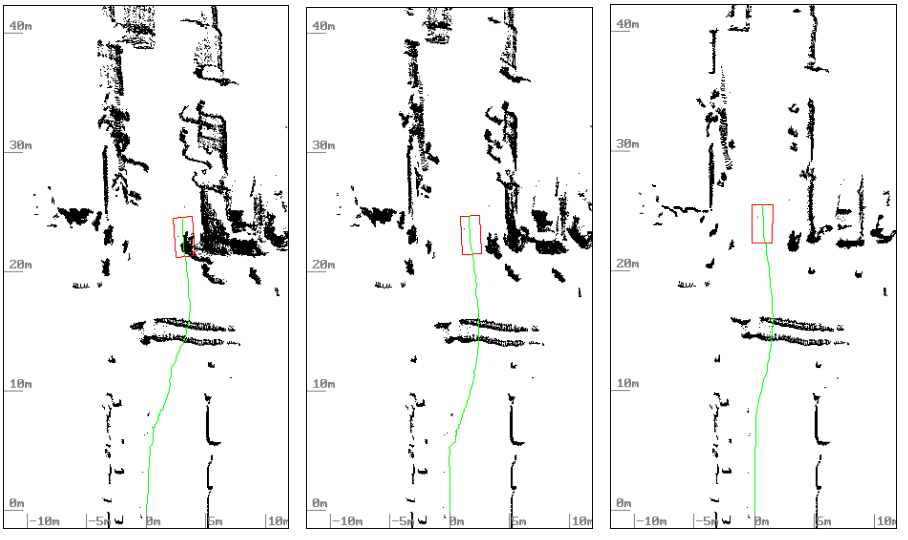
(a) Frame 139 (b) Frame 324

Fig. 1. Comparison SLAM results with three kind of method



(a) SLAM built by Edge Matching method (b) SLAM built by Maximum measurement probability method (c) SLAM built by RANSAC method

Fig. 2. SLAM built in Frame 139



(a) SLAM built by Edge Matching method (b) SLAM built by Maximum measurement probability method (c) SLAM built by RANSAC method

Fig. 3. SLAM built in frame 324

4.4 Comparative Analysis of SLAM Results

Figure 1 shows two camera images in frame 139 and 324. In frame 139, our vehicle was moving in a straightaway, and two pedestrians were ready for across the crossing from right to left. In frame 324, our vehicle was moving and turning right slightly to avoid two pedestrians. In this scene, there are also other walking pedestrians. Figure 2 shows the estimated SLAM results by three test methods. Since most of objects were static, our SLAM results show not significant difference. However the vehicle path, we can see, RANSAC method shows the best performance, because another method were very sensitive to the dynamic objects. In figure 3, since many dynamic objects (walking pedestrians) exist, the RANSAC results were much better than another methods.

5 Conclusions

This paper reviews a number of promising approaches and provides an overview of recent developments in the domain of SLAM, which aims at building a consistent map of the environment and at the same time determine the location of moving robot within this map. In this paper, we classify SLAM by the processes (occupancy grid mapping and localization estimation), sensors (visual SLAM, Lidar SLAM and SLAM by sensor fusion) and uncertainty calculation (probabilistic, evidence theoretic and possibilistic SLAMs). Possibilistic SLAM with RANSAC estimation shows better performance in a noisy environment to build the occupancy grid maps in our experiment.

References

[(Ayache 88)] Ayache, N., Faugeras, O.D.: Building, registrating, and fusing noisy visual maps. International Journal of Robotics Research 7(6), 45–65 (1988)

[(Besl 92)] Besl, P., Mckay, N.D.: A method for registration of 3D shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence 14(2), 239–256 (1992)

[(Biber 03)] Biber, P., Straser, W.: The normal distributions transform: A new approach to laser scan matching. In: IROS 2003, vol. 3, pp. 2743–2748 (2003)

[(Bosse 02)] Bosse, M., Rikoski, R., Leonard, J., Teller, S.: Vanishing points and 3d lines from omnidirectional video. In: ICIP 2002, vol. 3, pp. 513–516 (2002)

[(Bosse 04)] Bosse, M., Newman, P., Leonard, J., Teller, S.: Simultaneous localization and map building in large-scale cyclic environments using the Atlas framework. The International Journal of Robotics Research 23(12), 1113–1139 (2004)

[(Chen 1991)] Chen, Y., Medioni, G.: Object Modeling by Registration of Multiple RangeImage. In: Proc. IEEE Int. Conf. on Robotics and Automation (1991)

[(Cox 91)] Cox, I.J.: Blanche.an experiment in guidance and navigation of an autonomous robot vehicle. IEEE Transactions on Robotics and Automation 7(2), 193–203 (1991)

[(Dailey 05)] Dailey, M.N., Parnichkun, M.: Landmark-based simultaneous localization and mapping with stereo vision. In: Proc. Asian Conference on Industrial Automation and Robotics, Bangkok, Thailand (2005)

[(Davison 03)] Davison, A.J.: Real-time simultaneous localisation and mapping with a single camera. In: Proc. IEEE International Conference on Computer Vision, pp. 1403–1410 (2003)

[(Deans 05)] Deans, M.: Bearing-Only Localization and Mapping. PhD thesis, tech. report CMU-RI-TR-05-41, Robotics Institute, Carnegie Mellon University (2005)

[(Diosi 05)] Diosi, A., Kleeman, L.: Laser scan matching in polar coordinates with application to SLAM. Intelligent Robots and Systems, 3317–3322 (2005)

[(Duckett 00)] Duckett, T., Saffiotti, A.: Building globally consistent gridmaps from topologies. In: Proc. of the 6th Int. IFAC Symp. on Robot Control, SyROCO, Austria (2000)

[(Eade 06)] Eade, E., Drummond, T.: Scalable monocular SLAM. In: Proc. Conference on Computer Vision and Pattern Recognition, New York, USA, pp. 469–468 (2006)

[(Elfes 92)] Elfes, A.: Multi-source spatial data fusion using Bayesian reasoning. In: Abidi, M.A., Gonzalez, R.A. (eds.) Data fusion in robotics and machine intelligence, ch. 3. Academic Press, New York (1992)

[(Elinas 06)] Elinas, P., Sim, R.: Little. J.J.: SLAM: Stereo Vision SLAM Using the Rao-Blackwellised Particle Filter and a Novel Mixture Proposal Distribution. In: Proc. of ICRA 2006, pp. 1564–1570 (2006)

[(Fox 99)] Fox, D., Burgard, W., Thrun, S.: Probabilistic methods for mobile robot mapping. In: Proc. of the IJCAI 1999 Workshop on Adaptive Spatial Representations of Dynamic Environments (1999)

[(Gambino 96)] Gambino, F., Oriolo, G., Ulivi, G.: Comparison of three uncertainty calculus techniques for ultrasonic map building. In: Proc. SPIE Int. Symp. on Aerospace/Defense Sensing and Control, vol. 2761, pp. 249–260 (1996)

[(Garcia 04)] Garcia, M.A., Solanas, A.: 3D Simultaneous Localization and Modeling from Stereo Vision. In: Proc. of ICRA 2004, pp. 847–853 (2004)

[(Groecke 07)] Groecke, R., Asthana, A., Pettersson, N., Petersson, L.: Visual vehicle egomotion estimation using the fourier-mellin transform. In: IEEE Intelligent Vehicles Symposium, Istanbul, pp. 450–455 (2007)

[(Guivant 00)] Guivant, J., Nebot, E., Baiker, S.: High accuracy navigation using laser range sensors in outdoor applications. In: IEEE Int. conf. on Robotics and automation, vol. 4, pp. 3817–3822 (2000)

[(Gutmann 00)] Gutmann, J.S.: Robuste Navigation autonomer mobiler Systeme. PhD thesis, Universityy of Freiburg (2000)

[(Hahnel 03)] Hahnel, D., Burgard, W., Fox, D., Thrun, S.: An efficient fastSLAM algorithm for generating maps of large-scale cyclic environments from raw laser range measurements. In: IROS 2003, vol. 1, pp. 206–211 (2003)

[(Hartley 04)] Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press, Cambridge (2004)

[(Herath 07)] Herath, D.C., Kodagoda, K.R.S., Dissanayake, G.: Stereo Vision Based SLAM: Issues and Solutions, Vision Systems. In: Obinata, G., Dutta, A. (eds.) Advanced Robotic Systems, pp. 565–582 (2007)

[(Iocchi 00)] Iocchi, L., Konolige, K., Bajracharya, M.: Visually Realistic Mapping of a Planar Environment with Stereo. In: Proc. of ISER 2000 (2000)

[(Jeong 06)] Jeong, W.Y., Lee, K.M.: Visual SLAM with Line and Corner Features. In: Proc. of Intelligent Robots and Systems, pp. 2570–2575 (2006)

[(Lhuillier 05)] Lhuillier, M.: Automatic structure and motion using a catadioptric camera. In: IEEE Workshop on Omnidirectional Vision (2005)

[(Lingemann 04)] Lingemann, K., Surmann, H., Nuchter, A., Hertzberg, J.: Indor and outdoor localization for fast mobile robots. In: IROS 2004, vol. 3, pp. 2185–2190 (2004)

[(Lu 94)] Lu, F., Milios, E.E.: Robot pose estimation in unknown environments by matching 2D range scans. Computer Vision and Pattern Recognition, 935–938 (1994)

[(Lu 95)] Lu, F.: Shape Registration Using Optimization for Mobile Robot Navigation. PhD thesis, University of Toronto (1995)

[(Milford 08)] Milford, M.J., Wyeth, G.F.: Single camera vision-only slam on a suburban road network. In: IEEE International Conference on Robotics and Automation, ICRA 2008, pp. 3684–3689 (2008)

[(Milford 04)] Milford, M.J., Prasser, D., Wyeth, G.: Ratslam: A hippocampal model for simultaneous localization and mapping. In: IEEE International Conference on Robotics and Automation, USA, pp. 403–408 (2004)

[(Moravec 85)] Moravec, H.P., Elfes, A.: High Resolution Maps from Wide Angle Sonar. In: Proceedings of the IEEE International Conference on Robotics and Automation, March 1985, pp. 116–121 (1985)

[(Moravec 01)] Moravec, H.: DARPA MARS program research progress. Carnegie Mellon University (2001), http://www.frc.ri.cmu.edu/hpm/talks/Report.0107.html

[(Murray 00)] Murray, D., Little, J.J.: Using real-time stereo vision for mobile robot navigation. Autonomous Robots 8(2), 161–171 (2000)

[(Oriolo 99)] Oriolo, G., Ulivi, G., Vendittelli, M.: Real-time map building and navigation for autonomous robots in unknow environments. IEEE Transactions on Systems, Man, and Cybernetics 5 (1999)

[(Pupilli 05)] Pupilli, M., Calway, A.: Real-time camera tracking using a particle filter. In: Proc. British Machine Vision Conference, Oxford, UK, pp. 519–528 (2005)

[(Ribo 01)] Ribo, M., Pinz, A.: A comparison of three uncertainty calculi for building sonar-based occupancy grids. International Journal of Robotics and Autonomous Systems 35, 201–209 (2001)

[(Rosten 05)] Rosten, E., Drummond, T.: Fusing points and lines for high performance tracking. In: Proc. IEEE International Conference on Computer Vision, Beijing, China, pp. 1508–1515 (2005)

[(Scaramuzza 09)] Scaramuzza, D., Fraundorfer, F., Siegwart, R.: Real-Time Monocular Visual Odometry for On-Road Vehicles with 1-Point RANSAC. In: Proceedings of the IEEE International Conference on Robotics and Automation, Japan, pp. 4293–4299 (2009)

[(Se 01)] Se, S., Lowe, D., Little, J.: Local and Global Localization for Mobile Robots using Visual Landmarks. In: Proc. of IROS 2001 (2001)

[(Tardif 08)] Tardif, J., Pavlidis, Y., Daniilidis, K.: Monocular visual odometry in urban environments using an omnidirectional camera. In: IEEE IROS, pp. 2531–2538 (2008)

[(Thrun 00)] Thrun, S., Burgard, W., Fox, D.: A real-time algorithm for mobile robot mapping with applications to multi-robot and 3D mapping. In: ICRA 2000, vol. 1, pp. 321–328 (2000)

[(Thrun 03)] Thrun, S.: Learning occupancy grids with forward sensor models. Autonomous robots 15, 111–127 (2003)

[(Tomono 09)] Tomono, M.: Robust 3D SLAM with a stereo Camera Based on an Edge-Point ICP Algorithm. In: IEEE International Conference on robotics and Automation, Japan, May 2009, pp. 4306–4311 (2009)

[(Tomono 04)] Tomono, M.: A scan matching method using euclidean invariant signature for global localization and map building. In: ICRA 2004, pp. 866–871 (2004)

[(Vu 08)] Vu, T.D., Burlet, J., Aycard, O.: Mapping of environment, Detection and Tracking of Moving Objects using Occupancy Grids. In: Intelligent Vehicles Symposium, pp. 684–689. IEEE, Los Alamitos (2008)

[(Zhao 01)] Zhao, H., Shibasaki, R.: Reconstructing Urban 3D Model using Vehicle-borne Laser Range Scanners. In: Proc. of the Third Int. Conf. on 3-D Digital Imaging and Modeling, May 2001, pp. 349–356 (2001)

On the Analysis of Force Transmission Performance for the Thrust Systems of Shield Tunneling Machines

Kongshu Deng, Xiaoqiang Tang*, Liping Wang, Pingfa Feng, and Xu Chen

Department of Precision Instrument and Mechanology,
Tsinghua University, Beijing, People's Republic of China
tang-xq@mail.tsinghua.edu.cn

Abstract. Earth pressure balance is normally used in excavating tunneling due to high automation, quick construction, environmental friendliness, etc. Based on minimum mean square error, optimized mechanical models for non-group and group thrust systems of the earth pressure balance have been constructed. Through the offset load, the force transmission performances for the two thrust systems were studied, respectively. By applying force ellipse to the thrust system of a shield machine, the difference between the two configurations was analyzed. It provides an important theoretical foundation for the development and design of thrust systems.

Keywords: Earth pressure balance, Thrust systems, Force transmission, Force ellipse.

1 Introduction

In excavating tunneling for subways and power cables, shield tunneling machines are often used. At present, the shield tunneling is the most promising and competitive method for underground space exploration and utilization, characterized by quick and safe construction, high automation, environmental friendliness, etc [1].

The thrust system performs the task of driving shield ahead tunneling and controlling the pose of the shield, ensuring the shield can excavate along the given route [2]. However, the anisotropy of excavation interface rocks, the inconsistencies of resistance surrounding machine and its own weight lead to the offset loading. On the one hand, the big offset may make segments crack [3]; On the other hand, the shield machine extrudes the surrounding soil and the surface could be uplift. Then the collapse of houses and buildings on the ground would occur. Therefore, it is of primary importance to analyze and control the forces of the thrust system [4].

For the drive system of the shield, most of current studies focus on a given geological and construction requirements of the thrust system to estimate the tunneling parameters for providing the system design basis [5]. However, these studies are usually

* Corresponding author.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 268–278, 2009.

less involved in the structure of the shield and its law of force transmission. Based on the shield mechanical optimization model, the law of force transmission for hydraulic cylinders of the thrust system has been analyzed in more detail and the result could provide a theoretical foundation and support for the design of shield tunneling machines and automatic control for thrust systems.

2 Mechanical Model of Thrust Systems

In the tunneling process, the shield machine will be resisted by the drag force F_z which is produced by the soil mass. The resistive torques M_x and M_y will be produced due to its gravity, unevenness of excavation interface socks and the extrusion force of the curve tunneling. By overcoming the external resistive force F_z and the moment M_x and M_y, the driving system can control the main machine to pitch, sway and go straightly along the expected line.

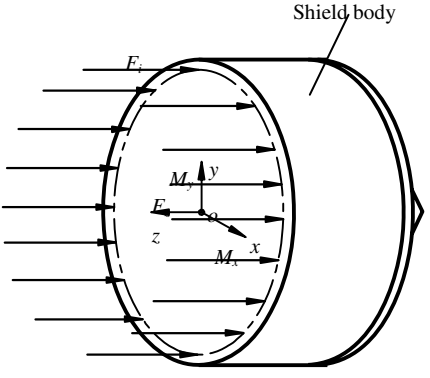


Fig. 1. Mechanical model of thrust system

As shown in Fig. 1, the circular center of shield hydraulic cylinders is considered as the coordinate origin, z-axis as being in parallel with the center line of the hydraulic cylinder and opposite to the driving direction, and y-axis as being perpendicular to z-axis and pointing to the ground. The x-axis is determined by the right-hand rule, and then a Cartesian coordinate system is constructed. In Fig.1, F_z is the resultant resistance in the z direction; M_x and M_y are the drag torques in the x direction and y direction, respectively; F_i (i = 1, 2,..., N) is the thrust of the i hydraulic cylinder; the system has N hydraulic cylinders [6].

According to the mechanical model in Fig. 1, the following force balance equations can be given:

$$\sum_{i=1}^{N} F_i + F_z = 0 \tag{1}$$

$$\sum_{i=1}^{N} F_i x_i - M_y = 0 \tag{2}$$

$$\sum_{i=1}^{N} F_i y_i + M_x = 0 \tag{3}$$

Where, F_i denotes the thrust of the i hydraulic cylinder, (x_i,y_i) is the coordinate of the point of the force applied.

3 Analysis of the Force Distribution Law for the Driving System

3.1 Force Distribution Law of the Non-group System

By controlling the pressure of the hydraulic cylinders in automatic systems, the configuration is usually non-group [6]. To achieve the expected position of the shield machine, the point of the resultant force would be controlled by adjusting the force of each cylinder, and then the shield machine can excavate along the planed route.

In order to force uniformly on segments, the optimization function is as follows

$$\Delta = \frac{1}{2}\sum_{i=1}^{N}\left(F_i - \bar{F}\right)^2 \tag{4}$$

In Eq. (4), the $\bar{F}$ can be expressed as

$$\bar{F} = \frac{1}{N}\sum_{i=1}^{N} F_i = -\frac{1}{N}F_z \tag{5}$$

Where N is the number of hydraulic cylinders of the thrust system.

According to Eqs. (1)-(4), the Lagrange function can be derived

$$L = \Delta + \lambda_1\left(\sum_{i=1}^{N} F_i + F_z\right) + \lambda_2\left(\sum_{i=1}^{N} x_i F_i - M_y\right) + \lambda_3\left(\sum_{i=1}^{N} y_i F_i + M_x\right) \tag{6}$$

The partial differential form of Eq. (6) can be obtained

$$\frac{\partial L}{\partial F_i} = F_i + \frac{F_z}{N} + \lambda_1 + \lambda_2 x_i + \lambda_3 y_i = 0, \left(i = 1, 2, ..., N\right) \tag{7}$$

$$\frac{\partial L}{\partial \lambda_i} = 0, \left(i = 1, 2, ..., 3\right) \tag{8}$$

Letting z_i=-F_i, Eq. (7) can be transferred into

$$-\lambda_2 x_i - \lambda_3 y_i + z_i = \frac{F_z}{N} + \lambda_1 \tag{9}$$

Due to the hydraulic cylinders with uniform spacing along the circumference, the following equation can be obtained

$$x_i^2 + y_i^2 = r^2 \tag{10}$$

where r is the radius of the layout of the hydraulic cylinders.

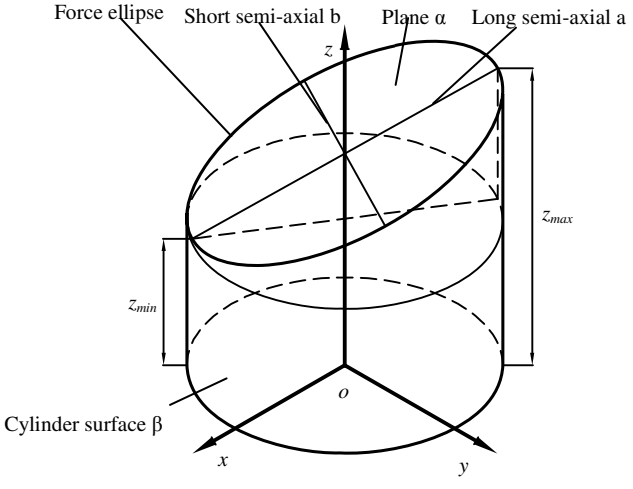


Fig. 2. Non-group construction force ellipse model

As shown in Fig.2, the plane α is constituted with Eq. (9) and the cylinder surface β is made with Eq. (10). The space point (x_i, y_i, z_i) is on the closed spatial ellipse curve line of intersection between the plane α and the cylinder surface β. z_i is the value of the force of the ith hydraulic cylinder at the stress point coordinates (x_i, y_i)

According to Eqs. (8)-(10), z_i can be solved as

$$z_i = -\frac{2M_y}{Nr^2} x_i + \frac{2M_x}{Nr^2} y_i + \frac{F_z}{N}, \left(i = 1, 2, .., N\right) \tag{11}$$

From Eqs. (10)-(11), Eqs.(12) and (13) are given by

$$z_{\max} = \frac{2\sqrt{M_x^2 + M_y^2}}{Nr} + \frac{F_z}{N} \tag{12}$$

$$z_{\min} = -\frac{2\sqrt{M_x^2 + M_y^2}}{Nr} + \frac{F_z}{N} \tag{13}$$

According to Eqs. (12)-(13), the parameters of the force ellipse can be obtained

$$a = \sqrt{\frac{4\left(M_x^2 + M_y^2\right)}{N^2 r^2} + r^2} \tag{14}$$

$$b = r \tag{15}$$

$$e = \sqrt{1 - \frac{1}{\frac{4\left(M_x^2 + M_y^2\right)}{N^2 r^4} + 1}} \tag{16}$$

The eccentricity of the force ellipse suggests the degree of the offset load, and the various parameters on the impact of e will be discussed as follows:

(1) Equation (16) shows that with increasing resistance torque M_x or M_y, e values gradually increase. It means that the level of the offset loading enhances and the degree of even force on segments remits. The segments are vulnerable to crack when the offset load is increased. Consequently, the bigger one can be obtained in engineering applications through several continuous little deviations correct in order to decrease the drag torque and the offset load.

(2) With the increasing of the radius of the layout of cylinders, the eccentricity becomes smaller. It is shown that the level of even load on segments is better. As a result, there is a data of eccentricity between the center of cylinders and pressure pads. In other words, it will increase the force radius in this way.

(3) The thrust F_z has no bearing upon the ellipse eccentricity.

3.2 Force Distribution Law of the Four-Group System

In order to reduce the complexity of control, all hydraulic cylinders on the existing shield thrust system are usually divided into four groups as A, B, C and D in the vertical and horizontal directions. In the excavation process, the shield machine can sway, pitch or push forward by controlling the pressure and velocity of each group individually. Then, the line of shield machine tunneling can be ensured to conform to the design of the tunnel alignment [7].

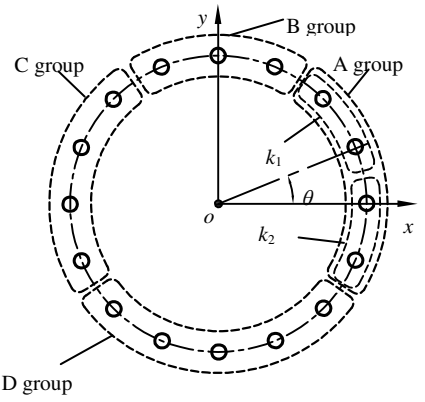


Fig. 3. Four-group of thrust cylinders in shield machine force

As shown in Fig. 3, all the hydraulic cylinders in Fig.1 are divided into four groups. A has n_1 cylinders; B has n_2; C has n_3; D has n_4. Among A group, k_1 hydraulic cylinders are in the first quadrant and k_2 in the fourth quadrant. The angle between the center line of the first cylinder and the x-axis is θ and the number of left group is equal to the right.

Similarly, the force balance Eqs. (17)-(19) can be obtained

$$\sum_{i=1}^{4} n_i F_i + F_z = 0 \tag{17}$$

$$\sum_{i=1}^{4} a_i F_i + \frac{M_x}{r} = 0 \tag{18}$$

$$\sum_{i=1}^{4} b_i F_i - \frac{M_y}{r} = 0 \tag{19}$$

Where

$$a_1 = \sum_{i=1}^{k_1} \sin\left(\theta + \frac{2(i-1)\pi}{N}\right) + \sum_{i=N-k_2+1}^{N} \sin\left(\theta + \frac{2(i-1)\pi}{N}\right)$$

$$a_2 = \sum_{i=k_1+1}^{k_1+n_2} \sin\left(\theta + \frac{2(i-1)\pi}{N}\right)$$

$$a_3 = \sum_{i=k_1+n_2+1}^{k_1+n_2+n_3} \sin\left(\theta + \frac{2(i-1)\pi}{N}\right),$$

$$a_4 = \sum_{i=k_1+n_2+n_3+1}^{N-k_2} \sin\left(\theta + \frac{2(i-1)\pi}{N}\right),$$

$$b_1 = \sum_{i=1}^{k_1} \cos\left(\theta + \frac{2(i-1)\pi}{N}\right) + \sum_{i=n-k_2+1}^{N} \cos\left(\theta + \frac{2(i-1)\pi}{N}\right),$$

$$b_2 = \sum_{i=k_1+1}^{k_1+n_2} \cos\left(\theta + \frac{2(i-1)\pi}{N}\right),$$

$$b_3 = \sum_{i=k_1+n_2+1}^{k_1+n_2+n_3} \cos\left(\theta + \frac{2(i-1)\pi}{N}\right),$$

$$b_4 = \sum_{i=k_1+n_2+n_3+1}^{N-k_2} \cos\left(\theta + \frac{2(i-1)\pi}{N}\right),$$

For simplicity, the optimization function can be constructed as following:

$$\Delta = \frac{1}{2}\sum_{i=1}^{n}\left(F_i - \bar{F}\right)^2 \tag{20}$$

Where $\bar{F} = \frac{1}{4}\sum_{i=1}^{4} F_i$

According to Eqs. (17)-(20), the Lagrange function can be expressed as

$$L = \frac{1}{2}\sum_{i=1}^{n}\left(F_i - \bar{F}\right)^2 + \lambda_1\left(\sum_{i=1}^{4} n_i F_i + F_z\right) + \lambda_2\left(\sum_{i=1}^{4} a_i F_i + \frac{M_x}{r}\right) + \lambda_3\left(\sum_{i=1}^{4} b_i F_i - \frac{M_y}{r}\right) \tag{21}$$

Of $F_i(i=1,2,3,4)$ and $\lambda_i(i=1,2,3)$ for the partial derivative, the following equations can be obtained

$$\frac{\partial L}{\partial F_i} = F_i - \overline{F} + n_i\lambda_1 + a_i\lambda_2 + b_i\lambda_3 = 0, (i = 1,2,3,4) \tag{22}$$

$$\frac{\partial L}{\partial \lambda_i} = 0, (i = 1,2,3) \tag{23}$$

Letting z_i=-F_i(i=1,2,3,4) and solving Eqs. (22)-(23), the solutions can be obtained,

$$z_1 = \frac{F_z}{N} - \frac{M_x\left(3n_2a_2 + n_2a_4 + n_4a_2 + 3n_4a_4\right)}{\left(3a_2^2 + 2a_2a_4 + 3a_4^2\right)Nr} - \frac{M_y}{2b_1r} \tag{24}$$

$$z_2 = \frac{F_z}{N} + \frac{2M_x\left(3n_1a_2 + n_1a_4 + n_4a_2 - n_4a_4\right)}{\left(3a_2^2 + 2a_2a_4 + 3a_4^2\right)Nr} \tag{25}$$

$$z_3 = \frac{F_z}{N} - \frac{M_x\left(3n_2a_2 + n_2a_4 + n_4a_2 + 3n_4a_4\right)}{\left(3a_2^2 + 2a_2a_4 + 3a_4^2\right)Nr} + \frac{M_y}{2b_1r} \tag{26}$$

$$z_4 = \frac{F_z}{N} + \frac{2M_x\left(n_1a_2 + 3n_1a_4 + n_2a_4 - n_2a_2\right)}{\left(3a_2^2 + 2a_2a_4 + 3a_4^2\right)Nr} \tag{27}$$

Defining z_{fmax}=max(z_1,z_2,z_3,z_4) and z_{fmin}=min(z_1,z_2,z_3,z_4) as the two points of long axis of the force ellipse for four-group sytem and two other as points of the ellipse, the spatial force ellipse can be given in Fig. 4.

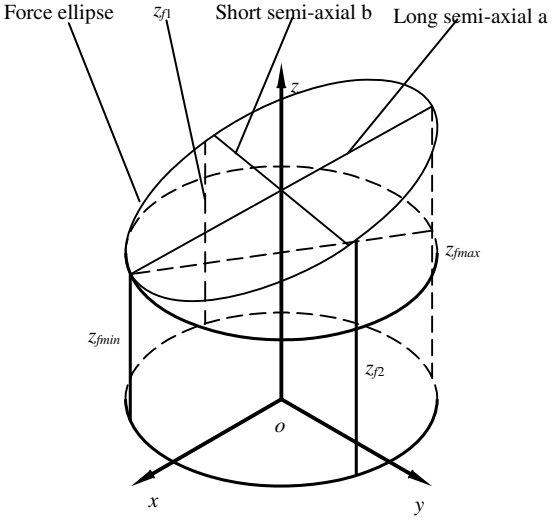


Fig. 4. Four-group mechanical ellipse model

According to the theoretical ellipse formula, the eccentricity of force ellipse can be described as

$$e_f = \sqrt{1 - \frac{1}{\frac{\left(z_{f\max} - z_{f\min}\right)^2}{4r^2} + 1}} \tag{28}$$

From Eqs. (24)-(28), it can be found

(1) Eq. (28) suggests that for non-group system, the eccentricity increases with the increasing resistive moment and the decreasing radius of arrangement. It also shows the thrust F_z does not affect the eccentricity e_f
(2) To make the machine sway motion, adjusting the pressure of Groups A and C group cylinders, Groups B and D can be determined by the average value of Groups A and C .
(3) To make the shield pitch motion, adjusting the pressure of Groups B and D cylinders, the pressure in Groups B and D can be given

$$z_1 = z_3 = \frac{F_z}{n} - \frac{M_x\left(3n_2a_2 + n_2a_4 + n_4a_2 + 3n_4a_4\right)}{\left(3a_2^2 + 2a_2a_4 + 3a_4^2\right)nr} \tag{29}$$

If the number of Group B is equal to Group D, Eq. (30) can be reduced to

$$z_1 = z_3 = \frac{\left(z_2 + z_4\right)}{2} \tag{30}$$

Thus, for a four-group system, if the upper is equal to the lower and the left is equal to the right, their pressure satisfy the linear relation.

(4) Eqs. (24)-(27) indicate that Groups B and D only generate pitching moment while Groups A and D produce pitching torque besides swaying moment.

4 Case Study

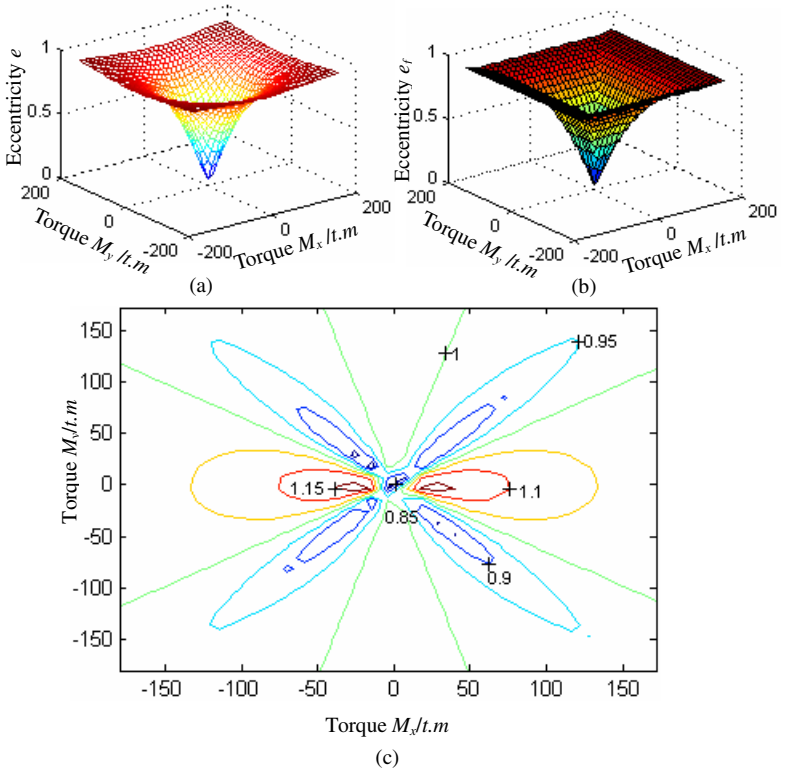
The basic structure parameters for a shield machine =6.34*m* are: *r*=2.85*m*; *N*=22 [8-9]. All cylinders are divided into four groups: A and C are six and B and D are five. In Fig. 5, it is shown that the eccentricity changes with the drag moment on non-group and four-group.

The value of e/e_f is defined as the eccentricity ratio e_r and the value symbols the eccentricity comparison between the non-group and the four-group under the same condition.

The following statements can be obtained from the Fig.5

(1) The elliptical eccentricity implies that the uniformity of stress on segments. It is shown that in Fig.5 (a) the eccentricity rises in circles with the increase of the resistive moment while in Fig.5 (b) the eccentricity rises in squares.

(2) As shown in Fig.5 (c), the value of eccentricity ration ranges from 0.85 to 1.15. Thus, although the pressure changes dramatically and the shearing force would produce easily on the non-group, the offset load has insignificant impact on both structures under the same condition. So, the four-group structure has been usually adopted in engineering applications.



(a) Eccentricity for non-group structure ;

(b) Eccentricity for four-group structure ;

(c) Contour for eccentricity ratio ;

Fig. 5. Force ellipse eccentricity comparison

5 Conclusion

This paper presents a force transmission optimization model for thrust systems in non-group and four-group structures according to the evenness of the pushing forces

on segments. By the force ellipse, the force transmission characteristics have been analyzed. Finally, the features of two structures have been compared through a case study. From the analyses above, the following result can be obtained:

(1) The force ellipse can be applied to the analysis of the thrust layout; Resistive moment is the key factor for two structures; Reducing torque and increasing radius can decrease the offset load.
(2) Due to the stepped pressure for the four-group, the segments are open to crack, and the offset load is not very large.
(3) According to the force transmission optimization model, the pressure can be provided for manipulators, and the excavation deviation could be avoided. Meanwhile, the optimization method can offer a theoretical foundation for the design and development of automatic control systems.
(4) The model has the feature of generalization in usage and four groups can be extended to more groups for the force optimization model.

Acknowledgments

This research is sponsored by the National 973 Program of China (No. 2007CB714007), National Natural Science Foundation of China (No.50605035, 10778625), the Important National Science & Technology Specific Projects of China (No.2009ZX04001-042-02).

References

1. Shogo, T., Yukihiro, T.: Automatic Detection and Characterization of Anomalous Objects in Shield Tunneling Method. In: 20th International Conference on Industrial Electronics, Control and Instrumentation, pp. 810–815. IEEE Press, Iecon (1994)
2. Zhou, Q., Chen, J., He, Z., Zhang, H.: Design of Intellectualized Controller of Shield Machine (in Chinese). Journal of Tongji University Natural Science 36, 76–80 (2008)
3. Tan, Z., Hong, K., Wan, J., Wang, M.: Shield Driven Tunneling in the Complex Layer: Direction Control of Shield and Preventing Segment Crack (in Chinese). Engineering Science 8, 92–96 (2006)
4. Maidl, B., Herrenknecht, M., Anheuser, L.: Mechanised Shield Tunnelling. Ernst & Sohn, Berlin (1996)
5. Deng K.S., Tang X.Q., Wang L.P., Chen X.: Construction and Analysis of Nonlinear Dynamic Model for the Hydraulic Thrust System of Earth Pressure Balance (in Chinese). Chinese High Technology Letters (accepted)
6. Zhang, F., Zhu, H., Fu, D.: Shield Tunneling Method (in Chinese). China Communications Press, Beijing (2004)
7. Hu, S., Gong, G.F., Yang, H.Y.: Pressure and Speed Control of Electro-hydraulic Drive for Shield Tunneling Machine. In: International Conference on Advanced Intelligent Mechatronics, pp. 314–317. IEEE Press, Xi'an (2008)
8. Wu, Y.: Application of Articulation Device of Shield Machine to Excavation of Tunnel with Minor Radius Bend Angle (in Chinese). Building Construction 27, 34–35 (2005)
9. Min, R., Huang, J.: φ 6.34m Earth Balance Pressure Machine (in Chinese). Building Construction 12, 46–49 (2005)

Passive Ultrasonic RFID Localization for Pipeline Pigs

Xubing Chen[1], Hanxin Chen[1], and Youlun Xiong[2]

[1] School of Mechanical and Electrical Engineering, Wuhan Institute of Technology, Wuhan 430073, P.R. China
[2] School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, P.R. China
bluegif@gmail.com

Abstract. For locating pipeline pigs, passive ultrasonic RFID tags are proposed to indicate pipeline segments in the paper. Ultrasonic transmitters and receivers are carried on the pig robot, and RFID tags are placed on the outside wall of long-distance pipelines. The built ultrasonic RFID localization model shows that ultrasonic transmitters and receivers work in their radiating near-field regions, while RFID tags work in their reactive near-field regions. These tags can also be discovered by hand-held readers when the maintenance required. Therefore, both requirements of inner localization and outer localization are satisfied simultaneously. What is more, system structures and operation steps of pig launching and pig receiving are provided. In discussion, three RFID tags are placed around the outside wall of each pipeline segment to improve the identification rate and in case of any potential malfunction. The methodology also provides a feasible solution to determine orientations of pipeline pigs and potential defects.

Keywords: Passive ultrasonic RFID, pipeline pig, inner localization, outer localization, orientation determination.

1 Introduction

As pipeline transportation has its merits of low cost, save energy, high security, and continuous supply, it is widely used for oil transportation in the world. But for reasons of abrasions, aging effects, geographical transformations, climatic changes, and human-caused damages, leakage accidents happen frequently [1]. These kinds of hazards have influenced our daily life, property securities and environments greatly. At the same time, serious waste of natural resources produced. It is very important to discover and locate pipeline defects before any potential leakage happens. To reach this goal, Chinese government has issued a decree in April, 2000. It is said that all major oil transportation pipelines in the country should be online inspected or detected every 3 to 5 years. As reasons of most pipelines are made of metal and running underground for years, they cannot be detected by human resources directly. Smart pipeline pigs are summoned to carry out these kinds of arduous tasks in most cases.

Smart pipeline pigs are autonomous robots inserted into and travelled throughout the full length of long-distance pipelines and driven by the product flow. Although each pipeline has its own set of characteristics which affect how and why pigging is

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 279–287, 2009.

used, there are basically three reasons to pig a pipeline [2]: to batch or separate dissimilar products; for displacement purposes; for internal inspections. As the oil flowing velocity is a variable related to temperature gradients, densities and compression performances of the crude oil, difficulties of locating a pipeline pig have been increased after it is inserted into the dark, viscous and hot pipelines.

Generally, there are two issues should be discussed in the field of locating pipeline pigs: inner localization and outer localization. Inner localization is used for pipeline pigs to tell inner locations of potential defects, such as distortions, metal peelings, cracks, and so on. While outer localization is used for workers to find where starting a digging. On the one hand, most methodologies for inner localization are reported in literatures: counting wheels, vision sensors, curvature sensors, rays, static magnetic field and multi-sensors fusion [3-4], extreme-low-frequency electromagnetic pulse [5], and so on. For long-distance pipelines, cumulative errors play important roles. Usually, welding seam tracking is employed to find which segment that the pig is located firstly, and counting wheels are used to count the relative location in the pipeline to avoid cumulative errors [6]. Missing of any welding seam will cause a series of dislocations after the missing point. On the other hand, new methodologies of outer localization are studied by some pioneers increasingly. RFID, Radio-Frequency Identification, has entered the views of researchers for its wide applications of identification and tracking, even from a distance over meters. It is reported that mobile robot localization based on RFID has been developed [7-9]. Electronic markers [10] and RFID markers [11] are used for outer localization of pipelines too.

But there is no methodology reported to provide a solution of both inner localization and outer localization for metallic pipelines simultaneously. Metal shields are great barriers for electromagnetic waves to cross. But the ultrasonic wave, a mechanical wave, has outstanding abilities of penetrating through metals, soils and liquids up to a thickness of 10 meters. They have wide employments in some industry applications, such as non-destructive detection [12]. Following this idea, ultrasound-RFID indoor positioning [13] and passive ultrasonic RFID elevator positioning [14] are researched. What is more, ultrasonic wave has a frequency range of 20 kHz to 10MHz that covers the frequency range of low frequency (LF) RFID. As a matter of course, passive ultrasonic RFID tags are proposed for both inner localization and outer localization of pipeline pigs in the paper.

The rest of the paper is organized as follows. First of all, the ultrasonic RFID localization model is introduced in Section 2. Next, the structures and operation steps of pipeline pig launching system and receiving system are described in Section 3. In Section 4, some questions of ultrasonic RFID localization are discussed, and the three tags' redundancy methodology is proposed to improve the identification rate of passive ultrasonic RFID localization system during its lifetime over years. Finally, the paper is concluded in Section 5.

2 Passive Ultrasonic RFID Localization Model

The proposed pipeline pig is contained of three major parts: the cleaning sphere head, the pig robot body and the polyurethane (PU) sealing tail (Fig.1). When the pig swims in pipelines, there will be a relative clean room between its cleaning sphere head and PU sealing tail. The clean room is good for the work process of ultrasonic devices.

Ultrasonic RFID tags are installed in the outside wall of each pipeline segment, and can be discovered by the pig when it is located in the sensitive area of RFID tags. As soon as one of the tags is discovered, counting wheels will be set to zero and start a new counting cycle. At these points, the location of a pipeline pig can be expressed as a combination of the global location of a RFID tag and the local location of a counting wheel along the long-distance pipeline.

$$L_{Pig} = L_i + L_{CountingWheel} \tag{1}$$

Where, L_i is the calibrated location of the discovered RFID tag, i is the ID code of the tag, and $L_{CountingWheel}$ is the local location of a counting wheel.

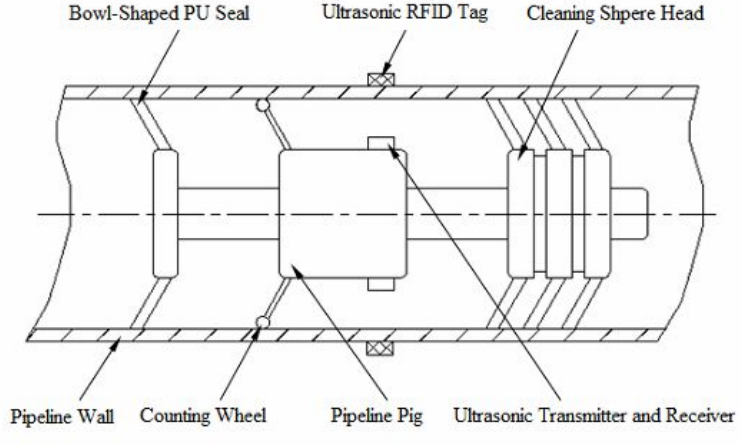


Fig. 1. Major parts of the pipeline pig's inner localization system

During construction, RFID tags are placed at the beginning part and near the welding seam of each pipeline segment firstly. After construction, their location will be calibrated by hand-held readers accurately. Once the calibration of all RFID tags completed, a location map of pipeline segments is forming. Global localization of a pipeline segment has been changed into a simple task of looking up a calibrated table. By this measure, even missing any pipeline segment, there will be no dislocation happened after the missing point. What is more, in virtue of using the same methodology to calibrate a pipeline and to discover a defect by workers, there is no cumulative error needed to be considered for global localization.

Generally, there are four steps existed in a complete discovering process (Fig. 2): First of all, one of the ultrasonic transmitter emits an energy concentrated and powerful penetrating ultrasonic wave; Next, energy transducer-I converts the mechanical wave into low frequency electromagnetic wave; Thirdly, the RFID tag absorbs energy of electromagnetic wave to reach its working voltage and emits its ID code through modulated electromagnetic wave; Finally, energy transducer-II converts the modulated electromagnetic wave to an ultrasonic wave for ultrasonic receiver discovering the coded pipeline ID.

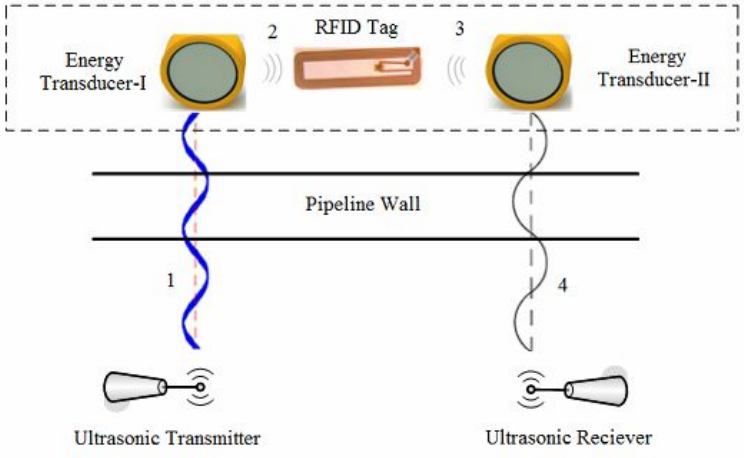


Fig. 2. Working sequence of an ultrasonic RFID tag

As everyone knows, there are three fields existed in the wave's direction of propagation: reactive near-field region, radiating near-field region and radiating far-field region. The reactive near-field is a field storing energy, where transformer model can be applied. In the other two fields, electromagnetic waves have shaken off from antennae, and the radar cross-section model can be applied. The borderline between reactive near-field region and radiating near-field region is [15]:

$$R_1 = \lambda / 2\pi \tag{2}$$

Where, R_1 is the radius of the borderline sphere between reactive near-field region and radiating near-field region, and λ is the wave length.

While the borderline between radiating near-field region and radiating far-field region is:

$$R_2 = 2D^2 / \lambda \tag{3}$$

Where, R_2 is the radius of the borderline sphere between radiating near-field region and radiating far-field region, and D is the diameter of antenna.

The power loss of RFID tag at a distance R from RFID reader is:

$$P_{Tag} = EIRP \times G_{Tag} \left(\frac{\lambda}{4\pi R} \right)^2 \tag{4}$$

Where, P_{Tag} is the power loss, $EIRP$ is the root mean square (RMS) value of radiating power of antenna, and G_{Tag} is the gain of the RFID tag.

And the received power of RFID Reader at a distance R from the RFID tag is:

$$P_{Rx} = \frac{P_{Tx} G_{Tx} G_{Rx} \lambda^2 \sigma}{(4\pi)^3 R^4} \tag{5}$$

Where, P_{Tag} is the received power, P_{Tx} is the transmitting power of RFID reader, G_{Tx} is the gain of transmitting antenna, G_{Rx} is the gain of receiving antenna, and σ is the radar cross section.

For ultrasonic waves of 20 kHz to 10 MHz, its wave length is about 17 to 0.034 mm, and R_1 is about 2.71 to 0.0054 mm. According to market research, the diameter of ultrasonic antenna for commercial usage is 16 to 70 mm. Therefore, R_2 is about 0.188 to 1814.814 m. Compared to the distance between ultrasonic transmitter loaded on the pig and the ultrasonic receiver installed on the outside wall of the pipeline, a conclusion can be reached obviously that ultrasonic transmitter and receiver work in their radiating near-field region.

After energy transmitter-I received ultrasonic wave, electromagnetic waves will function to arouse RFID tags. For electromagnetic waves of 20 kHz to 10 MHz, the wave length is about 15000 to 30 m, and R_1 is about 2388.54 to 4.78 m. As RFID tags are placed near the energy transmitters and the distance can be counted in centimeters, a similar conclusion can be reached that RFID tags work in their reactive near-field region.

3 Pig Launching and Pig Receiving

According to the pipeline construction standards, there should be some pump station arranged for pumping oil along long-distance pipelines. And in order to solve the pipeline cleaning issues, there always are some cleaning sphere launching and receiving systems constructed. As the pipeline pig proposed in the paper is followed after a cleaning sphere head in structure, it is naturally considered that pig launching and pig receiving will comply with homologous rules to make the devices and methodologies simplified. The operation steps of pig launching and pig receiving are described in detail as follows.

3.1 Pig Launching

As shown in Fig. 3, the integrated pipeline pig launching system is a bypass loop structure that is composed of a bypass valve-I (B), a bypass valve-II (C), a quick-open door, a relief valve (D), a pressure meter, a launch valve (A), and a pig passage indicator. Oil flows in the main pipeline upward and turns right as a routine. Before a pipeline pig is needed to be launched, the launching room is unpressured, valve A and valve C closed, but valve B and valve D opened. As soon as the launching room is empty of oil and the pressure meter reaches the normal atmospheric pressure, a pig launching process is ready.

Launching steps:

1. Open the quick-open door, push the pipeline pig into the segment of variable diameters, and make it fitting tightly with the pipeline.
2. Close the quick-open door correctly, and open valve C slowly to raise the inner pressure of launching room.
3. Close valve D as soon as the launching room is full of oil.
4. Open valve A, and keep valve C open to prepare for launching.
5. Half close valve B to increase the oil flow in valve C, then close valve B until the pipeline pig has been launched into the main pipeline and has been discovered by pig passage indicator.
6. Open valve B fully. After that, close valve A and C fully, and open valve D.

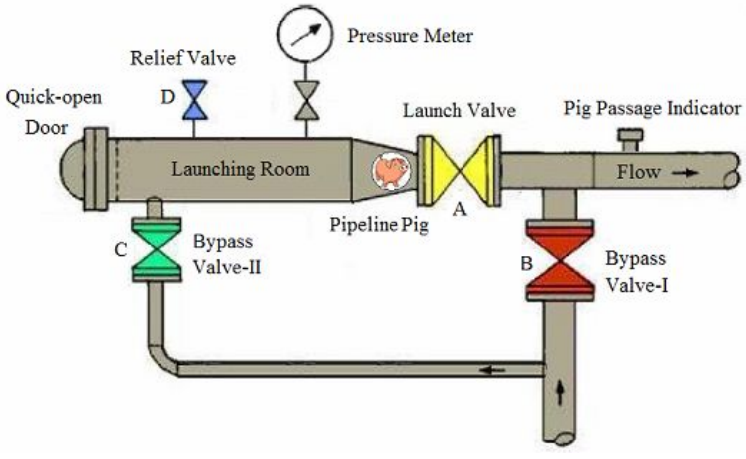


Fig. 3. Structure of a pipeline pig launching system

3.2 Pig Receiving

As shown in Fig. 4, the integrated pipeline pig receiving system is also a bypass loop structure that is composed of a bypass valve-I (B), a bypass valve-II (C), a quick-open door, a relief valve (D), an outlet valve (E), a pressure meter, a receive valve (A), and a pig passage indicator. Oil flows in the main pipeline rightward and turns down as a routine. Before a pipeline pig is needed to be received, the receiving room is unpressured, valves A and C closed, but valves B, D and E opened. As soon as the receiving room is empty and the pressure meter reaches the normal atmospheric pressure, a pig receiving process is ready.

Receiving steps:

1. Close valves D and E, open valve A and C to prepare for pig receiving.
2. Close valve B slowly to push pipeline pig into the middle of valve A and pipe tee and into the receiving room.

3. Open valve B, close valves A and C when pig passage indicator discovered the pipeline pig.
4. Open valves D and E to relief the inner pressure of pig receiver.
5. Open the quick-open door and pull the pipeline pig out after confirming there is no inner pressure in pig receiver.
6. Close the quick-open door and confirm that it is closed correctly.

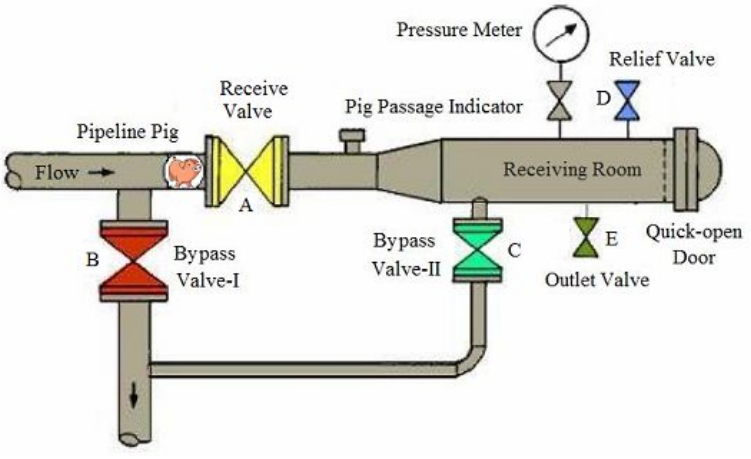


Fig. 4. Structure of a pipeline pig receiving system

4 Discussions

In case of potential malfunction, three RFID tags are placed around the outer wall of each pipeline segment to improve the identification rate. Their orientations are leftward, rightward and upward, as shown in Fig. 5. The identification code of a tag can be used to determine orientations of defects discovered in the pipeline segment. As shown in the figure, the pipeline is projected into a torus section. Define the upward direction as positive x direction, define the rightward direction as positive y direction, and define the inward direction as positive z direction, a right-hand coordinate system is forming. Suppose the top tag is located at 0°, the right tag is located at 90°, the left tag is located at 270°, and the orientation of a potential defect can be defined as angle α from x-axis to the defect. Even if one of the tags is malfunction, angle α can be deducted from any identified tag of the other two.

Another issue is the metal disturbances of pipelines. As the ultrasonic transmitter and receiver work in their radiating near-field region, there are metal disturbances needed to be considered when RFID tags placed. Metallic pipelines acts as capacitors or inductors in the radiating near-field region. And a combined antenna will be formed with RFID antenna and metallic pipelines. But the effect of interface capacitance contributes more than the inductance effect for the size differences in the three directions of pipelines. The interface capacitance can be computed as:

$$C = \varepsilon_0 \varepsilon A / \delta \tag{6}$$

Where, ε_0 is the permittivity of vacuum, ε is the relative permittivity, A is the effective coverage area, and δ is the distance between an ultrasonic transmitter and a RFID tag. In construction phase, only the δ can be easily controlled. Apparently, a big distance will be better for controlling the interface capacitance. But with the distance increasing, the electromagnetic radiation energy decreases too. Too less energy will cause passive ultrasonic RFID tags failing to emit. Therefore, there will be a proper gap reserved between RFID tags and metallic pipelines.

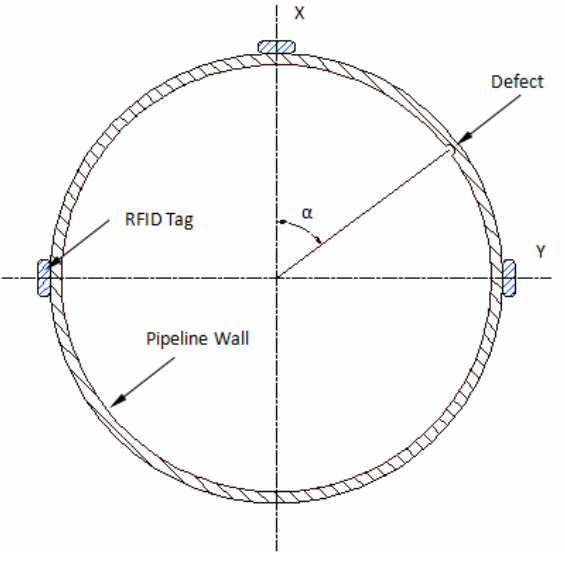


Fig. 5. Layout of three RFID tags outside the pipeline wall

5 Conclusions

Passive ultrasonic RFID is a combination application of mechanical waves and electromagnetic waves. The combination gives passive ultrasonic RFIDs outstanding penetrating capacity and excellent radio identification ability. These marvelous characters make them easy to be discovered by both pipeline pigs and maintenance workers. Analysis and computations show that the ultrasonic transmitter and receiver work in the radiating near-field regions, while the RFID tags work in their reactive near-field regions. In consideration for potential malfunction during years' running, three tags are placed around the outer wall of each pipeline segment to improve their identification rates and avoid of potential malfunctions. As orientations of the three tags are scheduled as top, left and right strictly, the methodology can be extended to determine orientations of pipeline pigs and discover defects.

Acknowledgments

The work is supported by National Natural Science Foundation of China under Grant No. 50775080 and 50835004. The authors are grateful to the editors and the anonymous reviewers for helpful comments.

References

1. Ren, W.J., Sun, Q.J., Lin, B.S., Hou, F.C.: Research on Pipeline Leak Positioning Based on Adaptive Immune Algorithm. Information and Control 36(5), 634–638 (2007)
2. About Pigs. Pipeline Pigging Products and Services Association (PPSA), http://www.ppsa-online.com/about-pigs.php
3. Li, J.Y., Chen, H.J., Zhang, X.H., Deng, Z.Q.: Localization Technique of Pipeline Robot Based on Multi-sensor Data Fusion. Control and Decision 21(6), 661–665 (2006)
4. Wang, Z.W., Cai, Q.X., Luan, N., Zhang, L.: Autonomous Localization Technique of Submarine In-pipe Robot based on Multi-sensor Data Fusion. Journal of Shanghai Jiao Tong University 42(10), 1707–1711 (2008)
5. Qi, H.M., Zhang, X.H., Chen, H.J., Ye, J.R.: Tracing and Localization System for Pipeline Robot. Mechatronics 19(1), 76–84 (2009)
6. Li, J.Y.: Research on Localization Technique of Pipeline Robot Based on Ultra Low Frequency Electromagnetic Wave. Ph.D. Dissertation, Harbin Institute of Technology (2006)
7. Lim, H.S., Choi, B.S., Lee, J.M.: An Efficient Localization Algorithm for Mobile Robots based on RFID System. In: SICE-ICASE International Joint Conference, Bexco, Busan, Korea, pp. 5945–5950 (2006)
8. Chae, H.S., Han, K.S.: Combination of RFID and Vision for Mobile Robot Localization. In: Intelligent Sensors, Sensor Networks and Information Processing Conference, pp. 75–80 (2006)
9. Wilson, P., Prashanth, D., Aghajan, H.: Utilizing RFID Signaling Scheme for Localization of Stationary Objects and Speed Estimation of Mobile Objects. In: IEEE International Conference on RFID, pp. 94–99 (2007)
10. Zhang, W.F., Shi, J.W., Zhu, S.X., Tao, Y.L.: Application of Detection System for Electronic Marker. Gas and Heat 26(2), 13–15 (2006)
11. 3M RFID-Based Underground Marking System Chosen for Shanghai World Exposition Site. Business Wire (2008), http://findarticles.com/p/articles/mi_m0EIN/is_2008_March_6/ai_n24377165/
12. Nondestructive testing of oil and gas steel pipeline. National Standards of the People's Republic of China. SY/T 4109-2005 (2005)
13. Holm, S.: Hybrid ultrasound-RFID indoor positioning: Combining the best of both worlds. In: IEEE International Conference on RFID, pp. 155–162 (2009)
14. Oh, J.H., Finn, A., Peng, P.Y., Kang, P.J.: Passive Ultrasonic RFID Elevator Positioning Reference System. US Patent No. 7441631 (2008)
15. Zhou, X.G., Wang, X.H., Wang, W.: System Design, Simulation and Application of RFID, pp. 9–12. Posts and Telecom Press (2008)

Allocation of Measurement Error of Shield's Pose

Minghua Pan, Bin Zhao, and Guoli Zhu

State Key Laboratory of Digital Manufacturing Equipment and Technology, Huazhong University of Science & Technology, Wuhan, 430074, China
glzhu@mail.hust.edu.cn

Abstract. To improve the measurement precision of the shield's pose, the factors which influence the measurement results are studied in this paper. Coordinate transformation formula is applied to calculate the shield's pose by transforming the relative coordinate of the shield's cutter head center in the laser target coordinate system to the world coordinate. The relative coordinate of the cutter head center is calibrated by transforming the formula inversely. The calibration error of the relative coordinate caused by the measurement error of the attitude angle is analyzed detailedly. The position measurement error of the cutter head center is deduced based on the principle of error transfer. Finally, the impact of each attitude angle's measurement error on the cutter head center is obtained by traversal algorithm in the variation range of attitude angle. According to the different influence degree, the measurement error of each attitude angle is proportionally allocated.

Keywords: Shield machine, measurement error, error transfer, error allocation.

1 Introduction

In the tunnel construction, the shield machine must be driven forward along the DTA (Designed Tunnel Axis) which is designed in the world coordinate system. In the process of tunnel driving, it is necessary to measure the pose of the shield machine real-timely to drive it accurately. The pose error of the shield machine can be obtained by comparing the measurement results with the DTA. What we concern in the tunnel construction is the horizontal deviation and vertical deviation[1] of the shield machine's cutter head center and tail center, so the main measurement task is to obtain the positions of the two points. For the measuring principles of the two points are the same, only the measurement error of the cutter head center is discussed in this paper.

The total station, which determines the position of the target by measuring the horizontal angle, the vertical angle and the range, is widely applied to the tunnel construction for its high precision in range measurement and angular measurement. Moreover, its powerful measurement performance of automatization and intelligentization, which is named Automatic Target Recognition[2-3], enables it possible to measure moving objects.

The laser target, which has many advantages such as small volume, convenient installation and so on, is used to measure the attitude angle of shield machine together with the total station. However, the measurement error of the attitude angle will result

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 288–296, 2009.

in the measurement error of the cutter head center's position. Therefore, it is necessary to analyze error transfer in the measurement process, and allocate the measurement error of attitude angle to improve the measurement precision of cutter head center.

In this paper, the measurement principle of the position of cutter head center is described in Sec 2. To measure the position of cutter head center, the calibration for the relative coordinate of the cutter head center in the laser target coordinate system and the analysis of the calibration error are studied in Sec. 3. After that, the measurement error of the cutter head center's position is also analyzed in Sec 4. In Sec 5, the impact of each attitude angle on the cutter head center is analyzed and the measurement error of each attitude angle is allocated reasonably. Sec 6 is the conclusion.

2 The Measurement Principle of the Position of Cutter Head Center

As shown in Figure 1, the shield machine looks like a cylinder and its axis is the line between the cutter head center and the tail center. The laser target is fixed in the shield machine and the target axis (a-axis) is nearly parallel to the shield axis. The total station, which is installed on the tunnel segments, aims at the laser target without any obstacle. It can emit laser beam to measure the world coordinate of the prism in the laser target, which is named (x_0, y_0, z_0). The horizontal angle γ of the a-axis in the world coordinate system can be obtained from the incident angle of the laser measured by the laser target and the horizontal angle of the laser. The roll angle α and the pitch angle β of the laser target can be measured by the dual-axes inclinometer which is installed in the laser target. One of the axes in the inclinometer is parallel to the a-axis and the other is perpendicular to it. The laser target coordinate system named o-abc is established and its origin point is (x_0, y_0, z_0). The coordinate system o-abc rotates together with the laser target and the shield machine when the shield changes its pose. The roll angel α is the one by which the laser target rotates around the a-axis and it will be zero when b-axis is horizontal. When a-axis is horizontal, the pitch angle β is zero.

For the laser target is fixed in the shied, the coordinate of the cutter head center (x_1, y_1, z_1) in the coordinate system o-abc is constant, but what we want to know is the world coordinate of the cutter head center, which is named (x, y, z). After the relative coordinate (x_1, y_1, z_1) in the laser target coordinate system is calibrated, the transformation matrix is deduced that will transfer the coordinate (x_1, y_1, z_1) to the world coordinate system. It can be thought of as a series of three rotations[4-6], each around an axis of world coordinate system. The first rotation is that (x_1, y_1, z_1) rotates around the X-axis by roll angle α, the second rotation is around the Y-axis by pitch angle β and the third one is around the Z-axis by horizontal angle γ. All the rotation matrixes are combined into one matrix named M. To obtain the world coordinate (x, y, z) after the three rotations, it is necessary to translate the origin point of o-abc coordinate system to the world coordinate system. The process of the rotations and translation of relative coordinate (x_1, y_1, z_1) is expressed by:

$$[x \quad y \quad z]=[x_1 \quad y_1 \quad z_1]\times M+[x_0 \quad y_0 \quad z_0] \tag{1}$$

where $M=\begin{bmatrix}1 & 0 & 0\\ 0 & \cos\alpha & -\sin\alpha\\ 0 & \sin\alpha & \cos\alpha\end{bmatrix}\times\begin{bmatrix}\cos\beta & 0 & \sin\beta\\ 0 & 1 & 0\\ -\sin\beta & 0 & \cos\beta\end{bmatrix}\times\begin{bmatrix}\cos\gamma & \sin\gamma & 0\\ -\sin\gamma & \cos\gamma & 0\\ 0 & 0 & 1\end{bmatrix}$

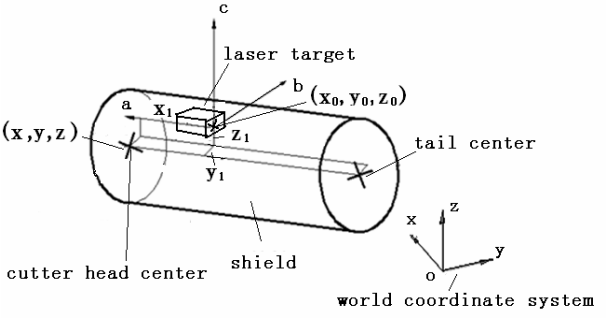


Fig. 1. The position and orientation of the laser target in the shield machine

3 Calibration Error of the Relative Coordinate

In order to measure the position of the cutter head center, the relative coordinate (x_1, y_1, z_1) needs to be calibrated firstly. When the shield machine is in static state, the world coordinate of the cutter head center (x, y, z) can be measured accurately by other methods[7]. At the same time, we can obtain the world coordinate of the laser target's prism (x_0, y_0, z_0) from the total station and the three attitude angles from the laser target. The theoretical value of x_1, y_1 and z_1 can be calculated by transforming formula (1) to :

$$[x_1 \quad y_1 \quad z_1]=[x-x_0 \quad y-y_0 \quad z-z_0]M^{-1} \tag{2}$$

where

$$M^{-1}=M_\gamma^{-1}M_\beta^{-1}M_\alpha^{-1}=\begin{bmatrix}\cos\gamma & -\sin\gamma & 0\\ \sin\gamma & \cos\gamma & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}\cos\beta & 0 & -\sin\beta\\ 0 & 1 & 0\\ \sin\beta & 0 & \cos\beta\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\ 0 & \cos\alpha & \sin\alpha\\ 0 & -\sin\alpha & \cos\alpha\end{bmatrix}$$

In this paper, $\Delta\alpha$, $\Delta\beta$ and$\Delta\gamma$ are used to describe the measurement error of the three attitude angles separately. When the measurement errors are considered, it means that the coordinate system o-abc rotates around axes of world coordinate system by the measurement error after it rotates by the attitude angle. So the rotation matrixes of the three measurement errors are:

$$M_{\Delta\alpha}^{-1}=\begin{bmatrix}1 & 0 & 0\\ 0 & \cos\Delta\alpha & \sin\Delta\alpha\\ 0 & -\sin\Delta\alpha & \cos\Delta\alpha\end{bmatrix}$$

$$M_{\Delta\beta}^{-1}=\begin{bmatrix}\cos\Delta\beta & 0 & -\sin\Delta\beta\\ 0 & 1 & 0\\ \sin\Delta\beta & 0 & \cos\Delta\beta\end{bmatrix}$$

$$M_{\Delta\gamma}^{-1}=\begin{bmatrix}\cos\Delta\gamma & -\sin\Delta\gamma & 0\\ \sin\Delta\gamma & \cos\Delta\gamma & 0\\ 0 & 0 & 1\end{bmatrix}$$

The matrix M^{-1} in formula (2) should be changed to

$$M_1^{-1}=M_\gamma^{-1}M_{\Delta\gamma}^{-1}M_\beta^{-1}M_{\Delta\beta}^{-1}M_\alpha^{-1}M_{\Delta\alpha}^{-1}$$

As we know, it is acceptable that $\sin\Delta\approx\Delta$ and $\cos\Delta\approx1$ when the Δ is small enough. The measurement error of attitude angle is as small as several milliradians, so the above matrixes can be expressed as follows:

$$M_{\Delta\alpha}^{-1}\approx\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & \Delta\alpha\\ 0 & -\Delta\alpha & 1\end{bmatrix},$$

$$M_{\Delta\beta}^{-1}\approx\begin{bmatrix}1 & 0 & -\Delta\beta\\ 0 & 1 & 0\\ \Delta\beta & 0 & 1\end{bmatrix},$$

$$M_{\Delta\gamma}^{-1}\approx\begin{bmatrix}1 & -\Delta\gamma & 0\\ \Delta\gamma & 1 & 0\\ 0 & 0 & 1\end{bmatrix}$$

According to the above analysis, the calibration error of x_1, y_1 and z_1 can be shown as

$$\begin{bmatrix}\Delta x_1 & \Delta y_1 & \Delta z_1\end{bmatrix}=\begin{bmatrix}x-x_0 & y-y_0 & z-z_0\end{bmatrix}M_1^{-1}-\begin{bmatrix}x-x_0 & y-y_0 & z-z_0\end{bmatrix}M^{-1} \tag{3}$$

If the high order terms of $\Delta\alpha$, $\Delta\beta$ and $\Delta\gamma$ are neglected, formula (3) can be expressed as follows:

$$\begin{cases}\Delta x_1=(-\Delta\beta\cos\gamma\sin\beta-\Delta\gamma\sin\gamma\cos\beta)(x-x_0)+(-\Delta\beta\sin\gamma\sin\beta+\Delta\gamma\cos\gamma\cos\beta)(y-y_0)\\ \qquad +\Delta\beta(z-z_0)\cos\beta\\ \Delta y_1=(\Delta\alpha\sin\gamma\sin\alpha-\Delta\gamma\cos\gamma\cos\alpha+\Delta\alpha\cos\gamma\sin\beta\cos\alpha+\Delta\beta\cos\gamma\cos\beta\sin\alpha\\ \qquad -\Delta\gamma\sin\gamma\sin\beta\sin\alpha)(x-x_0)+(-\Delta\alpha\cos\gamma\sin\alpha-\Delta\gamma\sin\gamma\cos\alpha+\Delta\alpha\sin\gamma\sin\beta\cos\alpha\\ \qquad +\Delta\beta\sin\gamma\cos\beta\sin\alpha+\Delta\gamma\cos\gamma\sin\beta\sin\alpha)(y-y_0)+(\Delta\beta\sin\beta\sin\alpha-\Delta\alpha\cos\beta\cos\alpha)(z-z_0)\\ \Delta z_1=(-\Delta\alpha\sin\gamma\cos\alpha-\Delta\gamma\cos\gamma\sin\alpha+\Delta\alpha\cos\gamma\sin\beta\sin\alpha-\Delta\beta\cos\alpha\cos\gamma\cos\beta\\ \qquad +\Delta\gamma\sin\gamma\sin\beta\cos\alpha)(x-x_0)+(\Delta\alpha\cos\gamma\cos\alpha-\Delta\gamma\sin\gamma\sin\alpha+\Delta\alpha\sin\gamma\sin\beta\sin\alpha\\ \qquad -\Delta\beta\sin\gamma\cos\beta\cos\alpha-\Delta\gamma\cos\gamma\sin\beta\cos\alpha)(y-y_0)+(-\Delta\alpha\cos\beta\sin\alpha-\Delta\beta\sin\beta\cos\alpha)(z-z_0)\end{cases}$$

When the relative coordinate (x_1, y_1, z_1) is in calibration, the shield is normally placed in horizontal direction ,the roll angle α and the pitch angle β are nearly zero. The horizontal angle γ is independent on (x_1, y_1, z_1) and can be set as zero. It is obvious that the components in the above expressions which include both angle error Δ and sine value of attitude angles are small enough to be neglected. Δx_1, Δy_1 and Δz_1 are expressed as follows:

$$\begin{cases} \Delta x_1 = \Delta\gamma(y - y_0) + \Delta\beta(z - z_0) \\ \Delta y_1 = -\Delta\gamma(x - x_0) - \Delta\alpha(z - z_0) \\ \Delta z_1 = -\Delta\beta(x - x_0) + \Delta\alpha(y - y_0) \end{cases} \tag{4}$$

As stated above, (x, y, z) and (x_0, y_0, z_0) are the world coordinate of the cutter head center and the prism in the laser target. When γ=0, $\alpha \approx 0$ and $\beta \approx 0$, it is easy to know that $x-x_0=x_1$, $y-y_0 \approx y_1$ and $z-z_0 \approx z_1$, so Δx_1, Δy_1 and Δz_1 can be expressed as:

$$\begin{cases} \Delta x_1 = y_1\Delta\gamma + z_1\Delta\beta \\ \Delta y_1 = -x_1\Delta\gamma - z_1\Delta\alpha \\ \Delta z_1 = -x_1\Delta\beta + y_1\Delta\alpha \end{cases} \tag{5}$$

For the value of x_1 is much larger than y_1 and z_1, $\Delta\alpha$ has less influence than $\Delta\beta$ and $\Delta\gamma$ in calibration of the relative coordinate. It can be found from formula (5) that the calibration error depends on the place where the laser target is installed in the shield machine. The farther the laser target is installed from the shield axis and the cutter head center, the larger the error of relative position will be.

4 Measurement Error Analysis for the Position of Cutter Head Center

In the tunnel construction, the total station emits laser beam to measure the world coordinate of the prism which is in the laser target, at the same time the computer will receive the data of the three attitude angles and calculate the position of the cutter head center by formula (1).

The error value of (x, y, z) can be deduced by formula (1). If the parameters in formula (1) are named as T_i (i=1,2,…), the measurement error of (x, y, z) can be expressed as follows by the principle of error transfer[7]:

$$\begin{cases} \Delta x = \sqrt{\sum_{i=1}^{n} (\partial x / \partial T_i \, \Delta T_i)^2} \\ \Delta y = \sqrt{\sum_{i=1}^{n} (\partial y / \partial T_i \, \Delta T_i)^2} \\ \Delta z = \sqrt{\sum_{i=1}^{n} (\partial z / \partial T_i \, \Delta T_i)^2} \end{cases} \tag{6}$$

Expression as follows shows the error components of coordinate x:

$$\begin{cases} \partial x/\partial x_1 \, \Delta x_1 = \cos\beta\cos\gamma(y_1\Delta\gamma + z_1\Delta\beta) \\ \partial x/\partial y_1 \, \Delta y_1 = (\cos\alpha\sin\gamma - \sin\alpha\sin\beta\cos\gamma)(x_1\Delta\gamma + z_1\Delta\alpha) \\ \partial x/\partial z_1 \, \Delta z_1 = (\cos\alpha\sin\beta\cos\gamma + \sin\alpha\sin\gamma)(x_1\Delta\beta - y_1\Delta\alpha) \end{cases} \tag{7}$$

$$\begin{cases} \dfrac{\partial x}{\partial\alpha}\Delta\alpha = ((\cos\alpha\sin\beta\cos\gamma + \sin\alpha\sin\gamma)y_1 + (\sin\alpha\sin\beta\cos\gamma - \cos\alpha\sin\gamma)z_1)\Delta\alpha \\ \dfrac{\partial x}{\partial\beta}\Delta\beta = (-x_1\sin\beta\cos\gamma + y_1\sin\alpha\cos\beta\cos\gamma - z_1\cos\alpha\cos\beta\cos\gamma)\Delta\beta \\ \dfrac{\partial x}{\partial\gamma}\Delta\gamma = (-x_1\cos\beta\sin\gamma - y_1(\sin\alpha\sin\beta\sin\gamma + \cos\alpha\cos\gamma) \\ \qquad + z_1(\cos\alpha\sin\beta\sin\gamma - \sin\alpha\cos\gamma))\Delta\gamma \end{cases} \tag{8}$$

With the same method, the error components of coordinate y and z can also be expressed. From formula (6), (7) and (8), we can obtain the measurement error of (x, y, z).

5 Analysis and Allocation of the Measurement Error

5.1 Analysis for Influence Degree of Each Attitude Angle

The error analysis is carried out with a shield machine designed by Mitsubishi Heavy Industries, Ltd. The shield is 10 meters long with the diameter of 6.34 meters. The relative coordinate of the cutter head center in the laser target coordinate system is x_1=5m, y_1=0.5m, z_1=2m. The variation range of the three attitude angles are set at $\alpha\in$ (-30°,30°), $\beta\in$(-15°,15°), $\gamma\in$(0°,360°). All the analysis and conclusions in the chapter 5 and 6 are based on the above parameters.

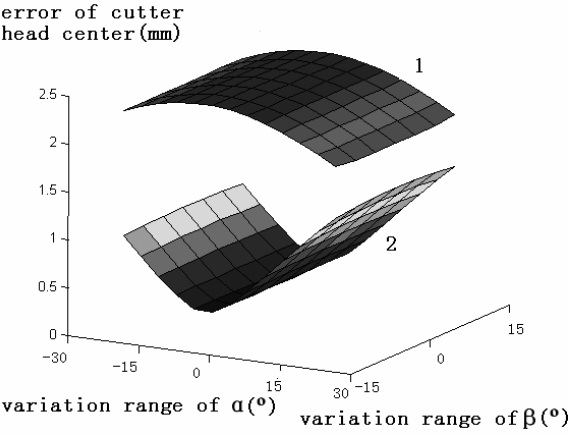


Fig. 2. Measurement error of the cutter head center: (1) error of coordinate x and y; (2) error of coordinate z

In order to analyze influence degree of each attitude angle's error to the measurement of the cutter head center, $\Delta\alpha$, $\Delta\beta$ and$\Delta\gamma$ should be considered separately. As the measurement accuracy of attitude angles reaches milliradian level, the measurement error of (x, y, z) can be calculated by formula (6), (7) and (8) when $\Delta\alpha$=1mrad, $\Delta\beta$=0 and $\Delta\gamma$=0. The maximum of measurement error is obtained by traversal algorithm in the variation range of the attitude angles and is shown in Fig.2.

When $\Delta\beta$=1mrad, $\Delta\alpha$=0 and $\Delta\gamma$=0, the maximum of measurement error of (x, y, z) is shown in Fig.3.

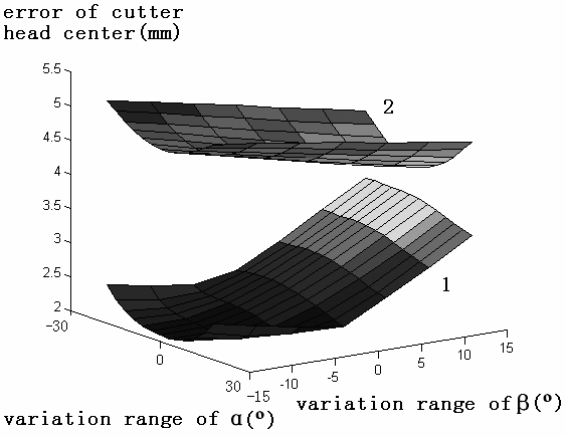


Fig. 3. Measurement error of cutter head center: (1) error of coordinate x and y; (2) error of coordinate z

When $\Delta\gamma$=1mrad, $\Delta\alpha$=0 and $\Delta\beta$=0, the maximum of measurement error of (x, y, z) is shown in Fig.4.

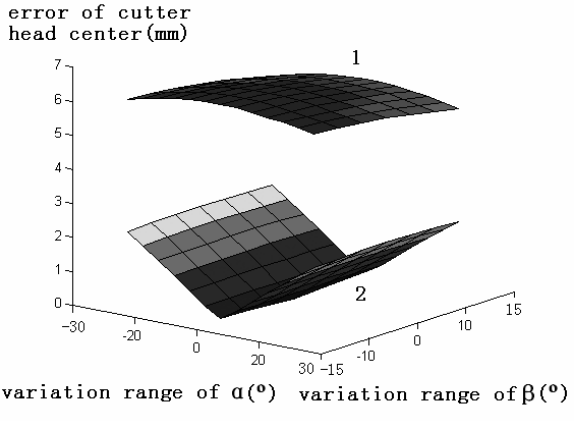


Fig. 4. Measurement error of cutter head center: (1) error of coordinate x and y; (2) error of coordinate z

As shown in Fig.2, Fig.3 and Fig.4, $\Delta\alpha$ and $\Delta\gamma$ have greater impact on x and y than on z while$\Delta\beta$ has greater impact on z than on x and y. By comprehensive consideration of the coordinate errors in three figures, the influence degree of $\Delta\alpha$ is nearly half of that of$\Delta\beta$ and $\Delta\gamma$.

Based on the analysis above, the range of $\Delta\alpha$ should be allocated as two times than $\Delta\beta$ and $\Delta\gamma$ to make the measurement error of x, y and z nearly be equal.

5.2 Error Allocation of Each Attitude Angle

In terms of the technical requirements in the tunnel construction, the tunnel transfixion error should be less than 30mm. Therefore, the horizontal position and vertical position measurement errors of the cutter head center should be less than 10mm. Vertical position measurement error is Δz and horizontal position measurement error is:

$$\delta = \sqrt{\Delta x^2 + \Delta y^2}$$

On the basis of the analysis in section 5.1, the calculation result by traversal algorithm in the variation range of the attitude angle is illustrated in Fig.5. It can be found that when $\Delta\beta=\Delta\gamma=1$mrad, $\Delta\alpha=2$mrads, the maximum of horizontal position measurement error is 10mm and vertical position measurement error is 6.7mm. Therefore, the maximum error of α is 2mrads, the maximum errors of β and γ are 1mrad.

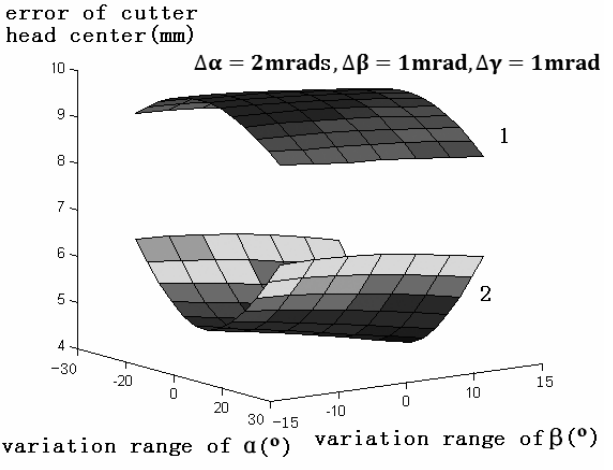


Fig. 5. Measurement error of the cutter head center: (1) error of horizontal position; (2) error of vertical position

6 Conclusion

The research on allocation for the measurement error of the cutter head center is presented in this paper. The conclusions are as follows:

a. The measurement errors of the three attitude angles have a great impact on the calibration error of the cutter head center's relative coordinate in the laser target coordinate system, the error of the roll angle α has less influence than that of the pitch angle β and the horizontal angle γ.

b. The best proportion of the angle error allocation is that the roll angle error $\Delta\alpha$ is twice than the horizontal angle error $\Delta\gamma$ and the pitch angle error $\Delta\beta$. When $\Delta\alpha$=2mrads, $\Delta\beta$=1mrad and $\Delta\gamma$=1mrad, the maximum of the total error is 10mm.

Acknowledgement

This work is supported by the National Basic Research Program of China (No. 2007CB714000).

References

1. Bob-Chow: Shield Tunnelling Technology. China Architecture & Building Press, Beijing (2004)
2. Yakar, M.: Digital elevation model generation by robotic total station instrument. J. Experimental Techniques. 33(2), 52–59 (2009)
3. Luo, Y., Zheng, Y., Xia, Z., Shaobo, W., Zhu, W.: The kinetic property and application of the total station instrument. In: Third International Symposium on Precision Mechanical Measurements, 62803E-1-6. SPIE Optical Engineering Press, Bellingham (2006)
4. Aviado, C., Gill, J., Redman, K., Ohl, R.: Methods for correlating autocollimation of theodolites and coordinate metrology in spacecraft systems. In: Optomechanical Technologies for Astronomy, 62733H-1-8. SPIE Optical Engineering Press, Bellingham (2006)
5. Pan, M., Zhu, G.: Study of Measure Methods of the Automatic Guiding System of Shield Machine. J. Construction Technology 34(6), 34–36 (2005)
6. Tang, Z., Zhao, J., Peng, G.: Measurement and calculation method for real-time attitude of tunnel boring machine. J. China Civil Engineering Journal 40(11), 92–97 (2007)
7. Zhang, X., Zhu, L.: Optimal Selection of Attitude Control Points of Shield Machine. J. China Mechanical Engineering 20(8), 902–906 (2009)
8. Han, C., Zhou, B., Yuan, X., Lian, F.: Error transfer and correction between multi-target measurements in a common clutter environment. J. Zidonghua Xuebao/Acta Automatica Sinica 33(7), 77–774 (2007)
9. Feng, D., et al.: On the Orientation Survey of the Shield in Subway Construction. J. Geotechnical Investigation and Surveying 5, 57–61 (2003)
10. Zhang, H., Gu, l.: On the measurement and the method of computation of the posture parameters of shields. J. Modern Tunnelling Technology 41(2), 14–20 (2004)

The Compliance Design on Thrust System of Shields Excavating in Various Geologic Conditions

Zhongpo Liu, Genliang Chen, Haidong Yu, and Xinmin Lai

State Key Laboratory of Mechanical System and Vibration, School of Mechanical Engineering, Shanghai Jiao Tong Unviersity, Shanghai, China, 200240
{skyliu,leungchan,hdyu,xmlai}@sjtu.edu.cn

Abstract. The redundant actuating thrust system is usually employed to comply with the complex load on the excavating face. The control strategy of these hydraulic cylinders is critical to the load transmission. The forces transmission model of the redundant actuating thrust system of shield machines was established based on the screw theory. The force transmission behavior was studied when the shields excavate in the heterogeneous geologic structures containing the hard rock and soft soil. The correlation between the group strategy of the thrust hydraulic cylinder and the load behavior on the cutterhead was discussed. A flexible group strategy on the redundant actuating hydraulic cylinders was proposed in terms of the various loads. It may improve the compliance ability of the thrust system excavating in the heterogeneous geologic structures.

Keywords: Flexible grouping, Load transmission, Various geologic conditions, Compliance.

1 Introduction

With the rapid development of the urbanization, the requirement of the tunnel for transportation is increased. The shield machines become the main excavation tool in modern tunneling construction due to the merit such as the improvement on construction safety and efficiency. Since the first shield machine appeared, much research work on their performances analysis was conducted by many researchers. The geologic structure along excavation route plays an important role in the shield design.

Many investigators have studied the correlation between geology conditions and the design of the cutters on the cutterhead. In these works, Boshkov [1] proposed a qualitative classification method on the geologic structures in the tunnel. Morrison[2] suggested a method for mining method selection. Laubscher [3] proposed a selection methodology of a corresponding large scale underground mine based on rock classification system.

As mentioned above, numerous studies have been done on estimating the shields performance mostly based on the interaction of rock properties and cutters [4], which are useful to the evaluation of the shield design. However, the resisting load on the cutterhead is closely relative with the thrust force. It is necessary to design the thrust

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 297–306, 2009.

system of shields in terms of the geological conditions. In this paper, the force transmission model of the thrust system was established. The flexible group strategy of the redundant thrust force in terms of the geologic structure is proposed, which is beneficial with the compliance design. Also, the thrust system of the shield machine also plays an important role within all design factors. Especially, the geology condition along the exaction route is complex which will result in the harmful bending moment on the cutter-head, boring direction deviated from the planed route and aggravating the wear of cutting tools, gears and bearings.

Therefore, processing the study of the force transmission effect of the thrust system is practical significant. The force transmission analysis on the thrust system which can be considered as a redundant parallel mechanism on the base of the way of cylinders installment was conducted making use of the screw theory. Based on the analysis, one to one corresponding control relationship between the input force of actuator joint and output force of the cutterhead was got, so the effect of variation of workloads on the each actuator can be obtained.

2 Force Transmission Analysis of Thrust System

The sketch map of the thrust system of shield machines is illustrated in Fig.1. In the thrust system, there are about sixteen hydraulic cylinders to provide the enormous actuating force in order to balance resistance on the cutterhead. Based on the screw theory, the actuating forces, provided by the hydraulic cylinders, can be synthesized to a resultant force screw with respect to the origin of the body-fixed framer on the head as

$$ {}_{O'}\hat{\boldsymbol{F}} = \sum_{i=1}^{16} f_i \ {}_{O'}\hat{\boldsymbol{S}}_i \tag{1} $$

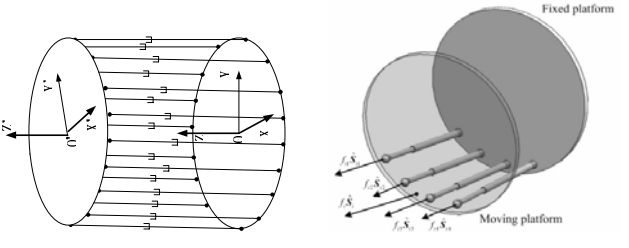


Fig. 1. Forces of thrust hydraulic cylinders

where ${}_{O'}\hat{\boldsymbol{F}}$ denotes the resultant force screw, combined both the force and torque, applied on the head related to the reference point O′, ${}_{O'}\hat{\boldsymbol{S}}_i$ represents the screw of driving force, actuated by the corresponding cylinder, with respect to the point O′, and f_i is the magnitude of the driving forces.

According to the principle of screw transformation, the driving force screw ${}_{O'}\hat{\boldsymbol{S}}_i$ can be derived from the line vector along the cylinders direction as

$$ {}_{O'}\hat{\boldsymbol{S}}_i = {}_{O'}\tilde{\boldsymbol{R}}_{P_i}\ {}_{P_i}\hat{\boldsymbol{S}}_i \tag{2} $$

where ${}_{P_i}\hat{\boldsymbol{S}}_i$ represents the original driving force screw with zero pitch, and ${}_{O'}\tilde{\boldsymbol{R}}_{P_i}$ is the 6×6 transformation matrix related to the reference points P_i and O', which can be obtained as

$$ {}_{O'}\tilde{\boldsymbol{R}}_{P_i} = \begin{bmatrix} \boldsymbol{E}_3 & \boldsymbol{0}_{3\times3} \\ {}_{O'}\tilde{\boldsymbol{\rho}}_{P_i}^T & \boldsymbol{E}_3 \end{bmatrix} \tag{3} $$

where $\boldsymbol{E}_3$ denotes the 3-order identity matrix, $\boldsymbol{0}_{3\times3}$ represents the 3×3 zero matrix, and ${}_{O'}\tilde{\boldsymbol{\rho}}_{P_i}^T$ is the skew-symmetric matrix associated with the vector ${}_{O'}\boldsymbol{\rho}_{P_i}$, pointed from P_i to O', which is given as

$$ {}_{O'}\tilde{\boldsymbol{\rho}}_{P_i} = \begin{bmatrix} 0 & -{}_{O'}z_{P_i} & {}_{O'}y_{P_i} \\ {}_{O'}z_{P_i} & 0 & -{}_{O'}x_{P_i} \\ -{}_{O'}y_{P_i} & {}_{O'}x_{P_i} & 0 \end{bmatrix} \tag{4} $$

where ${}_{O'}x_{P_i}$, ${}_{O'}y_{P_i}$, ${}_{O'}z_{P_i}$ are the coordinates of the vector ${}_{O'}\boldsymbol{\rho}_{P_i}$ in the reference frame $O\text{-}xyz$, which can be transformed from the body-fixed frame as

$$ {}_{O'}\boldsymbol{\rho}_{P_i} = {}^{O}\boldsymbol{A}^{O'}\ {}_{O'}\boldsymbol{\rho}_{P_i}^{O'} \tag{5} $$

where ${}_{O'}\boldsymbol{\rho}_{P_i}^{O'} = ({}_{O'}x_{P_i}^{O'}, {}_{O'}x_{P_i}^{O'}, {}_{O'}x_{P_i}^{O'})^T$ denotes the coordinate of the vector ${}_{O'}\boldsymbol{\rho}_{P_i}$ in the body-fixed frame, and ${}^{O}\boldsymbol{A}^{O'}$ represents the direction cosine matrix between the frame $O'\text{-}xyz$ and $O\text{-}xyz$, which can be obtained as

$$ {}^{O}\boldsymbol{A}^{O'} = \begin{bmatrix} C_\beta C_\gamma & -C_\beta S_\gamma & S_\beta \\ S_\alpha S_\beta C_\gamma + C_\alpha S_\gamma & C_\alpha C_\gamma - S_\alpha S_\beta S_\gamma & -S_\alpha C_\beta \\ S_\alpha S_\gamma - C_\alpha S_\beta C_\gamma & C_\alpha S_\beta S_\gamma + S_\alpha C_\gamma & C_\alpha C_\beta \end{bmatrix} \tag{6} $$

where α, β and γ are the Cardano angels, which means that the orientation of the moving frame could be described by three rotations motions, first is the rotation of α about the $O\text{-}x$ axis, then followed by another rotation of β about $O\text{-}y$ axis, and finally rotated about $O\text{-}z$ axis with γ. S_i and C_i ($i=\alpha,\beta,\gamma$) are their sine and cosine values.

On the other hand, the line vector of the driving forces of the hydraulic cylinders can be constructed as

$$ {}_{\mathrm{P}_i}\hat{\boldsymbol{S}}_i = \begin{bmatrix} {}_{\mathrm{P}_i}\boldsymbol{r}_{\mathrm{B}_i} \\ \boldsymbol{0}_{3\times1} \end{bmatrix} \tag{7} $$

where $\boldsymbol{0}_{3\times1}$ denotes the 3×1 zero matrix, and ${}_{\mathrm{P}_i}\boldsymbol{r}_{\mathrm{B}_i}$ represents the unit vector of the line $\mathrm{B}_i\mathrm{P}_i$, which can be obtained as

$$ {}_{\mathrm{P}_i}\boldsymbol{r}_{\mathrm{B}_i} = {}_{\mathrm{O}'}\boldsymbol{r}_{\mathrm{O}} + {}_{\mathrm{O}'}\boldsymbol{\rho}_{\mathrm{P}_i} - {}_{\mathrm{B}_i}\boldsymbol{\rho}_{\mathrm{O}} \tag{8} $$

where ${}_{\mathrm{O}'}\boldsymbol{r}_{\mathrm{O}}$ denotes the position vector of the frame $\mathrm{O}\text{-}xyz$ with respect to $\mathrm{O}'\text{-}xyz$, and ${}_{\mathrm{B}_i}\boldsymbol{\rho}_{\mathrm{O}}$ represents the vector of OB_i in the reference frame.

Then, Eq.1 can be rewritten in the matrix form as follows

$$ {}_{\mathrm{O}'}\hat{\boldsymbol{F}} = \boldsymbol{T}\boldsymbol{f} \tag{9} $$

where $\boldsymbol{T}$ is the 6×16 transmission matrix related the driving forces to the legs, and $\boldsymbol{f}$ denotes the vector combining all the driving forces actuated by the cylinders, both of which can be assembled as

$$ \boldsymbol{T} = \begin{bmatrix} {}_{\mathrm{O}'}\tilde{\boldsymbol{R}}_{\mathrm{P}_1\ \mathrm{P}_1}\hat{\boldsymbol{S}}_1, & {}_{\mathrm{O}'}\tilde{\boldsymbol{R}}_{\mathrm{P}_2\ \mathrm{P}_2}\hat{\boldsymbol{S}}_2, & \cdots, & {}_{\mathrm{O}'}\tilde{\boldsymbol{R}}_{\mathrm{P}_{16}\ \mathrm{P}_{16}}\hat{\boldsymbol{S}}_{16} \end{bmatrix} \tag{10} $$

$$ \boldsymbol{f} = \begin{bmatrix} f_1 & f_2 & \cdots & f_6 \end{bmatrix}^T \tag{11} $$

Therefore, the static equilibrium equation of the shield can be stated in the screw form as

$$ {}_{\mathrm{O}'}\hat{\boldsymbol{G}} + {}_{\mathrm{O}'}\hat{\boldsymbol{F}}_a + \boldsymbol{T}\boldsymbol{f} = \boldsymbol{0} \tag{12} $$

where ${}_{\mathrm{O}'}\hat{\boldsymbol{G}}$ represents the gravity screw of the cutterhead (gravity force and its torque) with respect to the point O', and ${}_{\mathrm{O}'}\hat{\boldsymbol{F}}_a$ is the resistance force screw during the excavating process [5-8].

When the resistance force screws are detected by the sensors or determined according to the work condition, the corresponding actuating forces of the cylinders can be obtained in terms of Eq.12 as

$$ \boldsymbol{f} = -\boldsymbol{T}^{+}({}_{\mathrm{O}'}\hat{\boldsymbol{G}} + {}_{\mathrm{O}'}\hat{\boldsymbol{F}}_a) \tag{13} $$

where $\boldsymbol{T}^{+}$ represents the pseudo inverse of the transmission matrix $\boldsymbol{T}$. However, it is relatively complicated to simultaneously control the sixteen actuated cylinders separately. Therefore, in order to simplify the control system and its strategy, the sixteen actuated hydraulic cylinders are generally divided into four groups.

And the cylinders in the same group are connected together, which means that the driving forces in the same group are with the same hydraulic pressure. Then, the vector $\boldsymbol{f}$ can be rewritten in the form as

$$\boldsymbol{f} = \boldsymbol{I}_g(g_1, g_2, g_3, g_4)\boldsymbol{F}_g \tag{14}$$

where $\boldsymbol{F}_g = \begin{bmatrix} f_{g1} & f_{g2} & f_{g3} & f_{g4} \end{bmatrix}^T$ denotes the driving force vector of the groups, and $\boldsymbol{I}_g$ is the 16×4 grouping matrix, whose elements are assigned to 0 or 1 on the basis of the grouping strategy. When the subscript of the row in $\boldsymbol{I}$ belongs to the group g_i, the i^{th} element of this row is assigned to 1, and the others are set to 0.

Thus, the static equilibrium equation of the thrust system can be simplified by substituting Eq.14 into Eq.12 as

$$_{O'}\hat{\boldsymbol{G}} + {}_{O'}\hat{\boldsymbol{F}}_a + \boldsymbol{T}\boldsymbol{I}_g\boldsymbol{F}_g = \boldsymbol{0} \tag{15}$$

Then, the driving forces, in groups, can be obtained by solving Eq.15 as

$$\boldsymbol{F}_g = -(\boldsymbol{T}\boldsymbol{I}_g)^+({}_{O'}\hat{\boldsymbol{G}} + {}_{O'}\hat{\boldsymbol{F}}_a) \tag{16}$$

In Eq.16, the resultant matrix $\boldsymbol{T}\boldsymbol{I}_g$ is with the order of 6×4, therefore it is easier to get its pseudo inverse than the 6×16 transmission matrix $\boldsymbol{T}$. Furthermore, in terms of the work principle of the shield machines, the forces on the O-xy plane are balanced by the constrained forces of the tunnel, and the cutting torque of the cutterhead is provided by the variable frequency motors. Consequently, the resistance force screws exerted on the cutterhead can be reduced as

$${}_{O'}\hat{\boldsymbol{F}} = {}_{O'}\hat{\boldsymbol{G}} + {}_{O'}\hat{\boldsymbol{F}}_a = \begin{bmatrix} 0 & 0 & F_z & M_x & M_y & 0 \end{bmatrix}^T \tag{17}$$

which can be proved a zero pitch screw easily as

$$h = \frac{F \cdot {}_{O'}M}{F \cdot F} = \frac{\begin{bmatrix} 0 & 0 & F_z \end{bmatrix}\begin{bmatrix} M_x & M_y & 0 \end{bmatrix}^T}{\begin{bmatrix} 0 & 0 & F_z \end{bmatrix}\begin{bmatrix} 0 & 0 & F_z \end{bmatrix}^T} = 0 \tag{18}$$

Therefore, the dimension of the equilibrium equations can be reduced to three by multiplying both sides of Eq.(15) with the reduction matrix as

$$\boldsymbol{R}_e\,{}_{O'}\hat{\boldsymbol{F}} + \boldsymbol{R}_e\boldsymbol{T}\boldsymbol{I}_g\boldsymbol{F}_g = \boldsymbol{0} \tag{19}$$

where $\boldsymbol{S}_l$ represents the 3×6 reduction matrix, which is given as

$$\boldsymbol{R}_e = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \tag{20}$$

Finally, the driving forces of the actuated hydraulic cylinders can be determined by solving the reduced equilibrium equations as

$$\boldsymbol{F}_g = -(\boldsymbol{R}_e \boldsymbol{TI}_g)^+ \boldsymbol{R}_{e\,O'} \hat{\boldsymbol{F}} \tag{21}$$

where the order of the resultant matrix $\boldsymbol{R}_e \boldsymbol{TI}_g$ has been reduced to 3×4.

Since the number of the equations is less than the dimension of the driving forces vector, an additional constraint should be employed for the thrust system, such as making one pair of the opposite groups has the same value of the driving forces. Then, the equilibrium equation can be finally determined as

$$\boldsymbol{R}_{e\,O'} \hat{\boldsymbol{F}} + \boldsymbol{R}_e \boldsymbol{TI}_g \boldsymbol{I}'_g \boldsymbol{F}'_g = \boldsymbol{0} \tag{22}$$

where $\boldsymbol{F}'_g = \begin{bmatrix} f_{g1} & f_{g2} & f_{g4} \end{bmatrix}^T$ is the reduction driving force vector, under the assumption that the driving force actuated in the third group is same as the first group (shown in Fig. 2). And $\boldsymbol{I}'_g$ is the coefficient matrix given as

$$\boldsymbol{I}'_g = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{23}$$

So far, the equilibrium equation has been stated by a set of statically determined linear equations. In Eq.(22), the number of unknowns equals to the number of equations, so the driving forces can be determined uniquely by solving it as

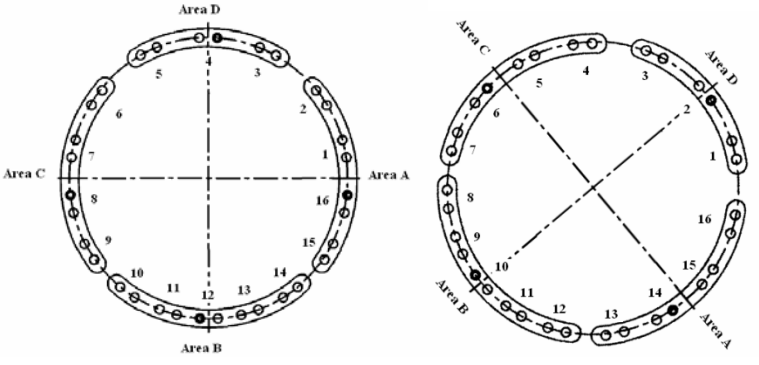
$$\boldsymbol{F}'_g = -(\boldsymbol{R}_e \boldsymbol{TI}_g \boldsymbol{I}'_g)^{-1} \boldsymbol{R}_{e\,O'} \hat{\boldsymbol{F}} \tag{24}$$

From Eq.(24), it is obvious that the driving forces of the hydraulic cylinders only depend on the grouping strategy of the driving system if the pose (position and orientation) and external forces of the head are specified according to the work condition. In other words, when the load and the pose of the head are determined, the difference in group division of the driving cylinders leads to the difference of the required driving forces of the actuators. Therefore, it provides the possibility to optimize the driving forces, according to different geologic structures, by selecting a proper way to group the cylinders.

3 Grouping Strategies on Hydraulic Cylinders

In order to provide the enormous thrust force, many hydraulic cylinders are usually employed in the shield machines. However, in the excavation process, it's impossible for the driver to control the pressure of each hydraulic cylinder simultaneous. It is necessary to divide these cylinders into several groups. If the cutting surface is composed of

two or more geologic structures, the grouping strategy of hydraulic cylinders in the thrust system should be taken into consideration in order to obtain the load compliance behavior for the various resistance forces on the cutterheads. In general, four groups, shown in Fig.2, are used in many shield machines, which are beneficial to be controlled by the drivers. Since the gravity, soil pressure and friction between the shield and ground are considered, the number of cylinders in the Area B is greater than that in the Area D. But the numbers of cylinders in the Area A and Area C are identical. This group strategy is limited when the geologic structures change in the planed direction.

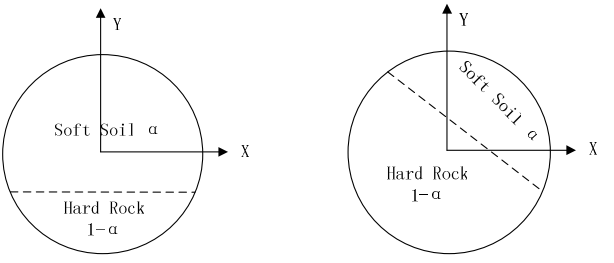


(a) Arrangement A (b) Arrangement B

Fig. 2. The common grouping method

The Fig.3 illustrates two geology structures, which are composed of two different geologic material and often encountered in the tunnel construction, will impose the bending moment on the cutter-head from different directions and bring boring direction shift from planed. A comparison of the adaptive ability between the fixed and the flexible grouping method was made considering two working conditions. In the following content, some numerical calculation will be conducted under the different grouping plan. The thrust force transmission effect was analyzed based on calculation results.

When the shield excavates in the geologic structure shown in Fig.3 (a), the bending moment at X direction will appear. In this case, the bending moment in the Y axis approximately equal to zero. The geology of type (a) in the Fig.3 will only result in the moment around the X axis. If the grouping method of Fig.2 is employed, the moment can be balanced via only change the pressure of cylinders in area B and C. But the type (b) of Fig.3 will result bending moments around the X and Y axis at the same time, to balance this kind of bending moment, the cylinder's pressure in all areas has to be adjusted. From the derived equation of the thrust system force transmission, the relationship between the end output force and the input force of each limb is quite complex. The complexity will cause the inconvenience in control system and the boring accuracy declination.



(a) The interface is horizontal (b) The interface is inclined

Fig. 3. Two kinds of geology composition of the cutting surface

As geology conditions in each sector of the planed excavation route were probed before the start of the construction of the tunnel. And variation of the geology condition does not change in high frequency; usually it changes once after several kilometers. So, it's applicable to change the grouping arrangement when the boring route of TBM enters the geology status from a different one.

In the numerical calculation, constraint force F_z and bending moments M_x and M_y were imposed on the cutterhead to simulate the work load under the geology status of type (b) in Fig.3. First, the force transmission equation of grouping arrangement type A in Fig.2 was utilized. Forces of each grouping area that balance the constraint force and moments were got. In Fig.4 and Fig.5, the varying trend of F_a, F_b, F_c and F_d in each grouping area was illustrated with the ascendant of M_x,

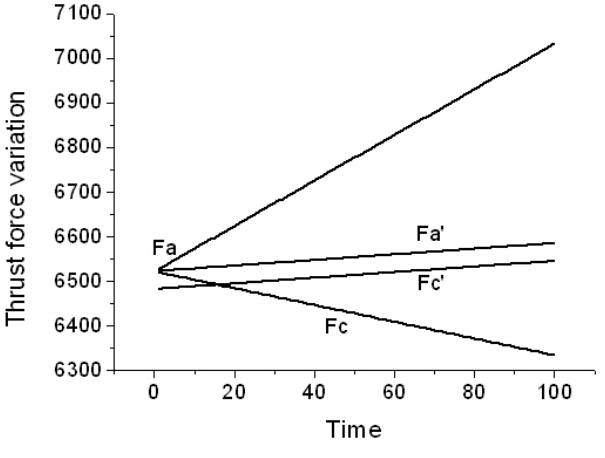


Fig. 4. The forces variation in each area of arrangement A and B

M_y . The conclusion can be draw that each area of force was affected by the bending moment in such kind of grouping method.

If the arrangement type A is adjusted to type B in Fig.2 by an applicable method that changing the connections between cylinders and valves, another force transmission effect can be got. The same calculation as last one with different parameters was conducted. Calculation results were demonstrated in Fig.4 and Fig.5.

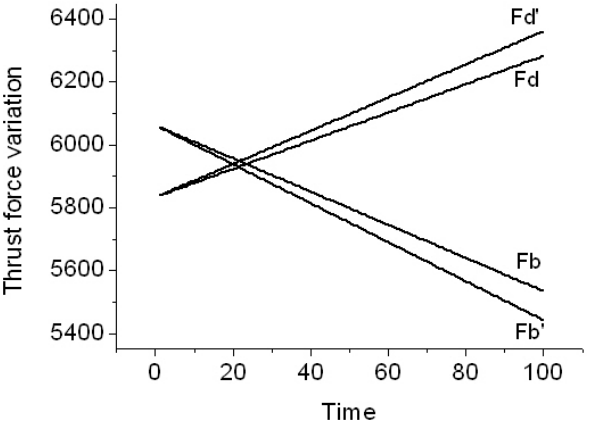


Fig. 5. The force variation of each area of arrangement A and B

As Fig.4 and Fig.5 show when the grouping arrangements are adjusted according to the work load on the cutterhead, only forces in one pair of area need to be adjusted. $F_a^{'}$ and $F_c^{'}$ need not be changed and only adjust $F_b^{'}$ and $F_d^{'}$ were enough to balance the moment on the cutterhead. Comparing the force transmission effect of two cylinder arrangement methods according to the numerical calculation results, proper grouping method can improve the control facility and reduced the accuracy error.

4 Conclusion

In this paper, we focused on the thrust system of the TBM. Screw theory was utilized to analysis the force transmission of thrust cylinders. The relationship of grouping method of hydraulic cylinders and the geology condition was discussed based on above analysis. It was proposed that the arrangement of cylinders schema should be designed from geology conditions of excavating environments for efficient tunnel construction. The numerical calculation was carried according to grouping method under one typical geology composition, and calculation results with one set of working load parameters imposed on the model demonstrated proper grouping method can

improve the construction efficiency. The scope of future works include development of better algorithm that can be used to find proper grouping method according to detailed geology data automatically.

Acknowledgement

The authors are grateful for supports from the National Fundamental Research Program of China (973) (2007CB714003) and the Ph.D. Programs Foundation of Ministry of Education of China (New Teacher) (200802481113).

References

1. Boshkov, S.H., Wright, F.D.: Basic and parametric criteria in the selection, design and development of underground mining systems. In: SME Mining Engineering Handbook. SME-AIME, New York (1973)
2. Morrison, R.G.K.: AW Philosophy of Ground Control. McGill University, Montreal (1976)
3. Laubscher, D.H.: Selection of Mass Underground Mining Methods. In: Design and Operation of Caving and Sublevel Stopping Mines. SME-AIME, New York (1981)
4. Nicholas, D.E.: Method Selection – A Numerical Approach. In: Design and Operation of Caving and Sublevel Stopping Mines. SME-AIME, New York (1981)
5. Huang, Z., Kong, L.F., Fang, Y.F.: Mechanism Theory and control of Parallel Robots. Machinery Industry Press, Beijing (1997)
6. Xiong, Y.L., Affi, Z., Fayet, M.: Robotics. Machinery Industry Press, Beijing (1993)
7. Huang, T., Wang, J.S., Whitehouse, D.J.: Closed form solution to the position workspace of Stewart parallel manipulators. J. Science in China, Series E: Technological Science 41, 394–403 (1998)
8. Merlet, J.P.: Jacobian, Manipulability, condition number, and accuracy of parallel robots. J. ASME Journal of Mechanical Design 128(1), 199–206 (2006)

Analysis for Dynamic Load Behavior of Shield Thrust System Considering Variable Boundary Constraints

Kaizhi Zhang, Haidong Yu, Zhongpo Liu, and Xinmin Lai

State Key Laboratory of Mechanical System and Vibration,
School of Mechanical Engineering, Shanghai Jiao Tong University,
Shanghai 200240, China
aaron@sjtu.edu.cn

Abstract. In this paper, the load transmission model of the shield thrust system is established taking into account the variable boundary constraints between the shield skin and the surrounding strata, which is obtained based on the finite element methods. The results show that the resisting moment on the shield skin from the geologic layer has a nonlinear relationship with the bending moment loads on the cutterhead when the material behavior of the strata is modeled by an elasto-plastic Mohr-Coulomb model. The dynamic load behavior of the shield thrust system is significantly influenced by the variable shield-strata boundary constraints, which may cause incorrect load predictions of the hydraulic thrust system and consequent snake-like motions of the shield machine.

Keywords: Shield machine, load transmission, variable boundary constraint, dynamic load behavior, finite element analysis.

1 Introduction

Shield tunneling method has found widespread applications in constructing tunnels for its high efficiency, safety and environmental protection in the latest decades. With the development of computer-aided automatic control systems, the operation complexity of the shield machine is effectively reduced. However, these systems are still largely dependent on the experiences of the operators and do not have a precise theoretical background [1]. Especially when the shield machine excavates in the composite strata, huge bending moment loads will be generated due to the severe uneven load at the excavation face. It is necessary to predict the load distribution of the shield thrust system and adjust the hydraulic propulsion forces in time. Nevertheless, current predictions and subsequent operations often lead to the machine direction's departure from or passage through the planned alignment all the time, that is, the "snake-like" motion. It may cause difficulties in segment lining and inner forces around the tunnel.

For the need of ground displacement control, much attention has been drawn on the interaction between the shield skin and the strata around. There are three basic models for the material behavior of the strata in finite element methods: linear elastic model by treating soil as a spring with a constant stiffness [2], Cam-Clay plastic model [3, 4] and Mohr-Coulomb elasto-plastic model [5, 6, 7], among which the

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 307–316, 2009.

elastic-perfectly-plastic model is most widely used for its preferable reliability and accuracy. However, this important factor is often improperly neglected for the prediction of thrust distribution. In this paper, the load transmission model of the shield thrust system is established and subsequently in virtue of finite element methods with the strata modeled by Mohr-Coulomb model, the variable boundary constraints between the shield skin and the surrounding strata are obtained. Based on the load transmission model proposed, the dynamic load behavior of the shield thrust system considering the variable boundary constraints is investigated and compared.

2 Load on the Shield Machine

Accurate load analysis is the precondition of proper thrust distribution and from an engineering point of view during excavation where the loads acts on are the cutterhead and the rear long shield skin.

The coordinate system is selected so that the x axis is right at the advance direction of the shield machine, the z axis is vertically upwards, and the x and y axes are on a horizontal plane. The origin of the coordinate system is selected as the center of the section point o on which the jack thrust acts. As shown in Fig.1, there are five loads acting on the shield machine, that is, load due to the self-weight of the cutterhead f_1 and the shield body f_2, load due to the jack thrust f_3, load acting at the face of cutterhead f_4 and load acting on the shield skin f_5, respectively. As the torque loads have no connection with the thrust distribution, they are ignored during the analysis process. Actually the torque loads are generated while the cutters mounted on the cutterhead scratch the excavation face and the driving motors will supply the torque needed.

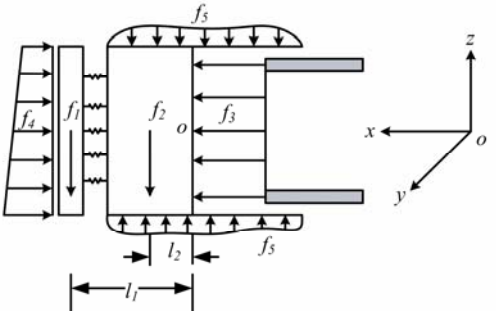


Fig. 1. Schematic of load on the shield machine

Then the load due to the self-weight of cutterhead f_1 and the shield skin f_2 could be simply obtained from

$$F_1 = \begin{bmatrix} 0 \\ 0 \\ -m_1 g \end{bmatrix}, F_2 = \begin{bmatrix} 0 \\ 0 \\ -m_2 g \end{bmatrix}, M_1 = \begin{bmatrix} 0 \\ m_1 g l_1 \\ 0 \end{bmatrix}, M_2 = \begin{bmatrix} 0 \\ m_2 g l_2 \\ 0 \end{bmatrix} \quad (1)$$

here m_1 and m_2 denote the self-weight of the cutterhead and the shield skin respectively, g is the gravity acceleration, l_1 and l_2 refer to the distances from the center of gravity of cutterhead and the shield skin to point o, respectively.

The jack thrust acts in the x axis direction where the jacks are mounted. The frictions on the interface between the jacking plates and the segments are not taken into account. Therefore, the load due to the jack thrust f_3 could be defined as

$$F_3 = \begin{bmatrix} F_{hx} \\ 0 \\ 0 \end{bmatrix}, M_3 = \begin{bmatrix} 0 \\ M_{hy} \\ M_{hz} \end{bmatrix} \tag{2}$$

where F_{hx} denotes the total hydraulic jack force of the thrust system and the force in the y and the z axes are not considered. M_{hy} and M_{hz} represent for the whole bending moments around the y and z axes respectively and the torque around the x axis is ignored during the analysis process.

The load on the face of cutterhead f_4 is determined by the mechanical properties of the geologic conditions at the excavation face and the cutting principle of various kinds of cutters. Since the penetration ability in various geologic structures differs from each other greatly, under the composite strata the uneven advance resistance will frequently appear on the cutterhead. The resulting bending moment loads could severely affect the advance direction which has to be rectified through adjustment of thrust distribution of the shield propulsion system. The load on the face of cutterhead f_4 could be written as

$$F_4 = \begin{bmatrix} -F_{rx} \\ 0 \\ 0 \end{bmatrix}, M_4 = \begin{bmatrix} 0 \\ M_{ry} \\ M_{rz} \end{bmatrix} \tag{3}$$

here F_{rx} denotes the total advance resistance and it acts at the opposite of the advance direction, M_{ry} and M_{rz} represent for the whole bending moments loads on the cutterhead around the y and z axes respectively with the torque around the x axis still ignored.

The load acting on the shield periphery is applied by the strata surrounding the shield skin. Since the shield skin is manufactured long enough to protect the apparatus and operators inside, it inclines to deviate away from the advance direction when suffering bending moment loads. The subsequent resisting moments resulted by the deformation of the strata would counteract a part of the loads on the cutterhead. Then the force acting on the shield periphery f_5 can be represented by

$$F_5 = \begin{bmatrix} F_{sx} \\ 0 \\ F_{sz} \end{bmatrix}, M_5 = \begin{bmatrix} 0 \\ -M_{sy} \\ -M_{sz} \end{bmatrix} \tag{4}$$

where F_{sx} and F_{sz} denote the friction advance resistance and supporting force respectively, M_{sy} and M_{sz} represent for the whole resisting moments around the y and z axes from the surrounding strata respectively. Since the resisting moments serve as an obstacle, they act at the negative x and z axes direction, respectively. It is to be noted

that generally the rotation of shield body is prohibited considering of safety, the static frictional torque is ignored here.

Since the advance speed of the shield machine is generally very slow, so the excavation process could be regrads as a quasi-static one, with the equibrillun equation presented by

$$\begin{aligned} &\sum_{i=1}^{5} F_i = 0 \\ &\sum_{i=1}^{5} M_i = 0 \end{aligned} \qquad (i = 1, 2, \cdots, 5) \tag{5}$$

From Eq. (5) the desired forces and bending moments of the thrust system could be solved by

$$\begin{aligned} F_{hx} &= F_{rx} + F_{sx} \\ M_{hy} &= -(m_1 g l_1 + m_2 g l_2) - M_{ry} + M_{sy} \\ M_{hz} &= -M_{rz} + M_{sz} \end{aligned} \tag{6}$$

It could be found that: a) due to the self-weight of the shield machine the thrust hydraulic system still had to supply resultant moments to prevent the machine from sinking into the soft ground even there were not bending moment loads on the cutterhead; b) the desired forces and bending moments of the thrust system are determined by the loads on the cutterhead and the reaction moments of the surrounding strata.

3 Thrust Distribution with Variable Boundary Constraints

3.1 Variable Boundary Constraints of the Strata

In order to investigate the boundary constraints between the shield skin and the surrounding strata, the finite element model of the shield machine taking into account various geologic conditions is established. The elastic-perfectly-plastic Mohr-Coloumb model widely used for ground and rock is employed to model the material behavior of the strata. The Mohr-Coloumb model is usually used for the materials with the classical Mohr-Coloumb yield criterion and allows the material to harden and/or soften isotropically. A smooth flow potential that has a hyperbolic shape in the meridional stress plane and a piecewise elliptic shape in the deviatoric stress plane is used with the linear elastic material model. It can be used for design applications in the geotechnical engineering area to simulate material response under essentially monotonic load[8]. In order to calculate the force acting on the shield periphery, the shield skin is assumed as a rigid body. The relationship between the bending moment loads on the cutterhead and the resisting moments resulted by the deformation of the strata is obtained via the commercial finite element software ABAQUS.

There are five kinds of geologic conditions with different mechanical properties considered in the simulation model as shown in Table 1. And the geometric parameters of the shield machine studied in this context are shown in Table 2.

Table 1. Mechanical properties of various geologic conditions

Geologic conditions	Specific weight	Young's modulus	Poisson's ratio	Friction angle	Cohesion
Name	γ/kNm^{-3}	E/MPa	μ	φ/°	c/kPa
rubble fill	19.7	15	0.35	20.3	27.14
sandy layer	19.0	33	0.28	35.0	15.0
soft clay	19.4	40	0.27	29.0	31.43
granite 1	24.0	60	0.25	37.0	498.6
granite 2	25.0	180	0.23	40.0	1059.7

Table 2. The geometric parameters of the shield machine

Item	Components	Value
Cutterhead	Diameter	6,285mm
	Width	450mm
	Mass m_1	105,183kg
	Center of gravity from the origin l_1	1,937mm
Shield skin	Diameter	6,280mm
	Mass m_2	286,533kg
	Length	6,300mm
	Center of gravity from the origin l_2	150mm
Thrust system	Numbers of jacks	16
	Groups	4
	Upper group	1
	Left group	4
	Right group	4
	Bottom group	7
	Cross-section area S	0.057m^2
	Center of jack from the origin r	2,797mm
	Angle between each jack	22.5°

3.2 Thrust Distribution of the Hydraulic System

It is by means of a large amount of hydraulic cylinders supplying thrust forces that the shield machine could overcome the advance resistance and go forward. Moreover, these cylinders are separated into several groups for the need of convenient control. For each group, the pressures in these hydraulic cylinders are uniform. The thrust moments will be generated to overcome the bending moment loads on the cuttrehead. As the four groups may be adjusted easily for the operators, it is widely employed in the division of hydraulic cylinders of the shield machine.

A typical application of four-group division is shown in Fig.2. There are sixteen hydraulic cylinders to supply sufficient jack thrust. And they are divided into four groups and the numbers are one, four, four and seven jacks in the upper, left, right, and the bottom group (viewed from the shield tail), respectively.

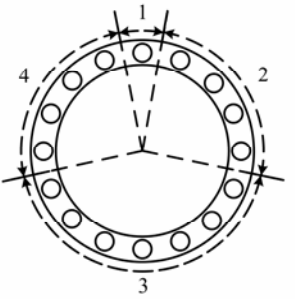


Fig. 2. Four-group division of the thrust system

Here under the coordinate system proposed in Fig.1, the four groups of hydraulic thrust cylinders are defined in the clockwise direction as group 1, 2, 3 and 4, respectively. Hydraulic pressures in the *j*th group is denoted by p_{hj} and the corresponding thrust supplied by each jack in the group could be obtained from

$$f_{hj} = p_{hj} \cdot S \tag{7}$$

where S denotes the cross-section area of each jack. If the hydraulic cylinders are mounted uniformly on the periphery, the value and direction of the total thrust of each jack group could be written as

$$\begin{aligned}
&F_{h1} = \begin{bmatrix} f_{h1} \\ 0 \\ 0 \end{bmatrix}, F_{h2} = \begin{bmatrix} 4f_{h2} \\ 0 \\ 0 \end{bmatrix}, F_{h3} = \begin{bmatrix} 4f_{h3} \\ 0 \\ 0 \end{bmatrix}, F_{h4} = \begin{bmatrix} 7f_{h4} \\ 0 \\ 0 \end{bmatrix} \\
&r_{h1} = \begin{bmatrix} 0 \\ 0 \\ r \end{bmatrix}, r_{h2} = \begin{bmatrix} 0 \\ -r(1+\cos\theta+\cos 2\theta+\cos 3\theta)/4 \\ r(\sin\theta+\sin 2\theta+\sin 3\theta)/4 \end{bmatrix}, \\
&r_{h3} = \begin{bmatrix} 0 \\ 0 \\ -r(1+2\sin\theta+2\sin 2\theta+2\sin 3\theta)/7 \end{bmatrix}, r_{h4} = \begin{bmatrix} 0 \\ r(1+\cos\theta+\cos 2\theta+\cos 3\theta)/4 \\ r(\sin\theta+\sin 2\theta+\sin 3\theta)/4 \end{bmatrix}
\end{aligned} \tag{8}$$

here F_{hj} and r_{hj} denote the value and the direction of the total thrust by the *j*th group, respectively. θ is the angle between each jack and r is the distance from the center of jack thrust to the origin.

In fact the desired thrusts and moments of the shield thrust system are F_3 and M_3 mentioned in the previous section, so it could be obtained that

$$\sum_{j=1}^{4} F_{hj} = F_3 = \begin{bmatrix} F_{rx} + F_{sx} \\ 0 \\ 0 \end{bmatrix}$$
$$\sum_{j=1}^{4} M_{hj} = M_3 = \begin{bmatrix} 0 \\ -(m_1 g l_1 + m_2 g l_2) - M_{ry} + M_{sy} \\ -M_{rz} + M_{sz} \end{bmatrix} \tag{9}$$

From Eq. 9, there are four variables which need to be solved, namely f_j (j=1, 2, 3, 4). If there are bending moment loads generated on the cutterhead, an additional equation should be introduced for the number of variables is greater than that of equations.

Actually during excavation in situ, there are certain control strategies to limit the change of hydraulic pressures in the thrust hydraulic system. Here we assume that the aim is the minimal change of the hydraulic pressures in four groups of cylinders, and then the problem could be regarded as a common optimal problem. On the other hand, when the operators have to adjust the hydraulic thrusts, the less adjustments the better the control effect will be. Then the optimal object could be represented as

$$\min[f(\mathbf{X})] = \min[(f_{h1} - c_{h1})^2 + (f_{h2} - c_{h2})^2 + (f_{h3} - c_{h3})^2 + (f_{h4} - c_{h4})^2] \tag{10}$$

where c_{h1}, c_{h2}, c_{h3} and c_{h4} denote the setting values of hydraulic thrusts in four groups of cylinders respectively and X is defined by

$$\mathbf{X} = [f_{h1}, f_{h2}, f_{h3}, f_{h4}]^T \tag{11}$$

If the Lagrange multipliers are introduced, then by combining Eq. (9-11) the value of hydraulic pressures of four groups of cylinders could be obtained.

4 Simulation Results

According to the process mentioned in the previous section, the variable constraints of the strata could be got via the finite element software ABAQUS. The thrust distribution with the variable constraints could be obtained by solving the simultaneous equations system.

4.1 Variable Constraints of the Strata

The comparisons of the resisting moments resulted by the various geologic conditions are shown in Fig.3. The results show that the constraints of the strata around the shield skin have a close correlation with the mechanical properties of the surrounding strata. When the stiffness of the strata decreases, the deviation from the design alignment of the shield machine will increase. Therefore, it is necessary that the advance direction should be controlled in real-time while the shield machine excavates in the soft clay strata. The correlations between the resisting moments from the deformation of the strata and the bending moment loads on the cutterhead could be fitted by squadratic curves as

$$M_{sy} = g(M_{ry}) = a_2 M_{ry}^2 + a_1 M_{ry} + a_0 \quad (a_2 > 0, 0 < a_1 < 1, a_0 > 0) \tag{12}$$

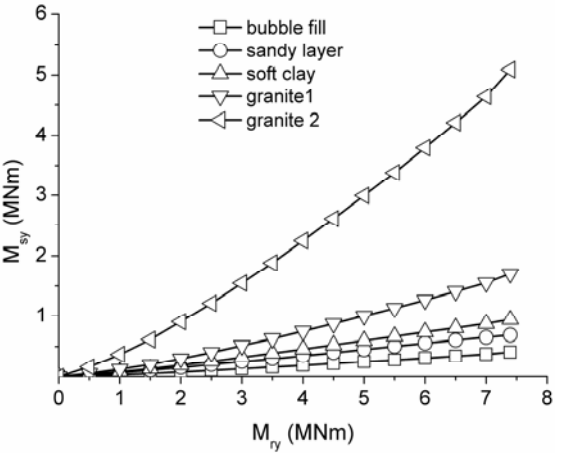


Fig. 3. Comparisons of the bending moments from the sratum

4.2 Thrust Distribution of the Hydraulic System

The thrust distribution with the variable constraints could be obtained by solving the simultaneous equations system if Eq. (9-12). It is to be noted that because the gravitational moments and the frictional force from the strata have little relationships with the dynamic load behavior of the thrust system, them are omitted during the calculation process. Taking the bottom group for example and the important moment loads M_{ry}, the load predictions are compared.

a) If the constraint of the strata is not considered, then we could get

$$\begin{bmatrix} f_{h1} \\ f_{h2} \\ f_{h3} \\ f_{h4} \end{bmatrix} = \Phi_r \begin{bmatrix} F_{rx} \\ M_{ry} \\ M_{rz} \end{bmatrix} + \Phi_c \begin{bmatrix} c_{h1} \\ c_{h2} \\ c_{h3} \\ c_{h4} \end{bmatrix} \tag{13}$$

where

$$\Phi_r = \begin{bmatrix} 0.01539 & -0.00912 & 0 \\ 0.05876 & -0.02847 & -0.05821 \\ 0.0735 & 0.03384 & 0 \\ 0.05876 & -0.02847 & 0.05821 \end{bmatrix}, \Phi_c = \begin{bmatrix} 0.9586 & -0.1399 & 0.02295 & -0.1399 \\ -0.1399 & 0.02042 & -0.003349 & 0.02042 \\ 0.02295 & -0.003349 & 0.0005492 & -0.003349 \\ -0.1399 & 0.02042 & -0.003349 & 0.02042 \end{bmatrix} \tag{14}$$

From Eq. (13-14) we could get the conclusion that when the constraint of the strata is not considered, the desired hydraulic thrusts increase along with the bending moment loads M_{ry} cause

$$\frac{\partial f_{h3}}{\partial M_{ry}} = \Phi_r(3,2) = k_{y3} = 0.03384 > 0 \tag{15}$$

b) If the constraint of the strata is considered in the analysis process, then the Eq.(15) would change to

$$\frac{\partial f_{h3}}{\partial M_{ry}} = k_{y3}\left[\frac{\partial (M_{ry} - M_{sy})}{\partial M_{ry}}\right] = k_{y3}(-2a_2 \cdot M_{ry} + 1 - a_1) \tag{16}$$

That means the tendacy of thrusts are no longer along with the bending moment loads on the cutterhead. While the bending moment loads are smaller than the critical value M_{ryc} defined as

$$M_{ryc} = \frac{1 - a_1}{2a_2} \tag{17}$$

the desired thrusts will increase to keep with the loads. However, once the loads are greater than M_{ryc}, the desired thrusts of the bottom group will decrease with the bending moment loads. Therefore, the hydraulic pressures of the bottom group should be decreased. What's more, the variable value of M_{ryc} is relevant with the mechanical properties of the geologic conditions.

5 Conclusions

Taking into account of the mechanical properties of various geologic structures, the interaction between the shield skin and the strata has been studied. Based on the analysis and the quasi-static equilibrium equations with a control strategy have been establised.Thrust distribution of the hydraulic system is both investigated considering the variable constraints of the strata or not.

The results show that the variable constraints of the strata have close relations to the mechanical properties of the geologic conditions and the geometric parameters of the shield machine. When the variable constraint is considered, the thrust distribution no longer coincides with the trend of bending moment loads on the cutterhead. While the bending moment loads are smaller than the critical value M_{ryc}, the thrusts in the bottom group will increase with the moment loads around the y axis. However, when the loads are greater than M_{ryc}, the thrusts in the bottom group will decrease. The results can provide the guidance for the proper prediction of the redundant thrust system of the shield machine.

Acknowledgments

The authors are grateful for supports from the National Fundamental Research Program of China (973) (2007CB714003) and the Ph.D. Programs Foundation of Ministry of Education of China (New Teacher) (200802481113).

References

1. Sugimoto, M., Sramoon, A.: Theoretical model of shield behavior during excavation. I: Theory. Journal of geotechnical and geoenvironmental engineering 128, 138–1552 (2002)
2. Finno, R.J., Clough, G.W.: Evaluation of soil response to EPB shield machine. Journal of Geotechnical Engineering 111, 155–173 (1985)

3. Kasper, T., Meschke, G.: A 3D finite element simulation model for TBM tunnelling in soft ground. International journal for numerical and analytical methods in geomechanics 28, 1441–1460 (2004)
4. Lee, K.M., Rowe, R.K.: Finite element modelling of the three-dimensional ground deformations due to tunnelling in soft cohesive soil: Part I - Method of analysis. Computers and Geotechnics 10, 87–109 (1990)
5. Bernat, S., Cambou, B.: Soil-structure interaction in shield tunnelling in soft soil. Computers and Geotechnics 22, 221–242 (1998)
6. Abu-Farsakh, M.Y., Voyiadjis, G.Z.: Computational model for the simulation of the shield tunneling process in cohesive soils. International journal for numerical and analytical methods in geomechanics 23, 23–44 (1999)
7. Sugimoto, M., Sramoon, A., Konishi, S., Sato, Y.: Simulation of shield tunneling behavior along a curved alignment in a multilayered ground. Journal of geotechnical and geoenvironmental engineering 133, 684–694 (2007)
8. ABAQUS, Inc. Abaqus Theory Manual, Version 6.8. Providence. Rhode Island: Dassault Systems (2008)

Evaluation and Optimization Method of High Speed Cutting Parameters Based on Cutting Process Simulation

Zhijun Wu, Daochun Xu, Pingfa Feng, Dingwen Yu, and Chenglong Zhang

Department of Precision Instruments and Mechanology, Tsinghua University,
100084 Beijing, China
wuzhijun@mail.tsinghua.edu.cn, xdc05@mails.tsinghua.edu.cn,
fengpf@mail.tsinghua.edu.cn, yudw@mail.tsinghua.edu.cn,
zcl08thu@gmail.com

Abstract. The choice of reasonable cutting parameters is very important for the application of high speed cutting technology. Cutting parameters optimization is a contradiction optimization problem of multi-objectives including cutting efficiency, tool wear and machining quality etc. In order to simplify the complicated problem, a new dimensionless comprehensive evaluation method of cutting parameters based on cutting multi-dimension characteristics including cutting force, temperature, feed rate and cutting depth is put forward in the paper. Then multi-level weight decomposition method is applied to determine the evaluation factors. The relations between multi-objectives and cutting force and temperature are emphatically analyzed. In combination with cutting process physical FEA (finite element analysis) simulation, the effect laws of cutting parameters to cutting force and temperature is researched. Furthermore cutting parameters are evaluated and optimized through the established comprehensive evaluation method. As shown the results: The results obtained by comprehensive evaluation method are match with the results by plenty of cutting experiments. The comprehensive evaluation method is suitable for actual production application because of its simple, effective and extensive characteristics.

Keywords: Cutting parameters optimization, comprehensive evaluation method, high speed cutting, cutting force, cutting temperature.

1 Introduction

With the wide application of high speed cutting (HSC) equipments in many machining domains, as important technical data of cutting technology, cutting parameters are closely related to the application potential of these equipments and reflected the technical level of technologic programming. How to obtain reasonable cutting parameters is related to multi-dimension technologic factors: movement characteristics including cutting speed, feed rate and cutting depth etc; physical characteristics including cutting force and temperature etc.; technologic objectives characteristics including production efficiency, tool wear, workpiece machining quality and technologic stability etc. So the optimization of cutting parameters belongs to multi-factors and multi-objectives problem.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 317–325, 2009.

About cutting parameters optimization, relevant scholars have done many researches from different cutting characteristics views: physical characteristics including cutting force [1, 2], residual stress [3] and etc; technological objectives including production efficiency [4], tool wear [5] and surface roughness [5,6]. Furthermore, some scholars utilized biology genetic algorithm [7] and particle swarm optimization (PSO) [8] and other mature algorithms to optimize cutting parameters. These algorithms can enlarge the search scope and develop the optimization accuracy. However, these researches are engaged only according to one or two cutting characteristics, and can't reflect the comprehensive effects.

There are complicated inter-relationships between multi-dimensional characteristics related to cutting performances. The change trends of different technologic objectives following with cutting parameters are often opposite. The optimal solution of one objective is often far away with the one of another objective. So it's necessary to introduce comprehensive evaluation method for this kind of multi-factors and multi-objective problem. How to simply model of cutting parameters comprehensive evaluation method through few initial input conditions as possible is the key research objective of this paper. Cutting force and temperature are most basic physical characteristics in machining process. The effect of cutting parameters to machining process can be converted to the effect of cutting force and temperature to technologic objectives. Namely, cutting force and temperature can be considered as initial input conditions of comprehensive evaluation method. Furthermore feed rate and cutting depth can also be considered as initial input conditions because they are closely related with production efficiency. So that it will obviously reduce the number of initial conditions, and further simplify the complicated comprehensive evaluation problem.

2 Comprehensive Evaluation Model of Cutting Parameters

Comprehensive evaluation method is organized by initial input conditions, technologic objectives and evaluation factors. In the paper, the initial input conditions are change degree of cutting force, cutting temperature, feed rate and cutting depth. The model of the cutting parameters comprehensive evaluation method is given by:

$$\begin{aligned} &\sum\nolimits_{Cutting-para.} = \sum_i K_{CFi} K_{Mi} \frac{F}{F_0} + \sum_i K_{CTi} K_{Mi} \frac{T}{T_0} - \sum_i K_{Cfi} \frac{f}{f_0} \times \frac{a_p}{a_{p0}} \\ &s.t \begin{Bmatrix} 0 < F < F_{critical}, \sigma_{F_{critical}} = \sigma_r / n \\ 0 < T < T_{melt} \end{Bmatrix} \end{aligned} \tag{1}$$

where K_{CFi} is factor of cutting force to No. i technologic goal; K_{CTi} factor of cutting temperature to No. i technologic goal; K_{Cfi} factor of feed rate and cutting depth to No. i technologic goal; K_{Mi} factor of No. i technologic goal to comprehensive goal; F, T cutting force and temperature of new cutting parameters; F_0, T_0 cutting force and temperature of original cutting parameters; f, a_p feed rate and cutting depth; f_0, a_{p0} original feed rate and cutting depth, $s.t$ indicates constraint conditions, σ_r is the fatigue strength of the tool material, n is the safety factor, and T_{melt} is the melting point of the workpiece material.

Because of the complexity of machining and the diversity of the technological goals, it is difficult to determine the evaluation factors accurately. Therefore some qualitative and quantitative conclusions in cutting theory are integrated together with the main factors of actual production conditions in order to determine the optimal evaluation factors.

To obtain the value of $K=K_C K_M$ as accurately as possible, a weight decomposition method was applied in the paper, which decomposes the value of K into weight factors of sub-goal versus total goal, and cutting force and temperature versus the sub-goal. The value of K_M is closely related to the machining conditions, requirements and parts structure, such as machining system stiffness variations, finish and rough machining, complicated and simple workpieces, piecework and batch production, etc. Therefore, the value of K_M should be determined according to the actual production conditions. The factors shown in Table 1 are determined for conditions of complicated workpieces, batch production and finish machining. Technological goals such as cutting efficiency, tool life, technological stability, surface roughness, machining precision, tool breakage and workpiece burn are selected, which are closely related to cutting parameters. The annotations of the weight factors are shown as follows (Table 1):

a) Cutting efficiency is a most important goal of cutting technologic programming. But the development of cutting efficiency is often accompanied with machine serious vibration, chip removal difficulties and other technologic problems. Otherwise for a more mature technology, the optimization space of cutting efficiency is limited. So the weight factor of cutting efficiency is considered as 0.2.
b) Tool wear is a most important index of tool cutting performance, which is closely related to machining cost. Its weight factor for the comprehensive goals is selected as 0.25.
c) Technological stability is a most important index of workpiece batch production. It indicates that the machining quality will appear unstable with increasing production numbers. However, only the tool wear condition is changed as the technological programming has been ordered, so tool wear is the main factor in technological stability. The weight factor of technological stability is thus selected as 0.25, which is same as that for tool life.
d) Surface roughness and machining precision are closely related to the machining quality. However, their super error phenomenon does not appear easily in the tolerance range. The oversize possibility increases with increasing cutting force and temperature, so their weight factors are both selected as 0.1.
e) Tool breakage and workpiece burn are limited conditions of machining technology. The possibility of these abnormal problems is not high, so their weight factors are selected as 0.1.

Table 1. Weight factors of technological goals to comprehensive goal (K_M)

Contents\No.	1	2	3	4	5	6
Technologic goals	Cutting efficiency	Tool life	Technologic stability	Surface roughness	Machining precision	Other
Weight factors	0.2[a]	0.25[b]	0.25[c]	0.1[d]	0.1[d]	0.1[e]

The value of K_C is determined according to qualitative and quantitative conclusions in cutting theory. The weight factors of cutting force, temperature, feed rate and depth of cutting versus the technological goals are shown in Table 2. The annotations of the weight factors are shown as follows:

a) There is a linear relationship between cutting efficiency and feed rate and cutting depth. So the weight factor of feed rate and cutting depth is 1. The effect of cutting parameters to other objectives has been converted to the effect of cutting force and temperature to them.
b) Cutting temperature is the main factor in tool life according to the cutting theory ($L=(C/T)^{1/n}$, where L is tool life, C and n are constants relevant to the tool and workpiece material, and T is cutting temperature.). Cutting force is the second most important cause of tool wear, as it increases the friction degree between tool and workpiece or chip. The factor for temperature was thus selected as 0.9, but the factor for cutting force is 0.1.
c) The selection basis for the technological stability is same as that for the tool life.
d) Cutting force directly affects the distortion of the workpiece, the vibration amplitudes of the spindle axis, and the rigid distortion of the tool system. These factors will directly affect the workpiece surface roughness and machining precision. Cutting temperature will indirectly affect the cutting force by tool wear. The factors of cutting force and temperature to surface roughness and machining precision are thus selected as 0.8 and 0.2.
e) For tool breakage, workpiece burn and other abnormal problems, cutting force and temperature will play the same role, and thus the factors are both 0.5.

Table 2. Factors of cutting force and temperature to technological goals (K_C)

Factors\Goals	Cutting efficiency	Tool life	Technologic stability	Surface roughness	Machining precision	Other
Cutting force	0	0.1[b]	0.1[c]	0.8[d]	0.8[d]	0.5[e]
Cutting temp.	0	0.9[b]	0.9[c]	0.2[d]	0.2[d]	0.5[e]
f and a_p	1[a]	0	0	0	0	0

The following formula will be obtained after substituting the factors of table 1 and 2 in the formula 1.

$$\sum_{Cutting\ para.} = 0.26\frac{F}{F_0} + 0.54\frac{T}{T_0} - 0.2\frac{f \times a_p}{f_0 \times a_{p0}} + 0.4 \tag{2}$$

The $\Sigma_{cutting\ para.}$ value for the original cutting parameters is 1. The '0.4' in the formula is set to balance the negative growth character of feed rate and cutting depth to cutting efficiency. The smaller the Σ value, then the better the cutting performance becomes. If the difference of the comprehensive value is small, then the new cutting parameters have inconspicuous differences from the original parameters. Otherwise, the original cutting parameters should be optimized. In the process of actual technologic programming, the programmer can select 2-3 optimized parameters groups, conduct

some cutting performance experiments, and determine the optimal cutting parameters. Customer can also add other input conditions into formula (2) according to actual necessary.

3 Case Study of the Comprehensive Evaluation Method

There are three current methods used to obtain cutting force and temperature data: cutting experiments, an empirical formula and cutting finite element analysis (FEA) simulation. Klamechi introduced the FEA method into the cutting domain for the first time in 1973 [9], and the cutting FEA method has gradually become the most effective method of obtaining cutting force and temperature data over the ensuing period [10]. Especially cutting temperature data is particularly difficult to obtain by experimental methods and is not accurately found by the empirical formula method. Cutting FEA simulation is thus used in the paper in order to obtain experimental data for different cutting parameters.

Table 3. Basic setting of the cutting FEA model

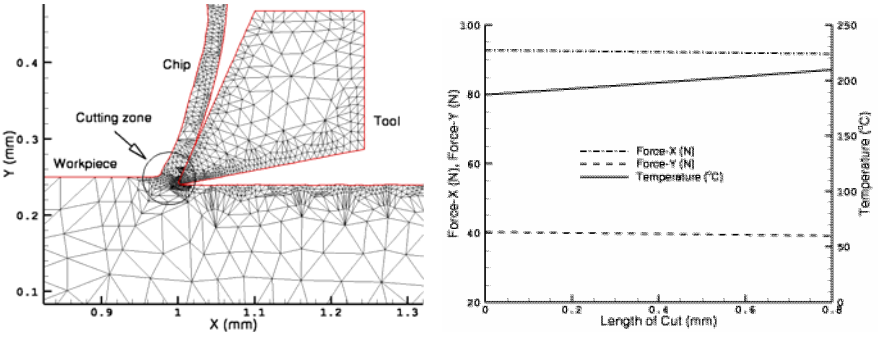
PROJECT	Project / Job Name = JOB1, Project / Job Description = Cutting Para. Opt., Units = SI, Process Type = turning, WorkPiece Type = STANDARD, Tool Type = STANDARD, Tool Material Type = STANDARD.
WORKPIECE	Workpiece length = 1 mm, Workpiece height = 0.25mm, Material = Al6061-T6
TOOL	Tool File = JOB11.twt, Rake angle = 24.0 , Rake length = 0.25mm, Relief angle = 11.0 , Relief length = 0.25mm, Cutting edge radius = 0.01mm, Tool Material = Carbide-General, Coating Layer No. = 0.
PROCESS	Depth of cut = 8.82 mm, Length of cut = 0.8mm, Feed = (0.005mm/z, 0.006 mm/z, 0.007 mm/z, 0.008 mm/z, 0.009 mm/z), Cutting speed = (880m/min, 1005m/min, 1131m/min, 1257m/min, 1382m/min), Initial temperature = 20.0 C, Friction coefficient = Default, Cutting mode = General, Coolant = OFF.
SIMULATION	Simulation Mode = Rapid, Steady State Analysis = 0, Avg. Length Of Cut = 10.0, Chip Breakage = 0, Max. number of nodes = 12000, Max Element Size = 0.1 mm, Min Element Size = 0.02 mm, Fraction of Radius = 0.6, Fraction of Feed = 0.1, Mesh Refine = 2, Mesh Coarse = 6, Output Frame = 30, Number of Threads = 1.

Basic parameters and model settings are shown as Table 3. The FEA model of cutting and its results are shown in Fig. 1. The experimental data are collected from the stable data zone. The final results are the average values of four data located at different lengths of cutting (0.2mm, 0.4mm, 0.6mm, 0.8mm).

Cutting force and temperature curves are shown as Fig.2-5. As seen from the chart: Cutting force will have an upward trend followed with feed per tooth gradually, and cutting force of different cutting speed will have three trend stages: upward, stable and forward. Namely cutting force will decrease in high speed. Cutting temperature

changes less with feed per tooth. Cutting temperature will have an upward trend followed with cutting speed. As the cutting speed rise to a certain level, the change is not obvious.

The comprehensive evaluation values of cutting speed and feed per tooth are shown in Fig.6. As the cutting speed is 880m/min, the feed per tooth pros and cons to choose is 0.006-0.008mm/z. As the cutting speed is 1257m/min, the feed per tooth



(a) (b)

Fig. 1. Cutting FEA model (a) and simulation results (b)

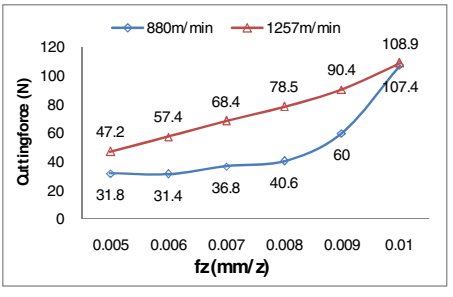


Fig. 2. Variation of cutting force with *fz*

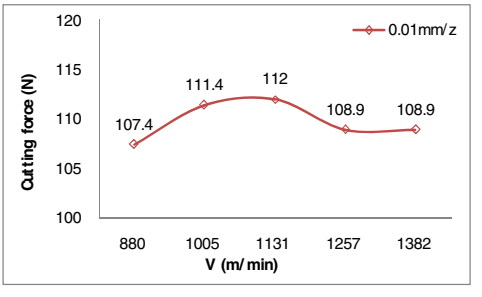


Fig. 3. Variation of cutting force with cutting speed

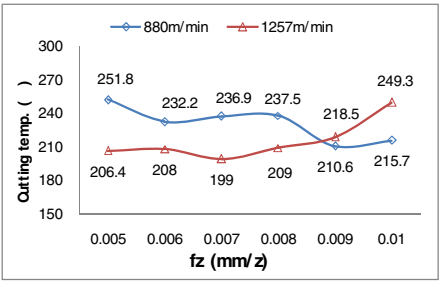


Fig. 4. Variation of cutting temp. with *fz*

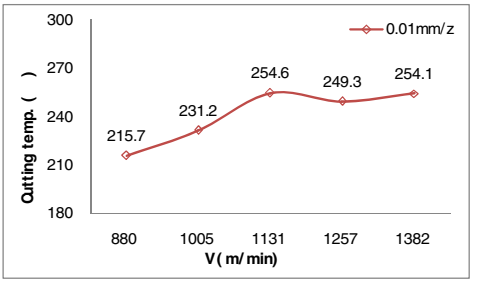


Fig. 5. Variation of cutting temp. with cutting speed

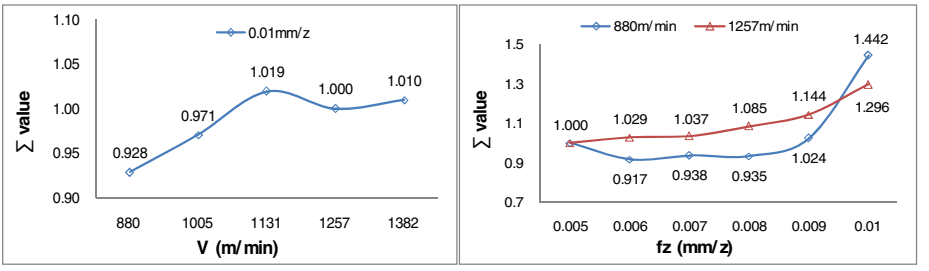


Fig. 6. Comprehensive evaluation value of single factor

pros and cons to choose is 0.005-0.007mm/z. As the feed per tooth is 0.01mm/z, the cutting speed pros and cons to choose is 880m/min. (880m/min, 0.006mm/z) is the better cutting parameter group.

Shown as above results, the optimal feed per tooth in different cutting speed is not same. So it's necessary to optimize the cutting parameters in the condition of combination cutting speed and feed. The comprehensive values are shown as Fig.7. No. 1-5 is the experiment group of (0.005mm/z, 0.006 mm/z, 0.007 mm/z, 0.008 mm/z, 0.009 mm/z) as the cutting velocity is 880m/min. No. 6-10 is the experiment group of (880m/min, 1005m/min, 1131m/min, 1257m/min, 1382m/min) as the feed rate per tooth is 0.01 mm/z. No. 11-15 is the experiment group of (0.005mm/z, 0.006 mm/z, 0.007 mm/z, 0.008 mm/z, 0.009 mm/z) as the cutting velocity is 1257m/min. As seen from the chart: The better combination is No.2-5 and No.11-13. No.2-5 is better than No.11-13. The optimal one in these 15 group parameters is No.5 (880m/min, 0.009mm/z). The result obtained by single factor evaluation isn't same with the one obtained by combination factors. It indicates that coupling effect exists in cutting parameter. The main reason of the effect was the existing of thermal-mechanical coupling.

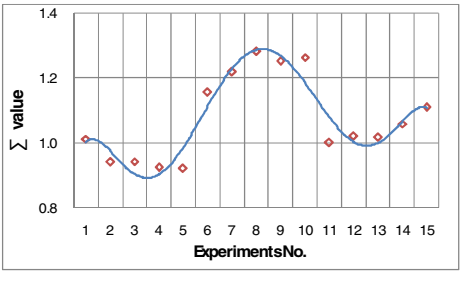


Fig. 7. ∑value of combination *v* and *fz*

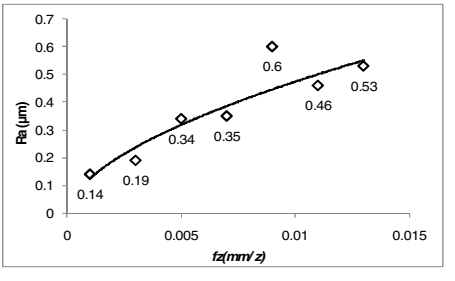


Fig. 8. Variation of surface roughness with *fz*

According to the results of comprehensive evaluation of single factor and combination factors, it can be concluded that the better cutting parameters are (880m/min, 0.006mm/z-0.009mm/z). Then cutting experiments of these better parameters were carried out in order to verify the evaluation results. The variation of surface roughness (*Ra*) with *fz* is shown in Fig.8. The *Ra* value will increase with *fz*. The surface roughness is in the range of precision machining and has little difference

among the *fz* (0.005mm/z-0.009mm/z). Because the *fz* is less than the cutting edge radius, the cutting locates in the range of micro-feed cutting and the main tool wear form is cutting edge wear. Namely the cutting edge radius will become bigger and bigger and the tool will become blunt. The variation of cutting edge radius (*ΔRn*) with *fz* is shown in Fig.9. *ΔRn* will decrease with the increasing of *fz* which denotes that higher *fz* is better parameters. The cutting efficiency results are shown in Fig.10. It also shows that higher *fz* is better parameters. According to the three kinds of experiment results, it can concluded that higher *fz* (among 0.005-0.009mm/z) is better parameters as the cutting speed is 880m/min. This conclusion is match with the comprehensive evaluation results shown as above research.

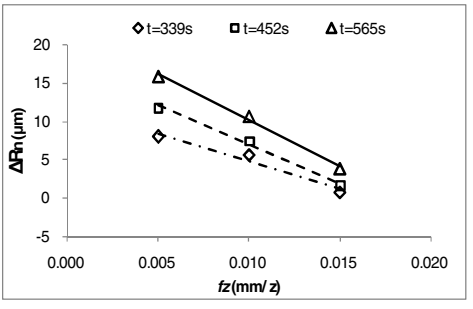


Fig. 9. Variation of cutting edge wear with *fz*

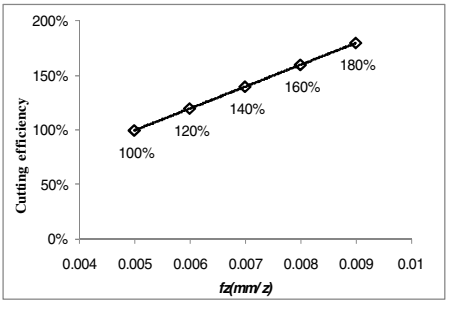


Fig. 10. Variation of cutting efficiency with *fz*

4 Conclusion

Cutting force, cutting temperature, feed rate and cutting depth are considered as initial input conditions. The multi-level weight decomposition method is utilized to determine the comprehensive evaluation factors accurately as possible in order to establish the dimensionless cutting parameters comprehensive evaluation method. Shown as the results:

1) The results obtained by comprehensive evaluation method match with the results by plenty of cutting experiments. The comprehensive evaluation method is suitable for actual production application because of its simple, effective and extensive characteristics.

2) (880m/min, 0.006-0.009mm/z) obtained by comprehensive evaluation is optimal combination parameters as $\phi 80$ saw milling tool machining a plane.

3) Compared with cutting experiment method, the combination method of comprehensive evaluation and cutting FEA simulation can greatly reduce cost of cutting experiments which will develop technologic programming efficiency and raise the overall technologic level of application enterprise.

Acknowledgments. The research was supported by National Key Basic Development Planning Projects (973 Projects) No.2007CB714007, China.

Reference

1. Pan, Y.Z., Ai, X., Tang, Z.T.: Optimization of tool geometry and cutting parameters based on a predictive model of cutting force (in Chinese). China Mechanical Engineering 19, 428–431 (2008)
2. Zhao, H., Li, J.G., Yao, Y.X.: Cutting parameters optimization for constant cutting force in milling. E-engineering & Digital enterprise technology 10, 483–487 (2008)
3. Zong, W.J., Li, D., Cheng, K., Sun, T., Liang, Y.C.: Finite element optimization of diamond tool geometry and cutting-process parameters based on surface residual stresses. Int. J. Adv. Manuf. Technol. 32, 666–674 (2007)
4. Wu, M.P., Zhai, J.J., Liao, W.: Research on NC Machining Parameter Optimization (in Chinese). China Mechanical Engineering 15, 235–237 (2004)
5. Aslan, E., Camuscu, N., Birgoren, B.: Design optimization of cutting parameters when turning hardened AISI 4140 steel (63 HRC) with Al2O3+TiCN mixed ceramic tool. Materials and Design 28, 1618–1622 (2007)
6. Bajic, D., Lela, B., Zivkovic, D.: Modeling of machined surface roughness and optimization of cutting parameters in face milling. Metalurgija 47, 331–334 (2008)
7. Shen, L.G., Han, Z.J., Zhang, K.: Planning cutting parameters based on biology genetic algorithm (in Chinese). China Mechanical Engineering 5, 34–35 (1994)
8. Li, J.G., Yao, Y.X., Gao, D.: Cutting parameters optimization by using particle swarm optimization (PSO). E-engineering & Digital enterprise technology 10, 879–883 (2008)
9. Klamechi, B.E.: Incipient chip formation in metal cutting: a three dimension finite-element analysis. PhD. Dissertation, University of Illinois at Urbana-Champaign, Urbana, IL (1973)
10. Klocke, F., Beck, T., Hoppe, S.: Examples of FEM application in manufacturing technology. J. Mater. Process. Technol. 120, 450–457 (2002)

The Reconfiguration Design of Redundant Thrust of Shields Excavating in Various Geologic Conditions

Haidong Yu, Zhongpo Liu, Kaizhi Zhang, and Xinmin Lai

State Key Laboratory of Mechanical System and Vibration,
School of Mechanical Engineering, Shanghai Jiao Tong University,
Shanghai, China, 200240
hdyu@sjtu.edu.cn, skyliu@sjtu.edu.cn, aaron@sjtu.edu.cn,
xmlai@sjtu.edu.cn

Abstract. The loads on the cutterhead will vary with the mechanical properties and the compositions of geologic structures on the excavation face. It is necessary that the reconfiguration design of the redundant thrust system for various geologic formations is conducted to improve the load compliance ability effectively, which is beneficial to decrease the excessive wear and the other accidents. In this paper, a 4-bar parallel manipulator is established in terms of the control characteristic of the redundant thrust system of the shield machines. The location parameters of legs on the platform and the base are used as the variables to obtain the force transmission matrix between the limbs and the platform. The force transmission behavior is studied when the shields excavate in the alignment direction. Based on the cutting principle, the total bending moment model exerted on the cutterhead is proposed. The behavior of bending moments during excavating in the various geologic formations is studied. The correlation between the bending moments and the configuration is discussed. The results show that the configurations of the thrust manipulator may be selected in terms of the geologic structures. It is useful for the compliance design of the thrust system of shields.

Keywords: Reconfiguration, thrust, bending moment, geologic structures, compliance.

1 Introduction

In recent years, the shield machines have found widespread applications in tunnel constructions and are used for excavating tunnels in almost all types of geologic structures due to their high efficiency, safety and environmental protection. The resistance thrust, bending moments on the cutterhead have close correlation with the geologic formations. In order to comply with the real time variation of the resistance loads, the redundant actuating parallel manipulator usually controlled by the four group hydraulic cylinders is usually used in the shield machines. However, the excessive wear and the other accidents such as the rapid stop frequently appear due to the unreasonable controlling parameters in the actuating systems. Therefore, the reconfiguration design

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 326–335, 2009.

of the thrust system should be considered in terms of the various geologic conditions to reduce the control difficulty of the redundant thrust system.

With the development of the block design in the mechanical system, the reconfiguration designs are extensively used in the machine tools, aerospace crafts, and other heavy mechanical systems. To extend the workspace and the kinematic of the parallel manipulators, much research work on the reconfiguration design in the parallel manipulations were conducted. Yang and Chen [1, 2] proposed a reconfiguration parallel manipulation with three limbs and six degree of freedom (DOF) based on the Podhorodeski's configuration theory. The work space can be changed rapidly in terms of the variation of the location of legs, typical pair of contact and lengths of the limbs. Hamlin [3] proposed a reconfiguration manipulation, Tetrobot, based on the truss structures. Ji [4] established a Stewart platform based on the reconfiguration blocks. The parameters identification was discussed. Yao [5] investigated the 6-SPS parallel machine tool with the Stewart platform block. The identification method on the location and pose of the joints was proposed. Gao [6] proposed a micro manipulation with decoupling parallel mechanism. Zhou [7] proposed a new micro manipulation with 3-PUU and 6-PUS. Nine kinds of kinematic limbs were established and their location correlations were taken into account. These studies can provide the idea and theory on the reconfiguration design of redundant parallel manipulation.

In recent years, the studies on the mechanical design of shield machines have been conducted widely with the development of tunneling construction. These studies mainly focus on the design of cutterheads, the equivalent load exerted on the cutterhead and control determination of the hydraulic systems. In 2008, Liu [8] proposed the reconfiguration design method of the cutters on the cutterhead to adapt the complex geologic formations. The modules of the various groups of the cutterhead are used in his work. The thrust system of shields is consisted of the parallel hydraulic cylinders, which are usually controlled by the four groups of hydraulic systems. Since the configuration of the thrust mechanism is fixed, the compliance for the complex load is heavily dependent on the hydraulic system. It easily causes the rapid blockage of the cutterhead. For example, the blockage frequently happened in the tunnel construction of the Shanggong Mountain because the unexpected geologic conditions [9]. Therefore, it is necessary to change the configuration of the thrust system to adapt the various geologic conditions.

In this paper, based on the configuration characteristic of thrust mechanism, a 4-bar SPS parallel mechanism was simplified for the thrust system of the shields. The force transmission matrix of this mechanism was established by using location parameters of the limbs. The load transmission behavior was studied when the shields excavates in the alignment directions.

2 Load Transmission Matrix

The photo and sketch map of the thrust system of shields is shown in Fig. 1. There are sixteen hydraulic cylinders which are divided into four groups controlled by the same hydraulic pressure in each group. At the origin position, there is no rotation at the two ends of the hydraulic cylinders. Consequently, the coordinates of the limbs at the base and platform are identical if they are described at the base and platform frames,

respectively. The redundant actuating thrust system can be simplified as shown in Fig. 2. Since the gravity of shields and the interaction between the shields and ground at the bottom are taken into account, the number of hydraulic cylinders in the bottom group is larger than that in the upper group. The numbers of hydraulic cylinders at the left and right groups are equal. Assuming the equivalent locations of every group are defined by using the parameters r_i and θ_i shown in Fig. 2, respectively, the coordinates of the i th leg at the base and platform can be written as

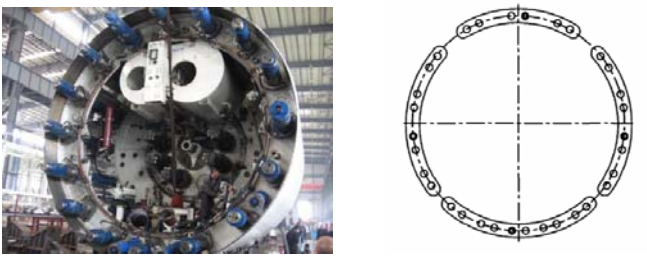


Fig. 1. The distribution of hydraulic cylinders

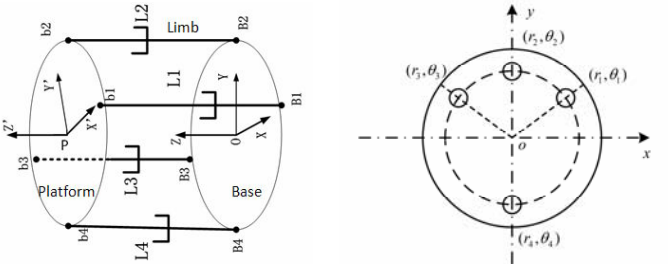


Fig. 2. The simplification model of thrust system

$$\begin{aligned}\mathbf{B}_i &= \begin{bmatrix} B_{ix} & B_{iy} & B_{iz} \end{bmatrix}^T = \begin{bmatrix} r_i\cos\theta_i & r_i\sin\theta_i & 0 \end{bmatrix}^T \\ \mathbf{p}_i &= \begin{bmatrix} p_{ix} & p_{iy} & p_{iz} \end{bmatrix}^T = \begin{bmatrix} r_i\cos\theta_i & r_i\sin\theta_i & 0 \end{bmatrix}^T \end{aligned} \tag{1}$$

Assuming $\mathbf{n}_i$ is the orientation vector of the i th leg. $\mathbf{P} = \begin{bmatrix} x & y & z \end{bmatrix}^T$ is the position vector of platform coordinate system's origin in the based coordinate system. The force $\mathbf{F}$ and $\mathbf{M}$ are applied at the center of the platform illustrated in Fig. 2. The forces in every leg are f_i, $i = 1,2,3,4$. The force equilibrium equation can be written as follows.

$$\begin{cases} f_1\mathbf{n}_1 + f_2\mathbf{n}_2 + f_3\mathbf{n}_3 + f_4\mathbf{n}_4 = \mathbf{F} \\ f_1\mathbf{n}_1 \times \mathbf{r}_1 + f_2\mathbf{n}_2 \times \mathbf{r}_2 + f_3\mathbf{n}_3 \times \mathbf{r}_3 + f_4\mathbf{n}_4 \times \mathbf{r}_4 = \mathbf{M} \end{cases} \tag{2}$$

Namely, the force transmission function can be written as matrix form.

$$[G][f]=\begin{bmatrix}\mathbf{F}\\ \mathbf{M}\end{bmatrix} \tag{3}$$

Here, G is the force transmission matrix and

$$[G]=\begin{bmatrix}\mathbf{n}_1 & \mathbf{n}_2 & \mathbf{n}_3 & \mathbf{n}_4\\ \mathbf{r}_1\times\mathbf{n}_1 & \mathbf{r}_2\times\mathbf{n}_2 & \mathbf{r}_3\times\mathbf{n}_3 & \mathbf{r}_4\times\mathbf{n}_4\end{bmatrix}=[J]^T \tag{4}$$

Assuming the corresponding angles of the coordinate axis of the base and platform are α,β,γ, the direction cosine matrix can be defined as

$$\mathbf{T}=\begin{bmatrix}T_{11} & T_{12} & T_{13}\\ T_{21} & T_{22} & T_{23}\\ T_{31} & T_{32} & T_{33}\end{bmatrix} \tag{5}$$

Therefore, the coordinates of the limbs on the platform can be obtained in the base frame.

$$\mathbf{p}_i'=\mathbf{T}\mathbf{p}_i+\mathbf{P} \tag{6}$$

The orientation of every limb can be written as

$$\mathbf{n}_i=\frac{\mathbf{p}_i'-\mathbf{B}_i}{\left|\mathbf{p}_i'-\mathbf{B}_i\right|}=\frac{\mathbf{T}\mathbf{p}_i+\mathbf{P}-\mathbf{B}_i}{\left|\mathbf{T}\mathbf{p}_i+\mathbf{P}-\mathbf{B}_i\right|} \tag{7}$$

The vector $\mathbf{r}_i$ can be calculated in the base frame.

$$\mathbf{r}_\mathbf{i}'=\mathbf{T}\mathbf{r}_i=\mathbf{T}\mathbf{p}_i \tag{8}$$

So $\mathbf{r}_i'\times\mathbf{n}_i$ can be written as

$$\mathbf{r}_\mathbf{i}'\times\mathbf{n}_i=\begin{bmatrix}0 & -r_{iz}' & r_{iy}'\\ r_{iz}' & 0 & -r_{ix}'\\ -r_{iy}' & r_{ix}' & 0\end{bmatrix}\begin{bmatrix}n_{ix}\\ n_{iy}\\ n_{iz}\end{bmatrix} \tag{9}$$

It is the most common case that the shields excavate in the alignment direction. Since the heterogeneous geologic structures usually appear on the excavation face, the resistance force, bending moments are applied on the cutterhead simultaneously. The cutterhead may be assumed to link with the platform. The orientations of each limb on the platform are parallel with the excavating direction. Namely, the direction cosine matrix is unit. The coordinates of the limbs on the platform can be obtained as

$$\mathbf{p}_i' = \mathbf{T}\mathbf{p}_i + \mathbf{P} = \begin{bmatrix} r_1\cos\theta_1 & r_1\sin\theta_1 & z \end{bmatrix}^T \tag{10}$$

In terms of the Eq. 7-9, the compositions in load transmission matrix when the shields excavate in the linear direction can be obtained as follows.

$$\mathbf{n}_i = \frac{\mathbf{b}_i' - \mathbf{B}_i}{\left|\mathbf{b}_i' - \mathbf{B}_i\right|} = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T \tag{11}$$

$$\mathbf{r}_i' = \mathbf{T}\mathbf{r}_i = \begin{bmatrix} r_i\cos\theta_i & r_i\sin\theta_i & 0 \end{bmatrix}^T \tag{12}$$

$$\mathbf{r}_i' \times \mathbf{n}_i = \begin{bmatrix} 0 & 0 & r_i\sin\theta_i \\ 0 & 0 & -r_i\cos\theta_i \\ -r_i\sin\theta_i & r_i\cos\theta_i & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} r_i\sin\theta_i \\ -r_i\cos\theta_i \\ 0 \end{bmatrix} \tag{13}$$

Therefore, the load transmission matrix between the limbs and loads on the cutterhead when the shields excavate in the alignment direction can be obtained as

$$[G] = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ r_1\sin\theta_1 & r_2\sin\theta_2 & r_3\sin\theta_3 & r_4\sin\theta_4 \\ -r_1\cos\theta_1 & -r_2\cos\theta_2 & -r_3\cos\theta_3 & -r_4\cos\theta_4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{14}$$

For the parallel manipulator shown in Fig. 2, it cannot bear the forces in the x and y directions and the bending moment in the z direction if the platform is parallel with the base. Due to the contact between the shield body and the ground and the larger friction force, the forces in x and y directions will be balanced by the ground. The bending moment around the z direction is mainly undergone by the variable frequency electric motors, which is independent on the thrust system. Consequently, the force transmission matrix may be used directly in the load behavior study. In terms of the design principle of shields, the constrained conditions can be described as follows.

$$\begin{aligned} &r_1 = r_3; \qquad \theta_1 + \theta_3 = \pi \\ &\theta_2 = \pi/2; \ \ \theta_4 = 3\pi/2 \end{aligned} \tag{15}$$

So the load transmission matrix can be simplified as

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ r_1\sin\theta_1 & r_2 & r_1\sin\theta_1 & -r_4 \\ -r_1\cos\theta_1 & 0 & -r_1\cos\theta_1 & 0 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \end{bmatrix} = \begin{bmatrix} F_z \\ M_x \\ M_y \end{bmatrix} \tag{16}$$

3 Bending Moments on the Cutterhead

Two layers of geologic formations containing the hard rock and soft soil are commonly met when the shields excavate in the tunnel construction shown in Fig.3. To satisfy the various geologic conditions, the disc cutter and drag bit are assembled on the cutterhead shown in Fig. 4. There are a great deal of research work on the load behavior for the individual disc cutter and drag bit in recent years. In 1993, Rostami gave the prediction model for the normal force of disc cutter which has high accurate and extensively used in the engineering design [10]. It can be written as

$$F_N = TR\phi C \sqrt[3]{\frac{\sigma_c^2 \sigma_t S}{\sqrt{RT}}} \cos\frac{\phi}{2} \tag{17}$$

here, F_N is the normal force of the disc cutter. T is the width of the cutter tip. R is the radius of the cutter. ϕ is the angle of the contact zone between the rock and disc cutter and $\phi = \cos^{-1}\left(\frac{R-P}{P}\right)$. P is the penetrating depth. C is the dimensionless parameter and approximately equal to 2.12. σ_c and σ_t are uniaxial compressive and tensile strength of rock. S is the span of the adjacent cutters.

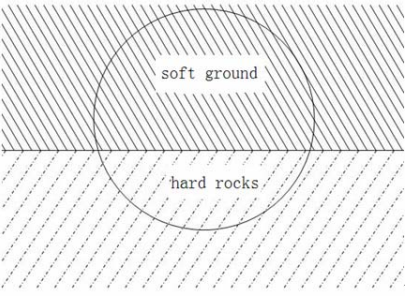


Fig. 3. The various geologic conditions

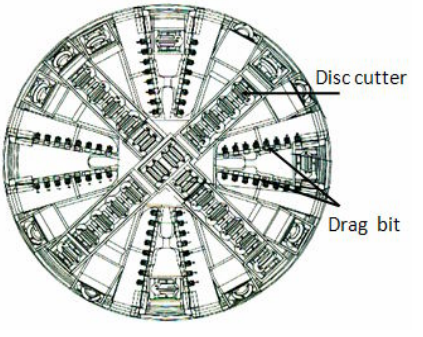


Fig. 4. The typical cutterhead

The cutter force of drag bit is given as follows [11].

$$F_v = N_0 \sin\alpha + \mu_0 N_0 \cos\alpha = \frac{C_1 S_1}{B}(\cos\alpha - \mu_0 \sin\alpha) \tag{18}$$

here, N_0 is the normal force on the cutter face. μ_0 is the friction coefficient between the cutter and soft ground. C_1 is the soil cohesive force of the shear fracture surface. The cutting angle α of the drag bit is defined as $\alpha = \beta + \gamma$. γ is the front angle of

the drag bit, respectively. β is the angle of the cutter tip. S_1 is area of the shear fracture face. B is an engineering parameter. B and S_1 can be defined respectively as follows.

$$B = (\cos\theta - \mu_1 \sin\theta)(\sin\alpha + \mu_0 \cos\alpha) + (\cos\alpha - \mu_0 \sin\alpha)(\sin\theta + \mu_1 \cos\theta) \tag{19}$$

$$S_1 = w \cdot h \cdot \tan\theta \tag{20}$$

Here, μ_1 is the friction coefficient on the shear fracture surface. w is the width of drag bit. h is the penetrating depth. θ is the angle between the shear failure face and cutting faces. In terms of the principle of the plastic mechanics and soil mechanics, $\theta = \frac{1}{2}(90 - \varphi)$, φ is the interior friction angle. The total bending moment when the cutterhead excavates in the various geologic structures can be obtained as follows.

$$\begin{cases} M_x = \sum_0^{m_1} F_{Ni} \cdot l_{ix} + \sum_0^{m_2} F_{vj} \cdot l_{jx} \\ M_y = \sum_0^{m_1} F_{Ni} \cdot l_{iy} + \sum_0^{m_2} F_{vj} \cdot l_{jy} \end{cases} \tag{21}$$

Where l_{ix} and l_{iy} are the distances from the i th disc cutter to the x and y axis, respectively. l_{jx} and l_{jy} are the distances from the j th drag bit to the x and y axis. m_1 and m_2 are the numbers of the disc cutters and drag bits in the hard rocks and soft soils. And they will vary when the cutterhead revolves in the heterogeneous geologic structures.

4 Discussion

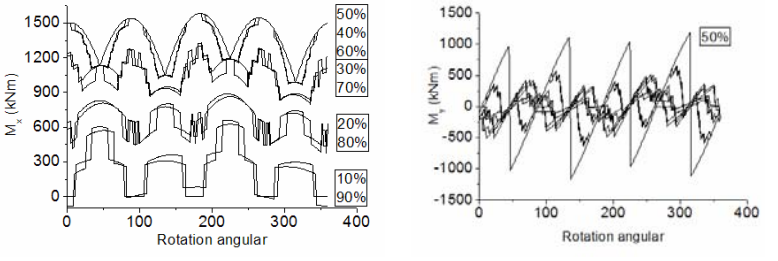
The bending moments on the cutterhead around the x and y axis are illustrated in Fig. 5. The bending moment M_y fluctuates around the axis $M_{y0} = 0$ shown in Fig. 5 (b). The amplitude and frequency of the moments are similar although the fractions of soil grounds are different. Since the mechanical properties and the fractions are different, the bending moment M_x on the cutterhead fluctuates around the axis $M_{xc} \neq 0$ shown in Fig.5(a). The values of the M_{xc} are close correlation with the fractions of the soil grounds. In terms of the load transmission matrix (Eq. 16), If $\theta_1 = \theta_3 = 0$, the bending moments M_x mainly provided by the hydraulic cylinders

in the upper and bottom groups. Since the numbers of hydraulic cylinders of the two groups are different, it is difficult to determine the control parameters of the hydraulic system. However, if $\theta_1 = \theta_3 \neq 0$, the bending moments M_x can be divided into two parts, the constant part M_{xc} and the fluctuation part M_{xf}. Namely

$$M_x = M_{xc} + M_{xf} \tag{22}$$

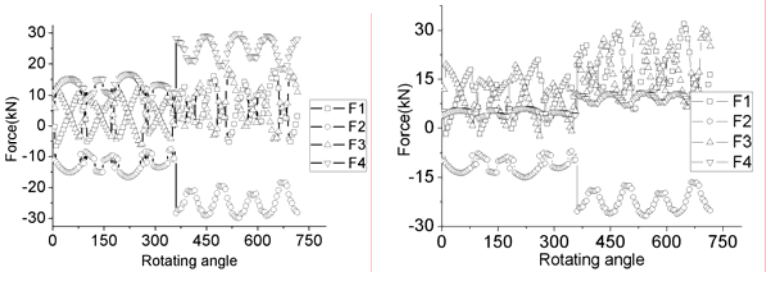
The constant part M_{xc} may be provided by the f_2 and f_4. In terms of the geologic conditions, the fixed pressures can be set. The fluctuation part M_{xf} is mainly provided by the f_1 and f_3. If the interface of the two geologic structures is parallel with the x axis and $M_y \approx 0$, then $f_1 = f_3$. In this case, the f_1 and f_3 will be increased or decreased simultaneously. It will degrade the difficulty of control.

The forces of the hydraulic cylinders in the various groups are shown in the Fig. 6 when the shields excavate in the geologic formations with 20% soil ground on the excavation face to that with 60% soil ground. There are two different configurations.



(a) (b)

Fig. 5. The bending moments on the cutterhead



(a) Case 1 (b) Case 2

Fig. 6. The force fluctuation in various hydraulic cylinders

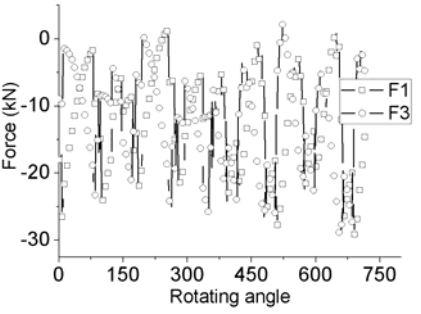


Fig. 7. The force fluctuation of cylinders in the first and third groups

One is $\theta_1 = \theta_3 = 0$ (Case 1)and the other is $\theta_1 = \theta_3 \neq 0$ (Case 2). The bending moment M_x will increase rapidly when the geologic structures vary. For the Case 1, the obvious force fluctuation of the hydraulic cylinders in the second and forth group appears which will increase the controlling difficulty of the hydraulic pressures. In the Case 2, the force fluctuation of the hydraulic cylinders will appear in the first and third group cylinders since they also can provide the bending moment around the x axis. However, if the moment is divided into two parts M_{xc} and M_{xf} and the constant parts is provided by the cylinders in the second and forth groups, the force fluctuations in the first and third groups will be confined in a fixed scope when the forces in the second and forth groups vary with the geologic formations shown in Fig.7. The controlling difficulty will reduce for the second configuration when the shields excavate in the heterogeneous geologic structures.

5 Conclusions

Based on the configuration characteristic of the redundant thrust system of shields, the load transmission matrix between the cutterhead and the hydraulic cylinders were given by using the location parameters of the limbs at the base and platform. The load transmission behaviors were discussed when the shields excavate in the linear directions.

In terms of the cutting principle of the individual disc cutter and the drag bit, the total bending moments at various directions were given. The bending moment behavior on the cutterhead was investigated when the shields excavate in the geologic conditions containing the various fractions of hard rock and soft soil. The results show that there is correlation between the bending moment and the configuration designs. The reconfiguration design for the thrust system of shields can effectively decrease the control difficulty of redundant actuating hydraulic systems. It provides the compliance strategy for the thrust system when the shields excavate in the complex geologic conditions.

Acknowledgment

The authors are grateful for supports from the National Fundamental Research Program of China (973) (2007CB714003) and National Natural Science Foundation in China (50905108).

References

1. Yang, G., Chen, I.M.: Design Consideration and Kinematic Modeling for Modular Reconfigurable Parallel Robots. In: 10th World Congress on The Theory of Machines and Mechanisms, Oulu, Finland (1999)
2. Yang, G., Chen, I.M.: Design and Kinematic Analysis of Modular Reconfigurable parallel Robots. In: Pro. of the IEEE Int. Conf. on Robotics and Automation, Detroit, Michigan, 5, pp. 2501–2507 (1999)
3. Hamlin, G.J., Sanderson, A.C.: TETROBOT, A Modular Approach to parallel Robotics. Kluwer Academic Publishers, Dordrecht (1998)
4. Ji, Z., Li, Z.: Identification of Placement Parameters for Modular Platform Manipulators. Journal of Robotics systems 16(4), 227–236 (1999)
5. Yao, Y., Yao, X., Li, Q., Wu, H.: Identification of Displacement Parameters for Modular 6-SPS Stewart Machine Tool. Journal of East China Shipbuilding Institute 15(1), 38–42 (2001)
6. Gao, F., Jin, Z.L., Liu, X.J.: A 6-DOF Decouping Parallel Micromanipulator, Patent, CN1258589A (2000)
7. Zhou, H.X., Gao, F., Zhang, Z.Y.: Research on Novel Reconfigurable Modular Parallel Micromanipulators. Nanotechnology and Precision Engineering 3, 204–207 (2005)
8. Liu, Z.J.: Cutterhead Design Methods of Rock Tunnel Boring Machine, Dalian University of Technology, Ph.D Thesis (2008)
9. Song, T.T., Xiao, Z.X., Su, H.Y., Lu, X.X.: Engineering geological analyses on 2.22 blockage accident in TBM construction of Shanggongshan tunnel. Chinese Journal of Rock Mechanics and Engineering 23, 4544–4546 (2004)
10. Rostami, J., Ozdemir, L.: A new model for performance prediction of hard rock TBMs. In: RETC Proceedings, pp. 793–809 (1993)
11. Tomisho, J.: Building machine. Kajima Institute Press, Tokyo (2001); Liu Z.J.: Cutterhead Design Methods of Rock Tunnel Boring Machine, Dalian University of Technology, Ph.D Thesis (2008)

Research of an Electric Laser System Based on the Non-diffracting Beam

Hui Chen and Bin Zhao

Department of Instrumentation, School of Mechanical Engineering,
Huazhong University of Science and Technology (HUST), 430074 Wuhan,
Hubei, People's Republic of China

Abstract. A novel electronic laser system is proposed. Based on the principle of the non-diffracting beam, a new method for measuring attitude angle has been developed in this paper, in combination with image processing technology. The system constitution and the measuring principle of the electronic laser system are introduced. As the non-diffracting beam is composed of Bessel fringe ring, the anti-interference ability and the measurement sensitivity is greatly improved by center location of the non-diffracting beam in digital image processing technology, compared with barycenter method. The electronic laser system is applicable to attitude angle measurement of the tunnel boring machine in tunnel construction and other similar occasions.

Keywords: Electronic laser system, non-diffracting beam, attitude angle measurement, image processing.

1 Introduction

With the rapid development of tunnel construction technology, shield method has been widely used in tunnel construction now. An increase of tunnel length leads to an increase in complexity of the tunnel design curve. In order to ensure the holing-through accuracy of tunnel, guiding technologies for the Tunnel Boring Machine (TBM) have developed correspondingly including laser guiding technology and gyroscope guiding technology. The automatic guiding systems commonly used in shield construction are the gyroscope system, ZED guiding system and VMT guiding system. A gyroscope is a device for measuring horizontal azimuth angle, pitch angle and roll angle. But it has a long time-delay, and the azimuth angle of the gyroscope drifts caused by the deviation from the gravity center and the moment of friction in the bearing [1]. Laser signals are detected by a photoelectric sensor in ZED system. The measurement accuracy will decrease with long distance measurement and low light intensity. It also requires a large amount of calculation after moving the total station [2]. In VMT system rotating components lead to a slow measurement speed. The accuracy and resolution of angle measurement is limited by the accuracy of the mechanical structure. So VMT system is difficult to meet the requirement of the higher-precision measurement [3].

Study on the electronic laser system is of great significance to produce the domestic-made guiding system of the TBM. A novel electronic laser system is

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 336–344, 2009.

proposed in this paper. The ranging laser of a total station is used as a light source. The azimuth of incident light is measured by the center positioning of the non-diffracting beam acquired by CCD camera in digital image processing technology. Horizontal azimuth in combination with pitch angle and roll angle measured by the electronic gradienter, three azimuths of the electric laser system are decided.

2 The Measuring Principle of Attitude Angles for the TBM

The main task of the automatic guiding system is to measure the drift while the TBM thrusting. The difference between the direction along which the TBM is moving and the axis of the TBM is called the drift. The position of the prism in the electronic laser system is measured by the total station. The distance between the prism and the center of shield section or shield tail is determinate, so the tunneling error of the TBM is obtained as long as the center positions of the shield section and shield tail are calculated through the attitude angles of the TBM.

The roll angle and pitch angle of the TBM are measured by the biaxial electronic gradienter. If the biaxial electronic gradienter shows that the angle in the direction close to the axis of the TBM is θx and the other angle is θy, then the actual pitch angle of the TBM is θx and the actual roll angle of the TBM is η. They satisfy the following relations.

$$(\cos\theta y)^2 = (\cos\eta)^2 + (\sin\eta \times \sin\theta x)^2 \tag{1}$$

As shown in Figure 1, in the measurement process the horizontal azimuth angle δ of the laser beam in the absolute coordinate system and the angle α between the laser beam and horizontal plane are measured directly by the total station. The angle between the laser beam and the axis of electronic laser system is regarded as the angle θ between the laser beam and the axis of the TBM. The literature [4] shows that the projection angle in the horizontal plane between the laser beam and the axis of the TBM is given by

$$\theta' = \arccos\left(\frac{\cos\theta - \sin\alpha\sin\beta}{\cos\alpha\cos\beta}\right) \tag{2}$$

where β is the angle between the horizontal plane and the axis of the TBM, namely the pitch angle of the TBM θx which is measured by the electronic gradienter. The horizontal azimuth angle of the TBM γ is the angle between the horizontal projection of the TBM axis and the measurement base axis (X-axis). They satisfy the following relations.

$$\gamma = \delta + \theta' \tag{3}$$

From equations (2) and (3), the horizontal azimuth angle γ of the TBM can be calculated by the angle θ between the laser beam and the axis of electronic laser system.

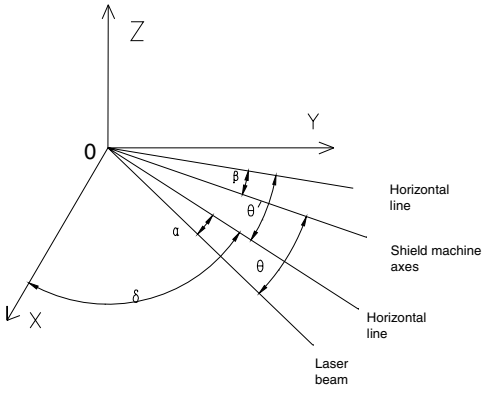


Fig. 1. Azimuth angles of the TBM

In the three attitude angles of the TBM, it is easy to measure the roll angle and the pitch angle by the biaxial electronic gradienter. So it is a key technology to measure the horizontal azimuth angle γ. A measurement equipment for the azimuth angle has been proposed in the literature [5]. The barycenter position of the laser spot acquired by CCD is calculated to get the azimuth angle. This method has the following shortcomings. Firstly the background light is added to calculate the light intensity. So the barycenter position of the laser spot is certainly influenced by the direction and strength of the background light, which results in the measurement error. Secondly the focal spot intensity distribution of the lens is influenced by various aberrations. So the energy of the focal spot is decentralized. When the distribution of the incident light in the aperture isn't uniform, the energy distribution of the focal spot changes subsequently. It also results in the measurement error.

So a new method for measuring attitude angle is proposed in this paper, based on the principle of non-diffracting beam, in combination with image processing technology.

3 System Composition of the Electric Laser System Based on the Non-diffracting Beam

The system composition of the electric laser system based on non-diffracting beam is shown in Figure 2. The tube is placed in the box. A glass plate is fixed in the front of the box, which is used to transmit the laser and protect the internal equipments in the box. The flat-top corner cube prism, the filter and the axicon are installed in the tube and along the axial direction of the tube. The CCD image sensor is located at the end of tube. The electronic gradienter is installed on the outer wall of the tube. The computer connected with the CCD image sensor and the electronic gradienter is used to calculate the center of non-diffracting beam and the attitude angles of the TBM.

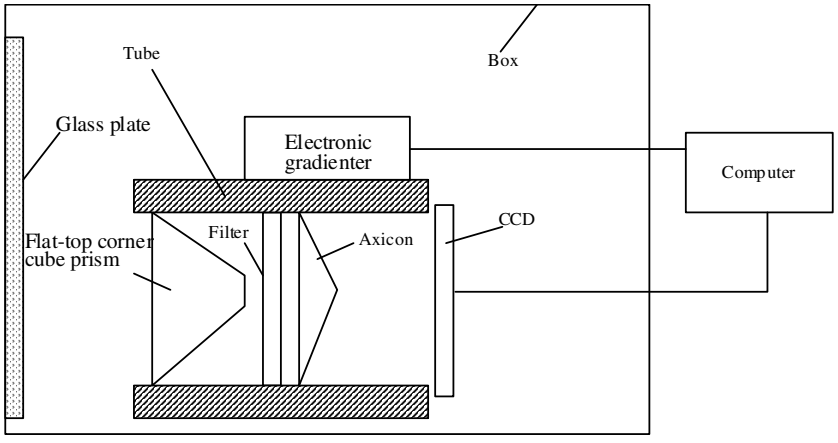


Fig. 2. Block diagram for electronic laser system based on the non-diffracting beam

4 The Measuring Principle for Horizontal Azimuth Angle of the Electric Laser System

Optical principle of the electric laser system based on the non-diffracting beam is shown in Figure 3. The electric laser system is composed of the flat-top corner cube prism, the filter, the axicon and the CCD image sensor. The photosensitive surface of the CCD image sensor is vertical to the axis of the axicon.

The flat-top corner cube prism is a corner cube whose top isn't a vertex, but a small plane. And the plane is parallel to the incident plane of the flat-top corner cube prism. The electric laser system is combined with the tracking prism by the flat-top corner cube prism. So the volume of the electric laser system is greatly reduced. Since the ranging laser of the total station is used as the light source directly, there's no need to install the laser collimator and the operation is simpler. The filter is used to transmit the specific wavelength laser, filter out the background light and lower the intensity of the laser beam incident on CCD.

During the measurement process, the ranging laser of the total station incidents at an arbitrary angle α. In the far distance the ranging laser is similar to the parallel beam. And it becomes a bunch of narrow parallel light after transmitting the small plane of the flat-top corner cube prism. The parallel beam becomes the non-diffracting beam after transmitting the filter and the axicon. The projection of non-diffracting beam on the CCD plane is a series of concentric rings. Figure 3 shows that the center offset of non-diffracting beam h is decided by the incident angle of the laser α. The center offset of non-diffracting beam h can be calculated in digital image processing technology. The incident angle of the laser α can be formulated by

$$\alpha = \arctan(h / L) \tag{4}$$

where L is the distance between the vertex of the axicon and the CCD plane.

So α is the angle between the laser beam and the axis of the electric laser system.

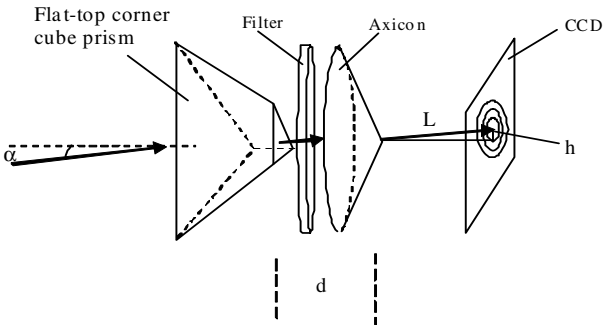


Fig. 3. Optical principle diagram for electronic laser system based on non-diffracting beam

The key technology of the measurement for the angle of the TBM based on the principle of the non-diffracting beam is how to locate the center of the non-diffracting beam. As the spot measured in the experiments contains stray light and noise, a global center detection algorithm is proposed in this paper and the total data of the spot are used in calculation.

5 The Measurement Principle of the Spot Center

In the electric laser system based on the non-diffracting beam, the projection of non-diffracting beam on the CCD plane is not a spot, but a series of concentric rings, as shown in figure 4(a). The transverse intensity distribution of the non-diffracting beam is in accord with the zero order Bessel function J_0. The two-dimensional transverse intensity distribution is shown in figure 4(b).

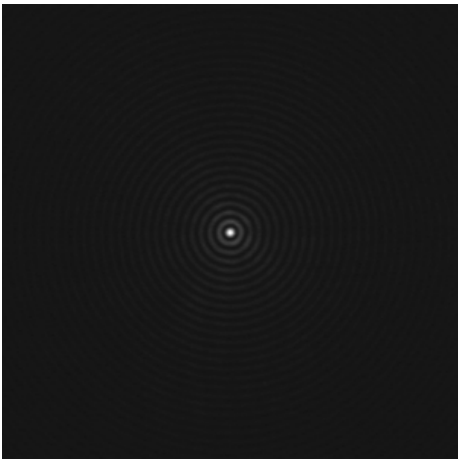


Fig. 4. (a) Experimental image of non-diffracting beam

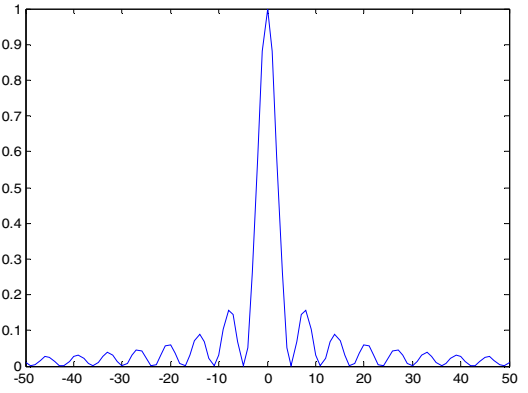


Fig. 4. (b) Bessel beam intensity distribution

Although any portion of non-diffracting beam may be distorted by various interferences, the global center remains changeless visually. Since the circles are concentric, a global center detection algorithm is proposed in this paper. The total data of the pattern are used in calculation. The method features good anti_interference ability, and is not sensitive to partial light intensity non-uniformity and error. The basic idea of the global center detection algorithm is as follow.

① Assume an initial center.

A two-dimension random number (a, b) is assumed to be the initial center. To improve the algorithm convergence rate, the point of the maximum light intensity or an arbitrary point in the ring closest to the center of the beam cross-section may be used as the initial center.

Average the data of each ring, to obtain the revised non-diffracting beam.

Compute the residual sum of squares between the original and revised non-diffracting beam, as indicated in Eq.(5).

$$Q' = \sum_{i\in[1,m],j\in[1,n]} \varepsilon_{i,j}^2 = \sum_{i\in[1,m],j\in[1,n]} [x_{(i,j)} - x'_{(i,j)}]^2 \quad (5)$$

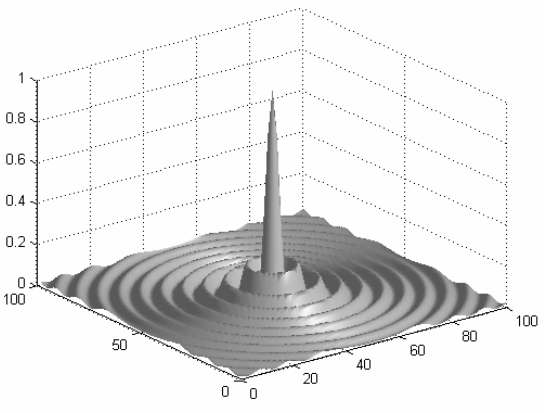
where m and n are the pixel number in one row and one column of non-diffracting beam respectively, x is the original data, and x' is the revised data.

Increase variable Δa and Δb on the basis of the initial center (a,b) to get the new center $(a+\Delta a, b+\Delta b)$.

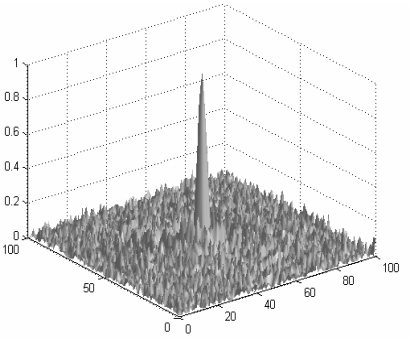
⑤ Find the variable Δa and Δb such that the residual sum of squares is minimized by using least squares optimization algorithm, as indicated in Eq.(6).

$$Q'_{\min} = f(\Delta a, \Delta b). \quad (6)$$

So the ultimate center $(a+\Delta a, b+\Delta b)$ is the global center.



(a)



(b)

Fig. 5. (a) Three-dimensional model of non-diffracting beam (b) Simulated diffraction pattern with Gaussian

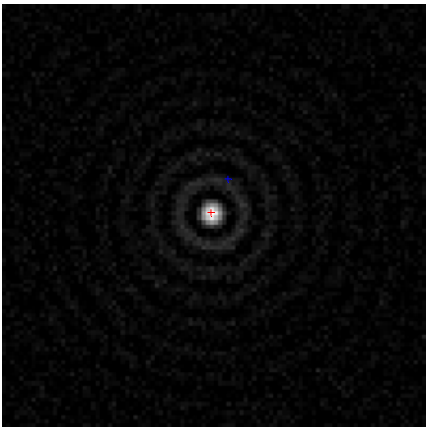


Fig. 6. Global center_ the cross symbol

In order to verify the performance of the global center detection algorithm, we have carried out the simulation experiment. The non-diffracting beam is simulated, as shown in Figure 5(a). The figure is of size 100×100 pixels, and the center is (50, 50). Gaussian noise with zero mean and a variance of 0.001 is added to the simulated non-diffracting beam, as shown in Figure 5(b). In figure 6, the blue cross represents the initial center (42, 54), and the red cross represents the global center (49.9997, 50.0008) which was found by the global center detection algorithm.

Different noises such as Gaussian noise, random noise, uniform noise, salt-and-pepper noise, exponential noise and erlang noise are added to the simulated non-diffracting beam respectively. The centers are shown in Table 1.

Table 1. Performance comparison of center detection algorithms correspond to different noises

Different noises added to simulated non-diffracting beam	Center of global center detection algorithm	Center of least squares circle fitting algorithm
Gaussian noise with zero mean and a variance of 0.001	(49.9997, 50.0008)	(49.8856, 49.9215)
Gaussian noise with zero mean and a variance of 0.01	(49.9965, 50.0066)	(49.8968, 50.1750)
random noise with zero mean and a variance of 0.2	(50.0061, 50.0081)	(49.9960, 50.2222)
uniform noise in the interval (0,0.1)	(50.0000, 50.0334)	(50.1871, 49.8272)
salt-and-pepper noise with noise density of 0.1	(50.0043, 49.9982)	(50.1152, 50.3571)
exponential noise with the mean of 0.03 and a variance of 0.001	(49.992,49.994)	(50.1385, 50.1049)
erlang noise with zero mean and a variance of 0.001	(50, 50.0000)	(49.9134, 50.0005)

The data of the ring closest to the center of the beam cross-section was used in the least squares circle fitting algorithm. Table 1 shows that the presented method is effective and possesses a higher sensitivity and anti-interference ability than the least squares circle fitting algorithm.

As long as the concentric circles are not hidden by the background light and not all rings are affected, the calculation accuracy of the global center will not be affected. A large number of rings of non-diffracting beam are used to calculate the global center in digital image processing method has a higher accuracy, stability of anti-interference and subpixel sensitivity resolution. Thus higher measurement accuracy can be achieved in the same number of CCD pixels.

6 Conclusions

In the electric laser system based on the non-diffracting beam, the ranging laser of the total station becomes the concentric ring-shaped non-diffracting beam after passing through the axicon. The incident angle of the ranging laser can be measured by the center position of the non-diffracting beam acquired by CCD in the image processing method. In combination with the pitch angle and the roll angle measured by the electronic gradienter, the total azimuths of the electric laser system are determined. The electric laser system based on the non-diffracting beam for attitude angle measurement of the TBM in tunnel construction has better anti-interference ability as well as higher measurement resolution.

Acknowledgments

The project is supported by a grant from the Major State Basic Research Development Program of China (973 Program) (No 2007CB714000).

References

1. Yang, H., Liu, Z.-q., Cai, Y.: Detection of Engineering Survey for Shield Tunnel. J. Geospatial Information 6, 36–39 (2005)
2. Yang, Q., Sha, M.-y.: Application of ZED-260 Laser Guiding System to TB880E IBM. J. Modern Tunnelling Technology 6, 29–32 (1998)
3. Liu, D.-l., Huang, Y.-k.: Guided system VMT of shield machine. J. Construction Machine 1, 60–61 (2003)
4. Zou, W.-j., Zhu, G.-l., Wu, X.-b.: Angle measuring system of laser based on array CCD. J. Opto-Electronic Engineering 33(10), 91–95 (2006)
5. Wu, X.-b., Zhu, G.-l., Zou, W.-j.: Research of a PSD-based Electric Laser System. J. Machinery & Electronics 7, 38–40 (2006)

Research on Multi-motor Synchronization Control for Cutter Head of Shield Machine Based on the Ring Coupled Control Strategy*

Jianzhong Sun[1], Ran Liu[1], Yaqin Luo[1], and Wei Sun[2]

[1] Department of Electrical and Electronics Engineering, Dalian University of Technology, Dalian 116023, Liaoning Province, China
[2] School of Mechanical Engineering, Dalian University of Technology, Dalian 116023, Liaoning Province, China
jzsun@dlut.edu.cn, psincos@163.com, sunwei@dlut.edu.cn

Abstract. Multi-motor synchronization control for cutter head of shield machine is studied in this paper. Firstly, a ring coupled control strategy for multi-motor synchronization control based on the idea of parallel control with compensation is proposed, and sliding mode variable structure control algorithm is used in the design of the controllers to ensure high robustness of the control system. Furthermore, synchronization control of a four-motor driving system with the proposed control strategy is simulated with MATLAB, and its effectiveness is verified by comparing its control performance with that of the traditional master-slave strategy. Finally, the ring coupled control strategy is applied to the cutter head driving of a prototype shield machine. The simulation results show that the proposed ring coupled control strategy has good dynamic performance and synchronization performance.

Keywords: Ring coupled control strategy, multi-motor synchronization control, cutter head driving of shield machine.

1 Introduction

Cutter head is the main working part of shield tunneling machine in tunnel construction. With constant increasing of the diameter and system power of shield machine, electrical driving with multi inverter-fed motors for the cutter head finds more and more applications [1]. Since the uneven load distribution on cutter head and the small differences in parameters of driving motors can lead to large differences of instantaneous speed among the driving motors, which may results in shaft broken accident when geological condition changes suddenly, large shield machine demands high synchronization performance.

At present, synchronization control for electrical driving of cutter head mainly includes the following two methods. One is average speed method which makes all the

* This work is supported by NBRPC Grant #2007CB714006.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 345–354, 2009.

motors' speed to be close to the average speed. The other is master-slave method which sets one motor as the main motor, and makes the others track the main motor's speed. The former has poor steady-state performance, and the later has low dynamic synchronization performance [1, 2].

Koren (1980) proposed to use the cross-coupling concept to solve the synchronization problem [3]. Most of the previous works on cross coupled control were found in the control of machine tools [4, 5]. The cross-coupling approach has been also found in robotics. Feng, Koren, and Borenstein (1993) applied the cross-coupling control to differential mobile robots to minimize the differential velocity error of two driving wheels [6]. Sun and Mills (2002) applied the cross-coupling concept to the coordination of two industrial manipulators [7]. Sun (2003) also applied the adaptive coupling control concept into the position synchronization of multi-robot for assembly tasks [8]. Xiao, Zhu and Liaw (2005) studied a generalized synchronization controller for multi axis motion systems by incorporating cross-coupling control [9]. Other synchronization approaches include fuzzy logic coupling control (Moore & Chen, 1995) [10], neuro-controller for synchronization (Lee & Jeon, 1998) [11] and QFT robust controller (Sun, Li and Guo, 2008) [12]. Nevertheless, the synchronization efforts so far are mostly applicable to biaxial systems, and have not yet applied in the control of cutter head driving for shield machines.

In the control of the system with more than two driving axes, typical cross coupling requires vast amount of online computation and inevitably increases the control complexity. Zhang, Shi and Cheng (2007) proposed adjacent cross-coupling approach to reduce the control complexity, by which the number of compensators of an n-axis system decreased from n^2 to $3n$ [13]. For further reduction of the control complexity, this paper puts forward a new ring coupled synchronization control strategy which reduces the number of compensators of an n-axis system to $2n$. The proposed control approach is applied to the control of cutter head driving for shield machines, and in order to deal with the problem resulted from the driving motor parameter differences, sliding mode robust controller is designed.

2 Ring Coupled Control Strategy

2.1 Basic Idea of the Strategy

The ring coupled control strategy is based on the thought of the parallel control with error compensation. As shown in Fig.1, to control n motors to rotate synchronously at a given speed, not only the errors between the given speed and the real speed for each motor (i.e. the tracking error), but also the errors between the speeds of the two adjacent motors (i.e. the synchronization error) must be considered. Define the speed difference between the first motor and the second motor as the synchronization error of the first motor, and couple the first motor with the second motor by compensating the first controller with the first motor's synchronization error. In the same way, couple the ith motor with the (i+1)th motor, and the nth motor with the first motor. At last, a coupling ring is formed through the synchronization errors.

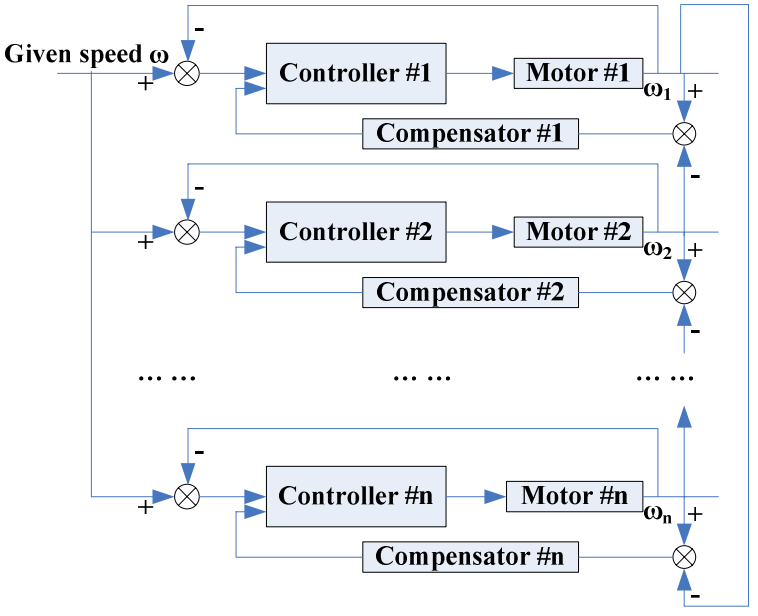


Fig. 1. Schematic diagram of the ring coupled control strategy

2.2 Motor Model

The mathematical model of induction motor in α-β coordinate system is as follows

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}\omega\\ \psi_{\alpha r}\\ i_{\alpha s}\\ i_{\beta s}\end{bmatrix}=\begin{bmatrix}\mu\psi_{\alpha r}i_{\beta s}-\dfrac{B}{J}\omega-\dfrac{1}{J}T_{\mathrm{L}}\\ -\alpha\psi_{\alpha r}+\alpha M i_{\alpha s}\\ -\gamma i_{\alpha s}+\alpha\beta\psi_{\alpha r}+n_{\mathrm{p}}\omega i_{\beta s}+\dfrac{\alpha M\, i_{\beta s}^{2}}{\Psi_{\alpha r}}\\ -\gamma i_{\beta s}+\beta n_{\mathrm{p}}\omega\psi_{\alpha r}-n_{\mathrm{p}}\omega i_{\alpha s}-\dfrac{\alpha M i_{\beta s}i_{\alpha s}}{\psi_{\alpha r}}\end{bmatrix}+\frac{1}{\sigma L_{\mathrm{s}}}\begin{bmatrix}0&0\\0&0\\1&0\\0&1\end{bmatrix}\begin{bmatrix}v_{\alpha s}\\ v_{\beta s}\end{bmatrix} \quad (1)$$

where ω is the mechanical angular velocity of motor; $\psi_{\alpha r}$ is rotor flux in α-axis, $i_{\alpha s}$ and $i_{\beta s}$ are stator currents in α and β-axis respectively, $v_{\alpha s}$ and $v_{\beta s}$ are stator voltages in α and β-axis respectively, L_{s} and L_{r} are inductances of stator an d rotor respectively; R_{s} and R_{r}

are resistances of stator and rotor respectively, M is the mutual inductance between stator and rotor; J is the rotor inertia; n_p is the number of pole pairs; B is the friction coefficient; T_L is the load torque. And μ, α, β, γ and σ are defined as

$$\mu = \frac{n_p M}{JL_r},\ \alpha = \frac{R_r}{L_r},\ \beta = \frac{M}{\sigma L_s L_r},\ \gamma = \frac{M^2 R_r}{\sigma L_s L_r^2} + \frac{R_s}{\sigma L_s},\ \sigma = 1 - \frac{M^2}{L_s L_r}$$

According to Equation (1), motion equation of induction motor is

$$\dot{\omega} = \mu \psi_{\alpha r} i_{\beta s} - \frac{B}{J}\omega - \frac{1}{J}T_L \tag{2}$$

let $a = \frac{B}{J}$, $u = \mu \psi_{\alpha r} i_{\beta s} - \frac{1}{J}T_L$, yields

$$\dot{\omega} + a\omega = u \tag{3}$$

Taking into account the change and uncertainty of the parameters of motors, Equation (3) can be changed as

$$\dot{\omega} + a\omega = u + d$$

where d denotes speed change caused by the parameters uncertainties.

So motion equation of the ith motor is as follows

$$\dot{\omega}_i(t) = -a_i \omega_i(t) + u_i(t) + d_i(t) \tag{4}$$

2.3 Control of the Tracking Error and the Synchronization Error

The tracking error of the ith motor can be written as

$$e_i(t) = \omega_i(t) - \omega_d(t) \tag{5}$$

where $\omega_d(t)$ is the given angular speed, so

$$\dot{e}_i(t) = \dot{\omega}_i(t) - \dot{\omega}_d(t) \tag{6}$$

If we want $e_i(t)$ to converge to zero in a progressive manner, it can be assumed that

$$\dot{e}_i(t) = k_i e_i(t) \tag{7}$$

where $k_i < 0$.

The sliding mode control approach is recognized as one of the efficient tools to design robust controllers for dynamical systems with uncertainties. The major advantage of sliding mode is low sensitivity to system parameter variations and disturbances which eliminates the necessity of exact modeling. In order to deal with the uncertainty caused by motor parameter variations and disturbances, we establishes the speed sliding mode hyperplane according to Lyapunov stability condition as follows

$$S = f_{\mathrm{i}} e_{\mathrm{i}}(t) \tag{8}$$

To ensure that all the states of the system can reach the sliding mode, the following conditions should be met

$$S \cdot \dot{S} < 0 \tag{9}$$

That is

$$S \cdot \dot{S} = S \cdot \left[f_{\mathrm{i}} \dot{e}_{\mathrm{i}}(t) \right] = S f_{\mathrm{i}} \cdot \left[-a_{\mathrm{i}} \omega_{\mathrm{i}}(t) + u_{\mathrm{i}}(t) + d_{\mathrm{i}}(t) - \dot{\omega}_{\mathrm{d}}(t) \right] < 0$$

Construct a function as follows

$$u_{\mathrm{i}}(t) = a_{\mathrm{i}} \omega_{\mathrm{i}}(t) + \dot{\omega}_{\mathrm{d}}(t) - \eta_{\mathrm{i}} \mathrm{sign}(S) \tag{10}$$

yields

$$S \cdot \dot{S} = S \cdot f_i \cdot \left[d_{\mathrm{i}}(t) - \eta_i \mathrm{sign}(S) \right] = -f_i \cdot \left[\eta_i - \left| d_{\mathrm{i}}(t) \right| \right] |S| \tag{11}$$

where $\mathrm{sign}(S)$ is the sign of *S*, η_{i} is the switching gain, and $\eta_{\mathrm{i}} \geq \left| d_{\mathrm{i}}(t) \right|$.

If $f_{\mathrm{i}} \geq 0$, then $S \cdot \dot{S} \leq 0$. Therefore, $S = 0$, the balance origin of the system, is also the globally asymptotically stable equilibrium point. That is, the tracking errors asymptotically converge to zero. This control method taking into account of the speed perturbation caused by parameter uncertainties of driving motors has strong robustness.

Similarly, we can prove that the synchronization errors asymptotically converge to zero.

2.4 Verification

To verify the proposed ring coupled control strategy, synchronization control of a four-motor driving system is studied using MATLAB. The control scheme uses PID compensators and sliding mode controllers. The simulation model of the controller is shown in Fig.2.

The parameters of the four driving motors are as follows: the rated power is 37kW, the rated voltage is 380V, frequency is 50Hz, stator resistance and impedance is 0.087Ω and 0.8e-3H respectively, rotor resistance and impedance is 0.228 Ω and 0.8e-3H respectively, mutual inductance is 34.7e-3H, rotor inertia is 1.662 kg·m^2, friction coefficient is 0.1 N·m·s, and the number of pole is 4. In order to consider load disturbance, the loads of the four motors as shown in Fig.3 (a), (b), (c) and (d) are set as constant, fluctuation, sudden change and random respectively.

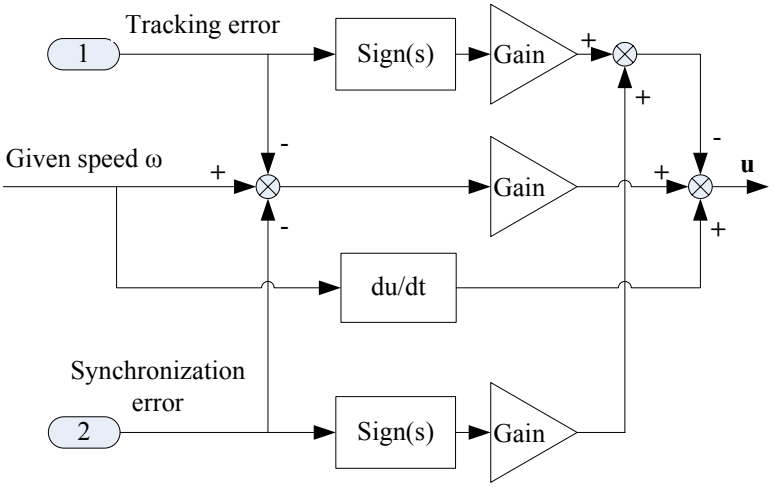


Fig. 2. The simulation model of the controller

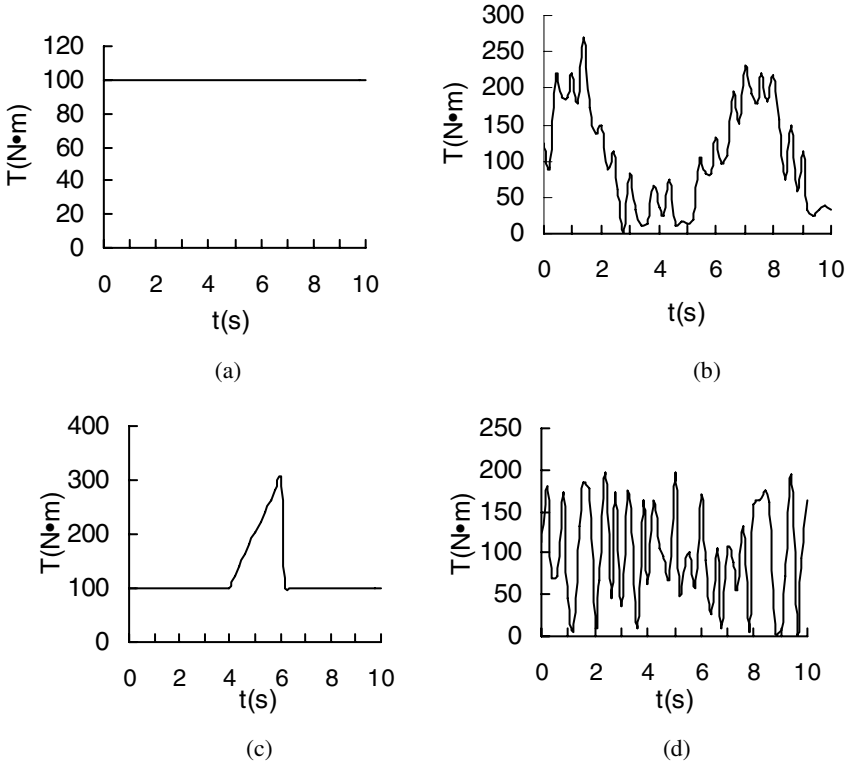


Fig. 3. (a) The load of the first motor (b) The load of the second motor (c) The load of the third motor (d) The load of the fourth motor

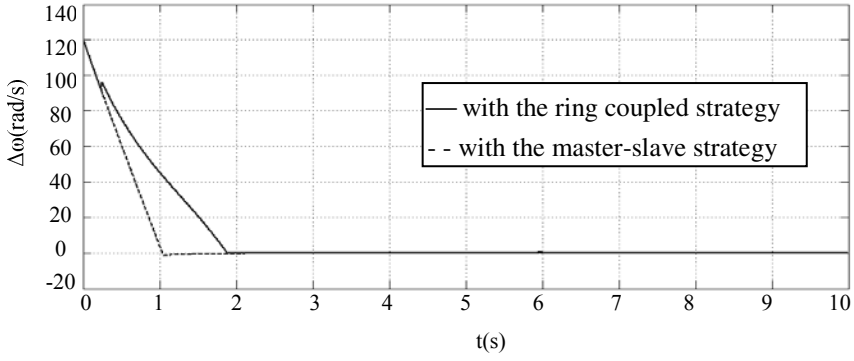


Fig. 4. The tracking error of the first motor

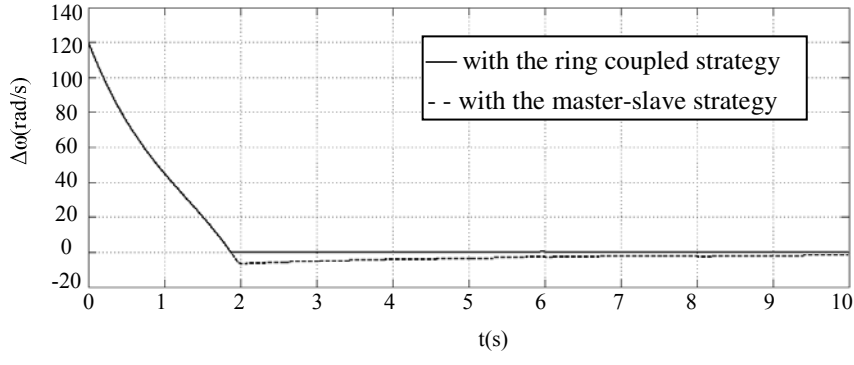


Fig. 5. The tracking error of the second motor

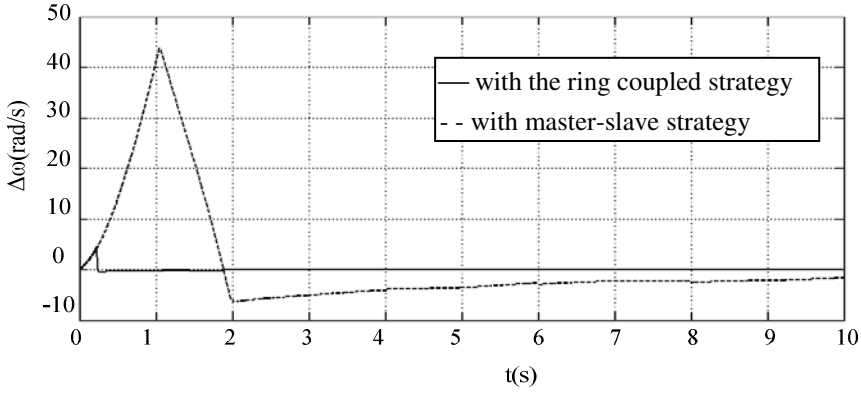


Fig. 6. The synchronization errors between the first motor and the second motor

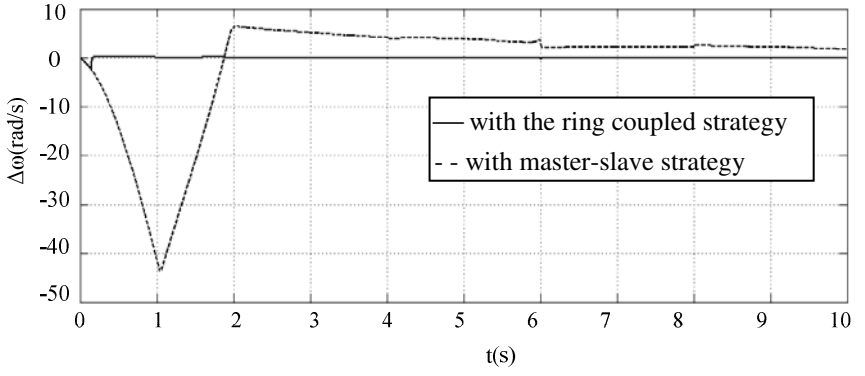


Fig. 7. The synchronization errors between the second motor and the third motor

The synchronization control of the four-motor system is simulated with the ring coupled control strategy and the conventional master-slave control strategy respectively. The tracking errors of the first and second motors are shown in Fig.4 and Fig.5, and the synchronization errors between the adjacent motors are shown in Fig.6 and Fig.7.

In the master-slave control, since the first motor is set as the main motor its speed reaches the given value fast, but other slave motors spend longer time to reach the given speed and are hard to keep constant speed. In the ring coupled control, although the loads of the four motors are very different, all the motors reach the synchronization speed rapidly and stably.

From the above results we can see that multi-motor system with the ring coupled control scheme has fast track performance, smaller synchronization errors, and high synchronization accuracy. The reason is that the sliding mode variable structure control method can effectively inhibit the impact of motor parameter variations and uncertainties on multi-motor synchronization. Besides, this scheme forms a coupling ring of multi-motor that ensures tracking errors of all motors have the same convergence speed, and the same is true to the synchronization errors between motors.

3 Application in Driving of Cutter Head of Shield Machine

The proposed ring couple control strategy was applied to the synchronization control for the driving of cutter head of a prototype shield machine. The joint mechanical and electrical simulation is performed with MATLAB, in which an Adams based subsystem is used for the computation of the mechanical system. The simulation results of the tracking error and synchronization error are shown in Fig.8.

From the results it can be seen that the synchronization control for the cutter head of the shield machine using the ring coupled control strategy can achieve good dynamic performance and synchronization performance.

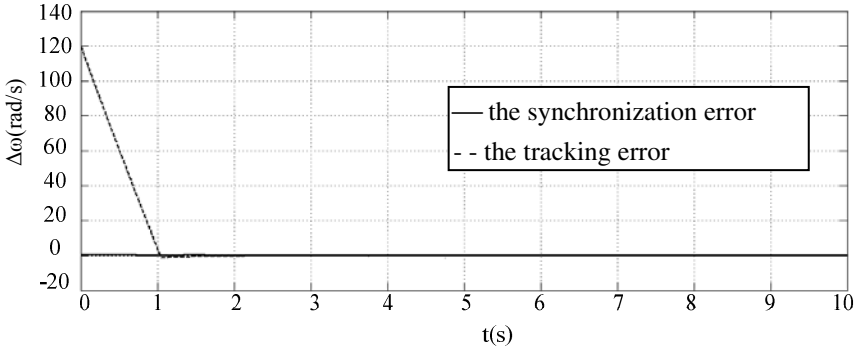


Fig. 8. The tracking error and the synchronization error

In practical engineering, it is necessary for the motor driven shield machines to debug control parameters to eliminate the influences of multi-motor parameter variations and uncertainties before put into operation. Otherwise, shaft broken accident may occur during startup and at load sudden change. In this paper, the ring coupled control strategy and sliding mode variable structure control method deal with the problem caused by motor parameter variations and uncertainties well. The reliability of motor driven shield machine is increased to a certain extent with the synchronization performance improved.

4 Conclusion

This paper proposed the ring coupled control strategy for the synchronization control of the cutter head driving of shield machine. Through simulation analysis and comparison, it can be concluded that the control method based on the ring coupled strategy has higher synchronization accuracy, higher tracking accuracy and higher robustness than that with the traditional master-slave strategy.

References

1. Yuanqi, S., Qianwei, Z., Jianzhong, L.: Driving System for Cutter Head of Shield Tunneling Machine with Large Diameter. Road Machinery & Construction Mechanization 25, 18–20 (2008)
2. Pkrez-Pinal, Francisco, J., NGez, C., Alvarez, R., Cervantes, I.: Comparison of Multi-motor Synchronization Techniques. In: The 30th Annual Conference of the IEEE Industrial Electronics Society, pp. 1670–1675. IEEE Press, Busan (2004)
3. Koren, Y.: Cross-coupled biaxial computer controls for manufacturing systems. Transactions of the ASME 102, 265–272 (1980)
4. Kulkarni, P.K., Srinivasan, K.: Cross-coupled control of biaxial feed drive servomechanisms. ASME Journal of Dynamic Systems, Measurement, and Control 112, 225–232 (1990)

5. Tomizuka, M., Hu, J.S., Chiu, T.C.: Synchronization of two motion control axes under adaptive feed forward control. ASME Journal of Dynamic Systems, Measurement, and Control 114, 196–203 (1992)
6. Feng, L., Koren, Y., Borenstein, J.: Cross-coupling motion controller for mobile robots. IEEE Control Systems 13, 35–43 (1993)
7. Sun, D., Mills, J.K.: Adaptive synchronized control for coordination of two robot manipulators. In: Proceedings of IEEE International Conference on Robotics and Automation, Washington, DC, pp. 976–981 (2002)
8. Sun, D.: Position synchronization control of multiple motion axes with adaptive coupling control. Automatica 39, 997–1005 (2003)
9. Xiao, Y., Zhu, K., Liaw, H.C.: Generalized synchronization control of multi-axis motion systems. Control Engineering practice 13, 809–819 (2005)
10. Moore, P.R., Chen, C.M.: Fuzzy logic coupling and synchronized control of multiple independent servo-drives. Control Engineering Practice 3, 1697–1708 (1995)
11. Lee, H.C., Jeon, G.J.: A neuro-controller for synchronization of two motion axes. International Journal of Intelligent System 13, 571–586 (1998)
12. Yibiao, S., Mengmeng, L., Qingding, G.: Robust synchronization control for dual linear motors based on QFT. Journal of Shenyang University of Technology 30, 249–252 (2008)
13. Chenghui, Z., Qingsheng, S., Jin, C.: Synchronization Control Strategy in Multi-motor Systems Based on the Adjacent Coupling Error. In: Proceedings of the CSEE, vol. 27, pp. 59–63 (2007)

Position and Attitude Precision Analysis of Segment Erector of Shield Tunneling Machine

Hu Shi, Guofang Gong, Huayong Yang, and Rulin Zhou

Zhejiang University, State Key Laboratory of Fluid Power Transmission and Control, Hangzhou, Zheda Road No. 38, 310027, China
tigershi@zj.com

Abstract. Focusing on a segment erector of a shield tunneling machine developed with 6 degrees of freedom and controlled by electro-hydraulic proportional systems, kinematics for the segment erection process is presented. The perturbation method in error analysis is introduced to establish the position and attitude error model, considering a number of factors such as hydraulic drive and control accuracy, tolerance in manufacturing and assembly. Dynamic simulations are carried out to obtain the controlling precision of the electro-hydraulic drive systems. Formulas for calculating the position and attitude error of the grip hand of the segment erector are derived. The calculation results verify the practicality and effectiveness of error analysis, providing a foundation for practical designing, manufacturing and assembling of the segment erecting mechanism.

Keywords: Shield tunneling machine, segment erector, kinematics, position and attitude precision.

1 Introduction

Tunneling and underground space utilization are increasingly becoming the new technology in recent years, with the rapid urban development and the continuous extension of transport systems. In response to these needs, tunnel construction machines are required with high efficiency and safety. As a key part of tunneling machines, segment erector performs the task of circular tunnel formation by erecting the concrete segments into right positions [1][2].

As for the segment erector, efficiency and precision are focused generally. it is a 6-DOF robot in essence to meet the erecting requirements. Efficiency is related to the actuators as well as the backup systems. Precision affects the tunnel quality because of tolerance of tunnel axis. However, most studies have concentrated on the production of lining itself and mechanism design or dynamics. To the best of our knowledge, erecting precision is rarely researched as a special subject [3][4].

This paper develops the kinematics of the 6-DOF segment erector by means of coordinate transformation, and analyzes the electro-hydraulic drive and manufacture errors. Based on an actual erecting process, the position and attitude precision is obtained to evaluate the erecting work quantificationally.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 355–363, 2009.

2 Specifications of Segment Erector

A segment erector works to place the segment in the right position and adjust its attitude properly. It is a multi-actuator robotic system with electro-hydraulic drive, as shown in Fig. 1. There are dual cylinders to lift the segment in vertical direction, a sliding cylinder to pull or push the gripping mechanism in horizontal direction, a hydraulic motor to drive the gear ring, thereby the whole erecting equipment to reach every position in circumferential direction. Near the bottom of the erector, a mechanism is equipped to force the gripped segment to perform the roll, pitch and yaw motions so that the segment can be kept in expected attitude.

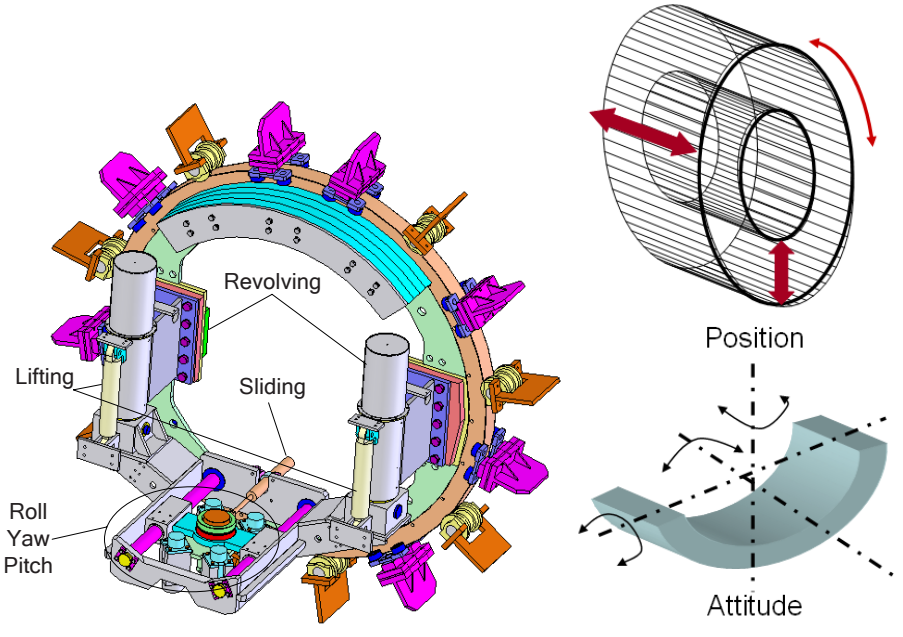


Fig. 1. Schematic of segment erector, position and attitude of segment

In fact, the segment will reach anywhere of a hollow cylindrical space with 3 DOFs attitudes of rotary actions. The erector has a relatively complicated and compound operation process. All motions are not independent, they depend on the preceding action and influence the sequent motion. So error transfer from one action to another is a vital problem to guarantee higher precision and efficiency in erection.

3 Kinematics Equations

To investigate the position and attitude of the segment, the motion of the grip hand of the erector would be considered because it directly reflects the action of the segment fixed on the grip hand. Fig. 2 shows the geometric model and the coordinate system of the segment erector. It has two prismatic joints and one rotary joint for positioning

the segment, another three rotary joints for posing. The position and attitude of the segment is dependent upon the joint angels θ_1, θ_4, θ_5, θ_6, and the joint displacements d_2, d_3. The structure parameters and the joint variables are listed in Table 1. Based on these models, the relations between joints are described as follows with coordinate transformation matrix [5].

$$A_1^0 = \begin{bmatrix} \cos\theta_1 & 0 & \sin\theta_1 & 0 \\ \sin\theta_1 & 0 & -\cos\theta_1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{1}$$

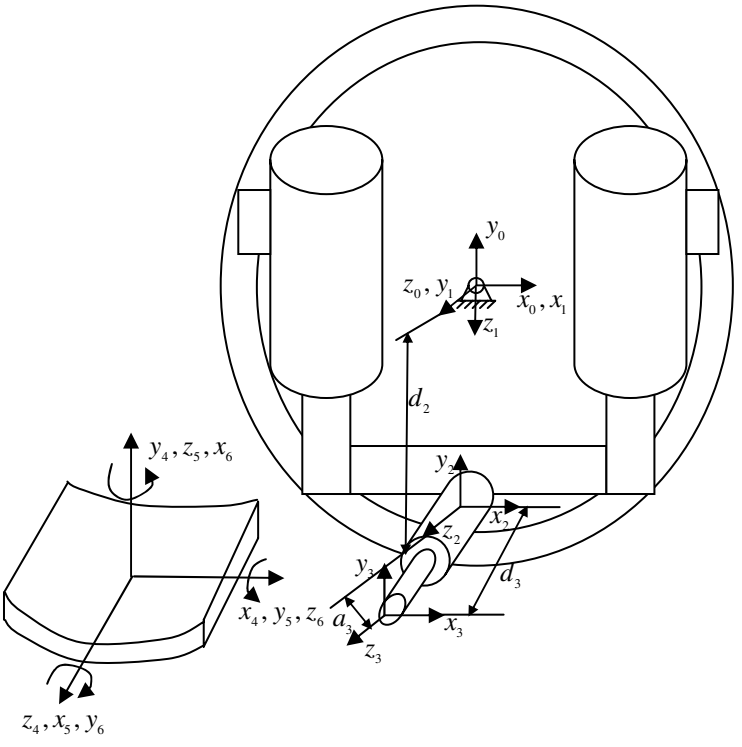


Fig. 2. Schematic diagram of coordinate system of erection

Table 1. D-H parameters of all joints of segment erector

i	θ_i	d_i	a_i	α_i
1	θ_1	0	0	π/2
2	0	d_2	0	-π/2
3	0	d_3	a_3	0
4	θ_4	0	0	0
5	θ_5	0	0	-π/2
6	θ_6	*0*	0	-π/2

$$A_2^1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & d_2 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{2}$$

$$A_3^2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & a_3 \\ 0 & 0 & 1 & d_3 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{3}$$

where A_i^{i-1} (i=1, 2, 3) represents the transformation from coordinate system i-1 to i.

For the attitude of the grip hand or the segment, it acts in agreement with RPY transformation. In other words, it follows the rotary motions of roll, pitch and yaw. So the attitude of the segment with respect to coordinate system 3 can be derived as

$$A_s^3 = \begin{bmatrix} \cos\theta_4\cos\theta_5 & \cos\theta_4\sin\theta_5\sin\theta_6 - \sin\theta_4\cos\theta_6 & \cos\theta_4\sin\theta_5\cos\theta_6 + \sin\theta_4\sin\theta_6 & 0 \\ \sin\theta_4\cos\theta_5 & \sin\theta_4\sin\theta_5\sin\theta_6 + \cos\theta_4\cos\theta_6 & \sin\theta_4\sin\theta_5\cos\theta_6 - \cos\theta_4\sin\theta_6 & 0 \\ -\sin\theta_5 & \cos\theta_5\sin\theta_6 & \cos\theta_5\cos\theta_6 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{4}$$

The position and attitude of the end of the grip hand with respect to the reference coordinate system {0} can be denoted as

$$A_s^0 = A_1^0 A_2^1 A_3^2 A_s^3$$

$$\begin{bmatrix} s1s5 + c1c4c5 & c1(c4s5s6 - c6s4) - c5s1s6 & c1(s4s6 + c4c6s5) - c5c6s1 & d2s1 - a3s1 \\ c4c5s1 - c1s5 & c1c5s6 - s1(c6s4 - c4s5s6) & s1(s4s6 + c4c6s5) + c1c5c6 & a3c1 - c1d2 \\ -s5 & c5s6 & c5c6 & d3 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{5}$$

where si (i=1, 4, 5, 6) is short for $\sin\theta_i$, and ci (i=1, 4, 5, 6) is short for $\cos\theta_i$.

Suppose the target position and attitude of the segment is

$$T = \begin{bmatrix} \boldsymbol{n} & \boldsymbol{o} & \boldsymbol{a} & \boldsymbol{p} \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} n_x & o_x & a_x & p_x \\ n_y & o_y & a_y & p_y \\ n_z & o_z & a_z & p_z \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{6}$$

where $\boldsymbol{n}$=[n_x, n_y, n_z]^T, $\boldsymbol{o}$=[o_x, o_y, o_z]^T, $\boldsymbol{a}$=[a_x, a_y, a_z]^T, $\boldsymbol{p}$=[p_x, p_y, p_z]^T are target position and attitude vectors. Then let the target matrix equal to the transformation matrix, that is

$$T = A_s^0. \tag{7}$$

Based on this equation, the kinematic inverse solutions for segment erection are obtained by equalizing every matrix element in the corresponding place. By this way, motion control for each actuator can be performed according to a given position and attitude object.

4 Precision Analysis

Perturbation method for error modeling in robotics is widely used for its easy realization and adaptability to multiple error sources. In the segment erecting kinematics, errors result from two factors generally: drive precision and manufacturing tolerance. Write equations (1)-(5) as the following format:

$$A_i^{i-1} = \begin{bmatrix} R_i^{i-1} & P_i^{i-1} \\ 0 & 1 \end{bmatrix} \text{ or } A_s^0 = \begin{bmatrix} R_s^0 & P_s^0 \\ 0 & 1 \end{bmatrix} . \tag{8}$$

Assume that there exist Δq_i displacement error and $\Delta\varphi_i$ angle error between joint i-1 and i, and the errors at the end of the transmission chain will be

$$\Delta q = \sum_{i=1}^{n} \Delta q_i + \sum_{i=1}^{m} \Delta\varphi_i \times p_i . \tag{9}$$

$$\Delta\varphi = \sum_{i=1}^{m} \Delta\varphi_i . \tag{10}$$

where n and m are numbers of linear and angular displacement [6].

4.1 Drive Errors

Motions of the segment erector are driven by electro-hydraulic systems. So the position and attitude precision of the segment is influenced by the control accuracy of every actuator in the electro-hydraulic system, and the error transfers from one drive part to another and accumulates continuously. Based on the above kinematics as well as the working principle of the segment erector, an automatically operated erecting process will be as shown in Fig. 3.

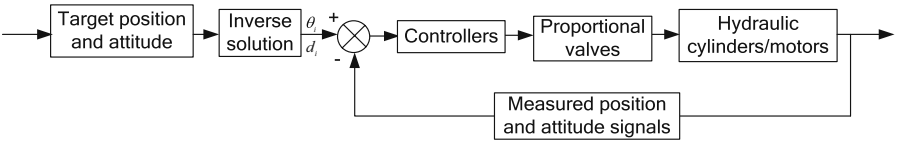


Fig. 3. Block diagram of segment erector control system

To obtain the errors generated in hydraulic transmission, as shown in Fig 4, a simulation model corresponding to the actual segment erector is created in AMESim software to simulate the hydraulic drive and control systems. The parameters are set in accordance with the actual system as shown in Table 2. During simulation, PID controllers are employed to construct closed loop control systems.

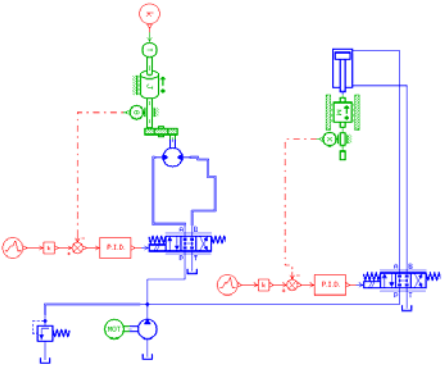


Fig. 4. Simulation model of electro-hydraulic control system of segment erector

Table 2. Parameters of electro-hydraulic control system of segment erector

Parameters	values
Mass of segment	1000kg
Size of lifting cylinder	Ø80/32×380
Size of sliding cylinder	Ø50/25×320
Displacement of rotary motor	37ml/r
Size of yaw cylinder	Ø50/25×40
Size of roll cylinder	Ø80/32×40
Size of pitch cylinder	Ø80/32×40
Frequency of proportional valve	25Hz
Maximum radius of gyration	1.5m

Table 3. Motion parameters and errors of simulation

Parameters	Values	Errors
Rotary angle	90°	0.003°
Lifting displacement	0.3m	0.5mm
Sliding displacement	0.2m	0.5mm
Yaw angle	2°	0.001°
Roll angle	2°	0.001°
Pitch angle	2°	0.001°

In fact, the attitude of segment is adjusted by the hydraulic linear actuators, and the results of measurements performed with the linear displacement sensors will be changed into the attitude angles of the segment. So the linear error will be translated into the angle error respectively.

Neglect the mechanical error sources such as motor backlash, simulation is carried out with the motion parameters listed in Table 3. Given a step signal, consider the output displacements of the actuators tracking the input signals, and take the error in steady stage as the positioning accuracy. Errors brought about by the electro-hydraulic control systems are presented in column 3 of Table 3.

4.2 Manufacturing Tolerance

Due to the movement of the joint and the process of machining and assembly, it is necessary to permit some clearance between the moving parts. However, the clearance may cause error when the joint is moving, and inaccurate position and attitude arises [7]. So machining error is another factor to be taken into account to deal with the precision of segment erection, this is also analyzed by the perturbation method.

4.3 Error Calculation

Error calculation is based on equations (9) and (10), which indicate the effect of the error of input parameter on the end of the whole transmission chain. Those two parts of error sources mentioned above should be calculated separately, because they propagate in different ways.

Assume that the joint parameters θ_i, d_i, α_i, a_i have small errors $\Delta\theta_i$, Δd_i, $\Delta\alpha_i$, Δa_i, then equations (9) and (10) can be changed into

$$\begin{aligned}\Delta q &= \sum_{i=1}^{s}(\Delta a_i + \Delta d_i + \Delta\theta_i \times P_{\theta_i} + \Delta\alpha_i \times P_{\alpha_i}) \\ &= R_s^0 \sum_{i=1}^{s}[R_i^{i-1}\begin{bmatrix}\Delta a_i \\ 0 \\ 0\end{bmatrix} + \begin{bmatrix}0 \\ 0 \\ d_i\end{bmatrix} + R_i^{i-1}\begin{bmatrix}\Delta\alpha_i \\ 0 \\ 0\end{bmatrix} \times \begin{bmatrix}0 \\ 0 \\ d_i\end{bmatrix} + (R_i^{i-1}\begin{bmatrix}\Delta a_i \\ 0 \\ 0\end{bmatrix} + \begin{bmatrix}0 \\ 0 \\ \Delta\theta_i\end{bmatrix}) \times P_i^{i-1}] \end{aligned} . \quad (11)$$

$$\Delta\varphi = \sum_{i=1}^{s}(\Delta\theta_i + \Delta\alpha_i) = R_s^0 \sum_{i=1}^{s}(R_i^{i-1}\begin{bmatrix}\Delta\alpha_i \\ 0 \\ 0\end{bmatrix} + \begin{bmatrix}0 \\ 0 \\ \Delta\theta_i\end{bmatrix}) . \quad (12)$$

Assume that the mechanical inaccuracy of component i with respect to i-1 in translation is $d_i^e = [\Delta u_i \quad \Delta v_i \quad \Delta w_i]^T$, and that in rotation is $\delta_i^e = [\Delta\eta_i \quad \Delta\xi_i \quad \Delta\zeta_i]^T$. Substituting them into equations (9) and (10), the position and attitude error induced by structure clearance can be obtained as

$$\Delta q = [dx \quad dy \quad dz]^T = R_s^0 \sum_{i=1}^{s} R_e^i (d_i^e + \delta_i^e \times P_s^i) . \quad (13)$$

$$\Delta\varphi = [\delta x \quad \delta y \quad \delta z]^T = R_s^0 \sum_{i=1}^{s} R_e^i \delta_e^i . \quad (14)$$

Choosing a set of structure parameters, drive and mechanical inaccuracy values, as shown in Table 4, substitute them into corresponding calculation equations respectively, the results are obtained as follows.

Position error by drive: $\Delta q_1 = [0.34 \quad 0.47 \quad -0.34]^T$.

Attitude error by drive: $\Delta\varphi_1 = [0.027 \quad 0.059 \quad -0.027]^T$.

Position error by clearance: $\Delta q_2 = [0.0675 \quad 0.1702 \quad -0.0675]^T$.

Attitude error by clearance: $\Delta\varphi_1 = [-0.02 \quad 0.039 \quad -0.02]^T$.

The parameters chosen for modeling, analysis and calculation are referring to an actual working condition. With position error less than 1mm and attitude error less than 0.1°, Such precision will definitely satisfy the segment erection requirements of shield tunneling.

Table 4. Structure parameters and errors of segment erector

i	θ_i (°)	d_i (m)	a_i (m)	α_i (°)	$\Delta\theta_i$ (°)	Δd_i (mm)	Δa_i (mm)	$\Delta\alpha_i$ (°)	Δu_i (mm)	Δv_i (mm)	Δw_i (mm)	$\Delta\eta_i$ (°)	$\Delta\xi_i$ (°)	$\Delta\zeta_i$ (°)
1	90	0	0	90	0.003	0	0.5	0.05	0.01	0.01	0.01	0.001	0.001	0.001
2	0	0.3	0	-90	0	0.5	0.5	0.05	0.01	0.01	0.01	0.001	0.001	0.001
3	0	0.2	0.1	0	0	0.5	0.5	0.05	0.01	0.01	0.01	0.001	0.001	0.001
4	2	0	0	0	0.001	0	0.5	0.05	0.01	0.01	0.01	0.001	0.001	0.001
5	2	0	0	-90	0.001	0	0.5	0.05	0.01	0.01	0.01	0.001	0.001	0.001
6	2	0	0	-90	0.001	0	0.5	0.05	0.01	0.01	0.01	0.001	0.001	0.001

5 Conclusion

In this paper, kinematics analysis of a prototype segment erector is conducted and the kinematics relations between moving parts and equations are derived with coordinate transformation matrix. The position and attitude precision of erection is analyzed by means of perturbation method in light of two types of error sources of drive and manufacture. The error calculation model established will provide a reference for the design and control of segment erection. Based on the theoretical analysis and the calculation results, the following conclusions can be drawn.

(1) The segment erecting mechanism and its drive and control systems can satisfy the requirements of tunneling with sufficient position and attitude precision.
(2) The error sources exert more influence on position precision than on attitude precision, so segment position should be placed more emphasis on in erection control.
(3) Errors induced by the control accuracy of the drive systems are lager than those arise from manufacturing tolerance, so there is no special requirement for mechanical processing, and electro-hydraulic drive accuracy should be the main control objective.

Acknowledgments

This work is supported by National Basic Research Program (973 Program) of China (No. 2007CB714004) and National High-tech Research and Development Program (863 Program) of China (No. 2007AA041806), Their support is gratefully appreciated.

The authors also wish to acknowledge the help of the engineers in Shenyang Heavy Machinery TBM Company.

References

1. Wada, M.: Automatic Segment Erection System for Shield Tunnels. Advanced Robotics 5(4), 429–443 (1991)
2. Yali, H., Kui, C.: Segment Erecting Technique of Shield Tunneling in Najing Metro. Tunnel Construction 23(2), 16–17 (2003)
3. Yasuo, T.: Automatic Segment Assembly Robot for Shield Tunneling Machine. Computer-Aided Civil and Infrastructure Engineering 10(5), 325–337 (2008)
4. Tanaka, Y.: Automatic Segment Assembly Robot for Shield Tunneling Machine. Microcomputers in Civil Engineering 10(5), 325–337 (1995)
5. Youlun, X.: Fundamental of Robotics. Huazhong University of Science and Technology Press, Wuhan (1996)
6. Guotai, J., Egorov, O.D., Yueqing, Y.: A Method for Calculating Position and Orientation Errors of Industrial Robot. Mechanical Science and Technology 21(1), 35–37 (1999)
7. Shuanghua, Y., Yilun, W., Lixun, Z., Ce, Z., Caixia, C.: Location Precision Analysis of Underwater Robot Auto-toolkit. Journal of Harbin Engineering University 22(2), 46–50 (2001)

An Improved Model of Loads Acting on Shield

Xiangtao Hu, Zhouping Yin, and Yongan Huang

State Key Laboratory of Digital Manufacturing Equipment and Technology,
Huazhong University of Science & Technology, Wuhan, China, 430074
hust_hoo@163.com

Abstract. Shield tunneling technologies have been developed for constructing tunnels in soft ground especially with groundwater. However, the automatic deviation correction systems are based on empirical relationships or theoretical model including a great deal of unknown parameters. In this paper, the shield-segment system is considered as underground parallel manipulators, so the shield behavior can be represented by the rotation of motion platform around the static platform. Consequently, load model acting on the shield is developed based on the proposed relationship between the shield behavior and the ground displacement. The examples of straight alignment and curve alignment were studied, comparing the results with the others model, and the results validate the load model studied here.

Keywords: Shield tunneling, shield behavior, ground displacement, automatic deviation correction.

1 Introduction

Shield tunneling technologies have been developed for constructing tunnels in soft ground especially under groundwater. Due to the large size of shield and complex working condition, such as geological diversity, multi-physics coupling, the loads distribution on the shield have great singularity and acute fluctuation, and easily leads to the deviations of shield from the planned alignment [1-3]. However, the present deviation correction control systems are based on empirical relationships or theoretical model including a great deal of unknown parameters. Consequently, there is narrow applicability, low precision and weak controllability. To improve the construction accuracy, it is necessary to model the loads acting on the shield, which influence the shield behavior,

Nowadays, the load model has been studied by both statistical and theoretical methods. Szecky proposed an approximated formula to calculate the jacks thrust by statistical analysis methods based on in situ data [4]. Shimizu and Suzuki and Shimizu et al. established the relationship between the shield attitude and the jacks moment [5]. Mitsutaka Sugimoto et al. studied the load model by theoretical methods, and recommended that the loads acting on the shield are combination of five forces: Self-weight, force on the shield tail, the jack thrust, load on the face, and load on the shield's periphery [6]. Aphichat Sramoon et al. believed that the excavated area, tail clearance, rotation direction of the cutter face, ground loosening at the shield crown, and dynamic

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 364–373, 2009.

equilibrium conditions are the factors affecting the shield behavior. In particular, the ground displacements around the shield and at cutter face are considered to be the predominant factor. All the factors above should be considered in load model [7]. Yongan Huang et al. presented a novel force planning method for underground articulated robot based on the interaction between soil, structure and motion [8]. From the viewpoint of engineering practices, these methods are useful, since the shield can be controlled by using up-to-date unknown parameters in the formula based on measured data during excavation [6]. However theoretical methods which formulations include a great deal of parameters are too complex to achieve engineering application and statistical methods can not control the shield behavior without case records.

To solve above problems, this paper develops the load model based on new proposed relationship between shield behavior and ground displacement. Furthermore, the examples of straight alignment and curve alignment have been carried out, and the validity of the model was discussed.

2 Static Analysis and Geometric Model

An earth pressure balance (EPB) shield with four jacks is studied in this paper, and there is a supporting shoe at the end of each jack connected with lever by joint. The loads acting on the shield are changing timely during excavation, the shield attitude is time-variant like snake motion. Since the axes of jacks and shield are parallel, the process can be described by the static analysis of a jack segregated from the shield system as shown in Fig.1. The jack attitude is composed of combinations of supporting shoe slide on the segment ring plane and rotation around the joint, which are affected much by the uncontrollable transversal loads.

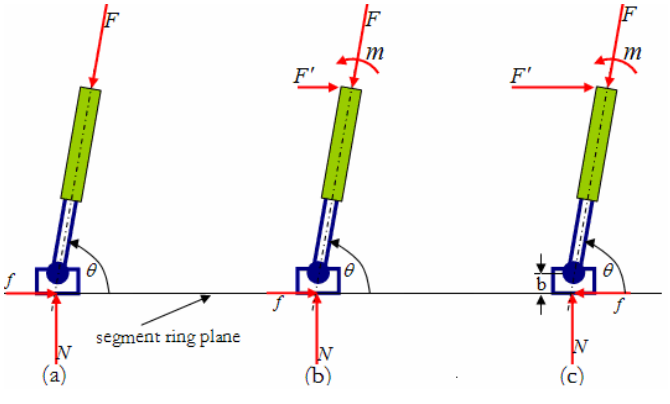


Fig. 1. The force diagram of jack

When the transversal load is balanced as illustrated in Fig.1 (a), slide cannot happen because the jack is self-locked, and the transversal loads satisfy inequation as follows:

$$F\cos\theta - \mu F\sin\theta < 0 \tag{1}$$

where F =the jack thrust, θ = the angle from the jack axis to the segment ring plane, μ =the coefficient of mobilized friction.

When the transversal load is unbalanced, the jacks' thrusts should be adjusted to produce a moment m for balancing the moment induced by the unbalanced force, as illustrated in Fig.1 (b) (c). The slide will occurr when the transversal loads satisfy inequation as:

$$F' - F\cos\theta - \mu F\sin\theta > 0 \tag{2}$$

And the rotation will occurr when the transversal loads satisfy inequation as:

$$\begin{cases} F' - F\cos\theta \le 0 \\ F'l\sin\theta - m \ne 0 \end{cases} \tag{3}$$

$$\text{Or}\begin{cases} F' - F\cos\theta > 0 \\ F'l\sin\theta + fb + Nb\cot\theta - m \ne 0 \end{cases} \tag{4}$$

where b =the distance from the joint to the segment ring plane, m =the moment induced by jacks' thrusts, f =the friction, N =the support force, F' = unbalanced transversal force.

Actually, the jack thrust is great, so that the friction is enough to resist the slide. Consequently, the shield-segment ring system can be considered as a kind of underground parallel manipulators which the machine body, the segment ring and the jacks are the motion platform, the static platform and the branched chains respectively, as illustrated in Fig.2. Then the shield behavior represents the rotation of motion platform around the static platform. To model each force, the global coordinate system $O-XYZ$ is selected so that the X axis is vertically upwards; the origin O is on the center of the segment ring plane, and the Y and Z axes are on a horizontal plane.

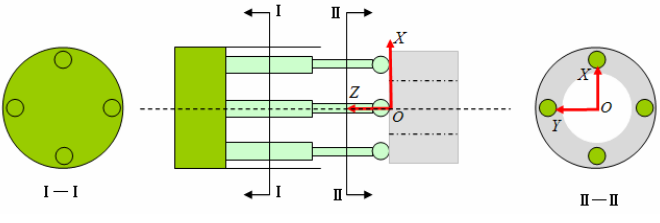


Fig. 2. The sketch map of EPB shield

From the viewpoint of geometric model, the deformation of shield and segment, the angle of θ and the attitude of segment ring are the factors affecting the shield behavior. For simplifying the model, the following were assumed:

1. The concrete segment ring is considered as a rigid body, and its attitude is fixed during excavation.
2. The shield self-rolling is not considered which is limited in practice.
3. The shield is considered as a rigid body.

3 Calculation of the Earth Pressure and the Ground Displacement

The shield machine meets the static equilibrium conditions for steadily advance resembling multi-legged walking robots. The difference between them is that the former loads are uncertain and the latter are determined [9]. The earth pressure is the main uncertain loads whose essence is the contact stress of soil and structure interaction. The pressure and distribution are affected by the shield stiffness, the geological formation, the overburden depth and so on. The relationship between the ground displacement and the earth pressure can be represented by [10]:

$$K_s(U_s)=\begin{cases}(K_{so}-K_{s\min})\tanh\left[\dfrac{a_sU_s}{K_{so}-K_{s\min}}\right]+K_{so}(U_s\le 0)\\ (K_{so}-K_{s\max})\tanh\left[\dfrac{a_sU_s}{K_{so}-K_{s\max}}\right]+K_{so}(U_s\ge 0)\end{cases} \tag{5}$$

where K = the coefficient of earth pressure which is defined as the earth pressure σ divided by the initial vertical earth pressure σ_{vo} ($K_s=\sigma_s/\sigma_{vo}$); U =the ground displacement; $s=v$ or h , subscripts $v=$ the vertical and h =the horizontal directions, respectively; a =the gradient of functions $K(U)$, which represents the coefficient of subgrade reaction k ($a_s=k_s/\sigma_{vo}$);and subscripts o = the initial, min = the lower limit, and max = the upper limit of the coefficient of earth pressure, respectively.

The coefficient of earth pressure in any direction K_ϑ can be interpolated by using K_h and K_v as:

$$K_\vartheta(U_n,\vartheta)=K_v(U_n,\vartheta)\cos^2\theta+K_h(U_n,\vartheta)\sin^2\vartheta \tag{6}$$

where ϑ = the angle measured from the P axis to calculation point.

The normal ground displacement around shield periphery is illustrated in Fig.3 which are from the simulation results of theoretical model [6]. In the case of straight line, the normal ground displacement is nearly zero, and there is a little displacement at the tail because of the knocking-head advancing. In the case of curve alignment, the soil at the concave side is compression, whereas at the convex side is extension, and displacement is almost symmetrical. The displacement is nearly zero at the cutter face since the soil is just excavated and is not disturbed by the shield attitude changing.

According to above analysis, the ground displacement around shield periphery is assumed as following:

$$\begin{aligned}U_n&=U_h=(-2/l_2\cdot x^2+2\cdot x)\theta\sin\vartheta(0<x<l_2)\\ U_n&=U_h=x\theta\sin\vartheta(-l_1<x<0)\end{aligned} \tag{7}$$

where l_1 =the distance from the tail to the XOY plane, l_2 =the distance from the cutter face to the XOY plane. Here only consider the horizontal yawing angle; the others can be dealt with by the same approach. In the case of curve alignment, the contour line diagram, according to the formulation of the displacement, is illustrated in Fig.4 based on artificial data, which coincide with the simulation results.

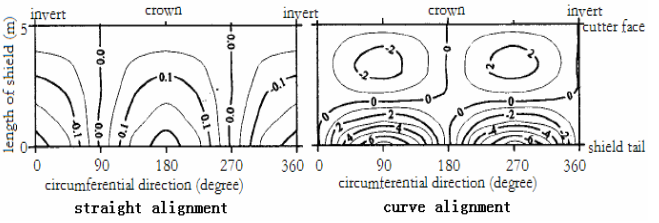


Fig. 3. Normal ground displacement around shield periphery (mm) [6]

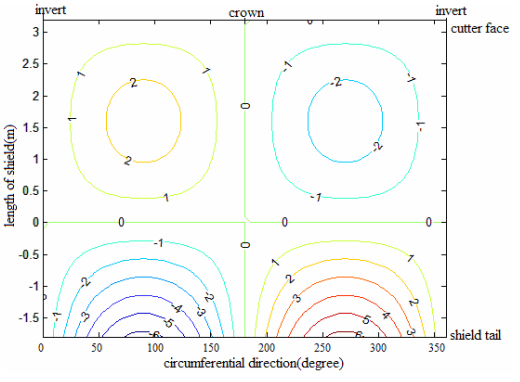


Fig. 4. Normal ground displacement around shield periphery (mm)

The ground displacement at the cutter face is variational with the shield attitude changing. The displacement at the cutter face is assumed as following:

$$U_h = r_c \theta \sin \vartheta, -r < r_c < r \tag{8}$$

where r =the shield radius, r_c =the distance from the calculation point to the R axis.

4 The Load Model

The loads acting on the shield can be divided into: Self-weight, force on the shield tail, the drive force, force acting at the face, and force acting on the shield periphery. Each force is a function of the ground properties and the shield behavior. All those forces compose equilibrium conditions, and drive force can be obtained by solve the equation. The load model is developed based on the same idea with reference [6]. The improvements are simplification of parameters and continuous models instead of discrete ones.

4.1 The Self-weight

The self- weight is composed of the shield machine and the soil in the chamber. The friction f_w and the pitching moment m_{1w} caused by the gravity can be defined as:

$$f_w = \mu w \qquad m_{1w} = w r_w \tag{9}$$

where μ =the coefficient of mobilized friction between the soil and the shield, r_w =the distance from the center of gravity to the global coordinate system, w =the self weight.

4.2 Force Acting at the Face

The force at the face can be divided into two components the earth pressure and the penetration [6]. The coefficient of earth pressure can be considered as K_h , and the earth pressure is proportional in the vertical direction. The penetration is the function of earth pressure, advance velocity and rotation speed, and can be obtained from:

$$\sigma_p = a\sigma_o v / \omega r \,,\ \sigma_o = K_h \rho g (h_0 - r_c \cos\theta) \tag{10}$$

where v =the advance velocity, ω = the cutter face rotation speed, ρ =the average density of the earth, g =the gravity acceleration, h_o =the distance from the origin of global coordinate system to the ground, r_c =the radius of the calculation point on cutter face.

The resistance f_c at cutter face and the moment are calculated as:

$$f_c = (1-\alpha_0)\int(\sigma_p + \sigma_o)ds \tag{11}$$

$$m_{1c} = (1-\alpha_0)\int(\sigma_p + \sigma_0) r_c \cos\vartheta ds \tag{12}$$

$$m_{2c} = (1-\alpha_0)\int(\sigma_p + \sigma_0) r_c \sin\vartheta ds \tag{13}$$

where α_0 =the open ratio of cutter face, ds =the area element, m_{1c} =the pitching moment, m_{2c} =the yawing moment.

4.3 Force Acting at the Shield Periphery

The normal earth pressure acting on the shield periphery can be defined as:

$$\sigma_n = K_\vartheta(U_n, \vartheta)\sigma_{vo} \,,\ \sigma_{vo} = \rho g (h_o - r\cos\vartheta) \tag{14}$$

The axial resistance f_{P} is composed of friction and mobilized cohesion, and the moments can be defined as:

$$f_{\mathrm{P}} = \iint(\mu\sigma_n + c)ds \tag{15}$$

$$m_{1\mathrm{P}} = \iint(\mu\sigma_n + c) r\cos\vartheta ds + \iint \sigma_n l \cos\vartheta ds \tag{16}$$

$$m_{2\mathrm{P}} = \iint(\mu\sigma_n + c) r\sin\vartheta ds + \iint \sigma_n l \sin\vartheta ds \tag{17}$$

where c =mobilized cohesion, l =the distance from the calculation point to the XOY plane, $m_{1\mathrm{P}}$ = yawing moment, $m_{2\mathrm{P}}$ = pitching moment.

4.4 Forces on Shield Tail

The forces on shield tail are composed of two parts: the force due the wire brush deformation and the applied grease pressure. The pressure σ_t at the tail can be found from:

$$\sigma_t = k_{wb}(t_{wb} - l\Delta\alpha \sin\vartheta) + \sigma_a \tag{18}$$

where k_{wb} =the spring constant of the wire brush, t_{wb} = the original height of the wire brush, σ_a =the applied grease pressure between the individual rows of wire brushes. The axial resistance f_t and moments can be obtained as follows:

$$f_t = \iint (\mu\sigma_t + c_{wc})ds \tag{19}$$

$$m_{1t} = \iint (\mu\sigma_t + c_{wc}) r\cos\vartheta ds + \iint \sigma_t l\cos\vartheta ds \tag{20}$$

$$m_{2t} = \iint (\mu\sigma_t + c_{wc}) r\sin\vartheta ds + \iint \sigma_t l\sin\vartheta ds \tag{21}$$

where c_{wc} =mobilized cohesion between the segment and the wire brush, m_{1t} = pitching moment, m_{2t} = yawing moment.

4.5 Drive Force Model

The shield which has quasistatic character is always governed by the static equilibrium equation, the drive force can be obtained by solve the equation. According to above analysis, the drive force and moments are defined as follows:

$$F = f_w + f_P + f_c + f_t \tag{22}$$

$$m_1 = m_{1c} + m_{1w} + m_{1P} + m_{1t} \tag{23}$$

$$m_2 = m_{2c} + m_{2P} + m_{2t} \tag{24}$$

where F =the drive force, m_1 =the drive pitching moment, m_2 =the drive yawing moment. Each jack thrust can be calculated as follows:

$$F_1 = (F - \Delta f_1 - \Delta f_2)/4 \tag{25}$$

$$F_2 = (F - \Delta f_1 - \Delta f_2)/4 + \Delta f_2 \tag{26}$$

$$F_3 = (F - \Delta f_1 - \Delta f_2)/4 + \Delta f_y \tag{27}$$

$$F_4 = (F - \Delta f_1 - \Delta f_2)/4 \tag{28}$$

where F_i =jack thrust; $\Delta f_1, \Delta f_2$ =the equivalent axis loads, $\Delta f_1 = m_1 / r'$, $\Delta f_2 = m_2 / r'$, r' = the radius of jack distribution.

5 Examples

The basic examples are composed of combinations of horizontal tunnel alignment and sandy ground. The calculation data are illustrated in Table 1, which are from the paper of Misutaka Sugimoto [6]. The results of improved model are shown in Table 2, and they indicat that:

1. In the composition of driving force, the resistance at the cutter face accounts for 59, and the resistance acting at the shield periphery accounts for 39. The proportions of yawing moments induced by the force at the cutter face and force acting on

Table 1. The calculation data [6]

Shield	Outer radius	2m
	Total length	5m
	Self-weight	1000kN
	Radius of jack distribution	1.75m
	Ctter face from global coordinate system origin	3.5m
	Center of gravity in machine coordinate system	(0 0 -0.5)m
	Length of jack	2.0m
	Number of jacks	4
Shield operation	Shield excavation velocity	0.04m/min
	Cutter face rotation speed	0.9rpm
Shield tail	Thickness of shield tail	0.14m
	Length of wire brush from tail	0.45m
	Spring constant for wire brush	100kN/m^3
	Grease pressure	0
Ground	Specific gravity	2.7
	cohesion	0
	Coeff. of earth pressure $K_{h\max}, K_{ho}, K_{h\min}$	0.3,0.5,5.0
	Coeff. of earth pressure $K_{v\max}, K_{vo}, K_{v\min}$	0.3,1.0,5.0
	Coeff. of ground reaction k_h, k_v	30,30MN/m^3
	Earth pressure at face constant	30
	Coeff. of the friction	0.1
tunnel	Horizontal curvature for straight alignment	0/m
	Horizontal curvature for curve alignment	0.002/m
	Overburden depth	20m
	Length of segment	1.0m

shield periphery are about 35 and 65 respectively. Those confirm that ground displacements around the shield and at cutter face are the predominant factors of the shield behavior.

2. There is a little pitching moment because of the earth pressure increasing with overburden depth increment from the crown to the bottom of shield.

Comparing the results with the model proposed by Misutaka Sugimoto (Model b), the following were found:

1. The same points are that the axis drive forces are almost equivalent for both alignments, and the yawing moments are very close each other.
2. However there are three differences: the proportions of each component in models due to different relationship between ground displacement and shield behavior, yawing moments for the reason of neglecting the influence of the rotation direction of cutter face in Model a, and the pitching moments because of not considering the frictions at jacks supporting shoe in Model b. Especially, the difference of pitching moments is great. Actually, there are some vertical frictions at supporting shoe because of the nonuniform earth pressure from the bottom to the crown of shield, which can produce a moment to decrease the unbalanced pitching moment. So the frictions should be considered.

Through the above analysis, the improved model has similar precision with the Model b at the axis and yawing component of driving force. Although there are some differences, the improved model is valid and more reasonable from the theoretical viewpoint.

Table 2. The results of drive force

	Force acting at the face		Force acting at the shield periphery		Forces on shield tail		Self- weight		Drive force	
model	a	b	a	b	a	b	a	b	a	b
straight alignment										
Axis loads	3789	5445	2494	941	0	0	100	0	6383	6386
Yawing moments	0	36	0	27	0	0	0	0	0	63
Pitching moments	-291	272	1455	1170	0	0	-1200	-500	-36	-942
Curve alignment										
Axis loads	3932	5445	2506	997	4	0	100	0	6542	6442
Yawing moments	-1010	-39	-1877	-2576	0	0	0	0	-2887	-2615
Pitching moments	-180	-272	1255	-1057	-4	0	-1200	500	-129	-829

Note: Model a is proposed in this paper, Model b is proposed by Misutaka Sugimoto [6].

6 Conclusions

In this paper, a load model is developed for real-time deviation correction. Through the static analysis of jack, the shield-segment ring system can be considered as a kind of underground parallel manipulators, and the shield behavior represents the rotation of

motion platform around the static platform. Thus the load model is proposed based on the relationship between the ground displacement and the shield behavior. The examples of straight alignment and curve alignment have been carried out. Comparing with theoretical model proposed by Misutaka Sugimoto, the results confirm that new model has similar precision and more reasonable. To great extent, it overcomes some limits and disadvantages of statistical model and theoretical model, such as over-reliance on empirical relationships, plentiful unknown parameters. The major contribution is that the load model lays the foundation for automatic deviation correction control. However, it is should be mentioned, there are some limitation, such as not consider all factors affecting the shield behavior, and the validity should be examined in the shield tunneling test site. In the future, the model needs to be further investigated by simulation and experiments. Besides the statistical and theoretical methods new approaches combined intelligent methods are also worth researching.

Acknowledgement

The authors acknowledge supports from the National Natural Science Foundation of China (50705035 and 50625516), the National Fundamental Research Program of China (973) (2007CB714000).

References

1. Loganathan, N., Poulos, H.G.: Centrifuge model testing of tunnelling induced ground and pile deformations. Geotechnique 50(3), 283–294 (2000)
2. Lee, G.T.K.: The Effects of advancing open face tunneling on an existing loaded pile. Journal of Geotechnical and Geoenvironmental Engineering 131(2), 193–201 (2005)
3. Ueno, T.: Direction control system of shield tunneling machines. Ishikawajima-Harima Engineering Review 37(3), 203–218 (1997)
4. Szechy, K.: The art of tunneling. Akademiai Kiado, Budapest (1966)
5. Shimizu, Y., Suzuki, M.: Study of the moving characteristics of a shield tunneling machines. J. Japan. Soc. Mech. Eng. 58(550), 155–161 (1992)
6. Sugimoto, M., Sramoon, A.: Theoretical model of shield behavior during excavation I: Theory. Journal of Geotechnical and Geoenvironmental Engineering 128(2), 138–155 (2002)
7. Sramoon, A., Sugimoto, M., Kayukawa, K.: Theoretical model of shield behavior during excavation II: Application. Journal of Geotechnical and Geoenvironmental Engineering 128(2), 156–165 (2002)
8. Huang, Y., Yin, Z., Hu, X., Xiong, Y.: A Novel Force Control Method for Quasi-Static Underactuated Multibody Systems. Advanced Robotics, 1–12 (2009)
9. Klein, C.A., Kittivatcharapong, S.: Optimal Force Distribution for the Legs of a Walking Machine with Friction Cone Constraints. IEEE Trans. On Robotics and Automation 6(1), 73–85 (1990)
10. Sramoon, A., Sugimoto, M.: Development of a ground reaction curve for shield tunnelling. In: Proc., Int. Symp. Geotechnical Aspects of Underground Construction in Soft Ground, Balkema, Rotterdam, The Netherlands, pp. 437–442 (1999)

Establishment of TBM Disc Cutter Dynamic Model for Vertical Vibration

Kui Zhang[1,2], Yi-min Xia[2], Qing Tan[1], Kai Wang[1], and Nian-en Yi[1]

[1] College of Mechanical and Electrical Engineering, Central South University, Changsha Hunan 410083, China
[2] Key Laboratory of Modern Complex Equipment Design and Extreme Manufacturing, Ministry of Education, Central South University, Changsha Hunan 410083, China
zhangkui414313@yahoo.com.cn, tanqing@mail.csu.edu.cn

Abstract. On the basis of reasonably simplifying the interaction process of TBM disc cutters and rock on tunneling surface, the shape of effective breaking pit is used as the lower boundary condition of vibration system and based on the equivalent stiffness method, two general differential equations of TBM disc cutter for vertical vibration are established, which are respectively suitable for hard rock, soft rock and soil. Taking an experimental disc cutter rolling granite as the example, through dynamic simulation, results show that vibration acceleration increases linearly with increased depth of the effective breaking pit and linear velocity of disc cutter and Nonlinearly decrease with increased width of the effective breaking pit; Besides satisfying the strength requirement, the selection of bearings and the design of cutter shaft should give dual attention to dynamics performance.

Keywords: TBM, disc cutter, vertical vibration, modeling, shape of effective breaking pit.

1 Preface

In tunnel construction, as a commonly used TBM (Tunnel Boring Machine) cutter, the disc cutter is the only part directly interacting with rock. The incentive energy inputted into the TBM dynamic system is almost all generated by disc cutter group in cutter head. It can be said that disc cutter's mechanical property is one of the most important characteristics to determine or affect the whole TBM dynamic performance. Once not satisfying the request, it will cause serious consequences for TBM [1-3]. The interest to conduct the quantitative analysis of TBM's dynamic performance in engineering, which stems from the demand for higher loading capacity and efficiency of breaking rock and longer lifetime of disc cutter, depends on the study of disc cutter's dynamic model to a great extent [4, 5].

But different with the traditional research, dynamic characteristics of disc cutter closely link with rock mechanics properties, rock's crack and joint, the discontinuity of geological conditions, construction parameters and so on. How to relate these factors with disc cutter is key subject of dynamic study. On the other hand, due to the

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 374–382, 2009.

complicated interaction between disc cutter and rock, it's very difficult to use a determined function to express the change law of resistance force. It might bring inconvenience to dynamic modeling for TBM cutter. It is a pity that there are still few related literatures reporting about it at home and abroad.

In order to solve the above problems, this paper reasonably simplifies the process of rolling rock with different lithology, firstly introduces the geotechnical model into the vibration analysis of disc cutter, and proposes to use the shape of effective breaking pit as the cutter's displacement excitation to respectively establish general differential equations of vertical vibration for rolling hard rock, soft rock and soil by disc cutter.

2 Introduction of Disc Cutter

As shown in Fig. 1a, a standard disc cutter consists of a cutter ring, a cutter body, a pair of tapered roller bearings (bearing for short), a cutter shaft and sealing device. Fig.1b shows an experimental disc cutter which is installed in the linear cutting test-bed at the College of Mechanical and Electrical Engineering of Central South University. It's similar to Fig. 1a, but the cutter ring and the cutter body is made into one body.

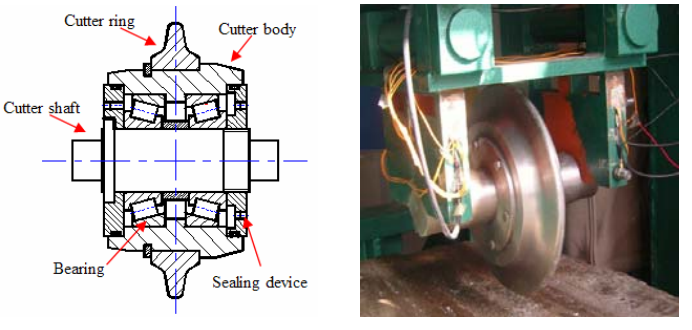


Fig. 1. Left: Assembly drawing of disc cutter; Right: experimental disc cutter in test-bed

3 Establishment of Vertical Vibration Model

3.1 Hard Rock

3.1.1 Basic Hypothesis

This article makes the following essential supposition and the simplification:

The cutter ring, the cutter body and the bearings are simplified as the rigid mass wheel (wheel for short) with equivalent spring without considering the effect of tool post. The initial rock surface is smooth, as shown in Fig. 2a. The cutter shaft is simplified to a homogeneous constant-section simply supported beam without considering the bending effect caused by side force. The amount of rock deformation under the cutter edge varies with different types of rock. When rolling hard rock, the

deformation amount is too little to be considered. Some other parts like sealing device are also not considered the effect on vibration because they don't receive any force. It is assumed that before the certain critical time t_0, it only happened the propagation of micro-crack and the formation of compacting nuclear beneath rock face and the elastic deformation of disc cutter by the normal pressure, as shown in Fig. 2b. When $t = t_0$, breaking pit instantaneously formed (Fig. 2c). In a step-crushing cycle, under the condition of small cutting depth, the cutter ring closely contact with rock face of the breaking pit because the cutter release elasticity, so the cutter rolls around the breaking pit and grinds the original contour of the pit to forms the final shape, as shown in Fig. 2d. The motion locus of the lowest point on the disc cutter namely the envelope is defined as the shape of effective breaking pit $q_e(t)$. This paper chooses it as the displacement boundary condition of disc cutter for vertical vibration.

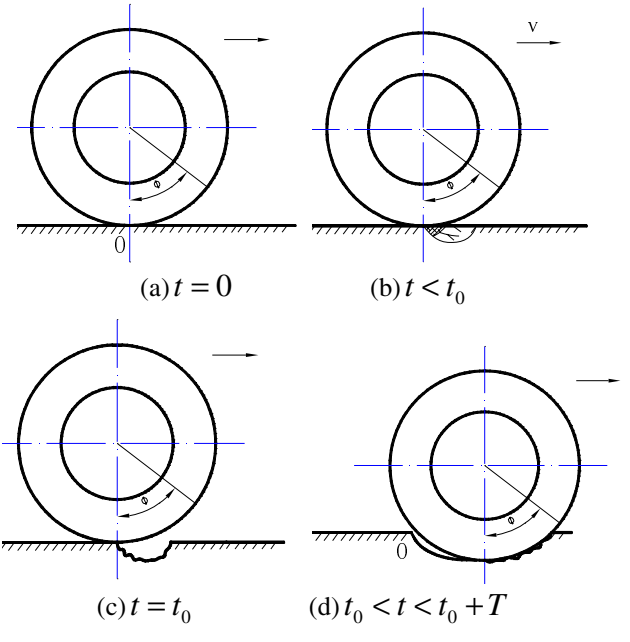


Fig. 2. Simplification process of rolling rock by disc cutter

In the above process, the breaking pit forms when cuttings flake off from the rock mass. The size and the distribution of the breaking pit have close relationship with the formation mechanism of cuttings. Literature [6] had carried spot investigation and experimental study on cuttings in tunneling engineering by different rock-breaking methods. Results show that the distribution of cuttings had similarity in certain condition which basically obeys the Rosin-Rammler distribution. Obviously the shape of the effective breaking pit also obeys some distribution, so the above method can build the relationship between rock mechanics properties and TBM tunneling performance with dynamic study.

3.1.2 Vertical Vibration Model of Rolling Hard Rock

According to Fig. 3. The dynamic equation can be build:

$$\begin{cases} m_1\ddot{z}_1 = k_p[z_2 - z_1] + c_p(\dot{z}_2 - \dot{z}_1) + k_1[q(t) - z_1] + c_1[\dot{q}(t) - \dot{z}_1] + m_1 g \\ m_2\ddot{z}_2 = -k_2 z_2 - c_2\dot{z}_2 + k_p[z_1 - z_2] + c_p(\dot{z}_1 - \dot{z}_2) + m_2 g \end{cases} \tag{1}$$

Where z_1 is the vertical displacement of the wheel, z_2 is the displacement of the cutter shaft; m_1 is the equivalent mass of wheel; m_2 is the equivalent mass of the cutter shaft relative to its centre, $m_2 = 17/35 m_{DZ}$, m_{DZ} is the mass of the shaft; k_1 is the equivalent stiffness of the wheel which is two times of the equivalent radial stiffness of one bearing k_r. Based on the equivalent stiffness method, k_r ($N \cdot \mu m^{-1}$) is expressed by the following equation [7]:

$$k_r = 7.253 l^{0.8} Z^{0.9} F_{a0}^{0.1} \cos^2\alpha / \sin^{0.1}\alpha \tag{2}$$

Where l is the length of rollers(mm); z is the number of rollers(mm); α is the contact angle(degree); F_{a0} is the axial pre tightening force(N), the minimum is determined by the following empirical equation:

$$F_{a0} = (30-40)d \tag{3}$$

Where d is the diameter of the cutter shaft (mm);

k_2 is the lateral stiffness of the cutter shaft. Base on the theory of material mechanic, k_2 can obtained by the following equation:

$$k_2 = 3E\pi R^4 / (3a^2 L - 4a^3) \tag{4}$$

Where E is the elastic modulus of the cutter shaft; R is the radius of the shaft's cross section; L is the length; a is the distance from the installation position of bearing to

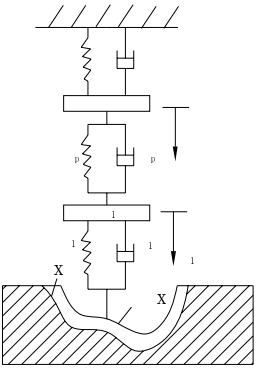


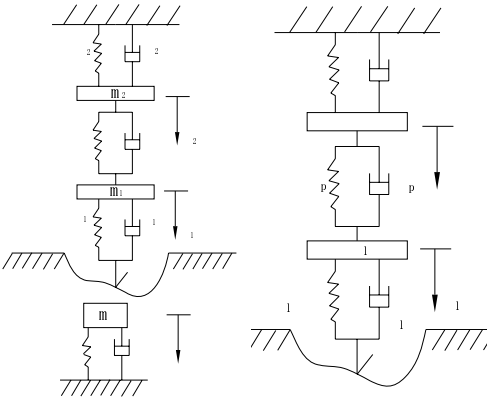
Fig. 3. Simplification model of rolling hard rock: $q_e(t)$ is the shape of effective breaking pit and $q_r(t)$ is the real shape

the fix position of shaft. k_p is the equivalent stiffness of oil film on clearance fit surface between the bearings and the cutter shaft; c_1 c_2 c_p is respectively the damping coefficient of bearings, cutter shaft and clearance fit surface; $q(t)$ is the activation function.

3.2 Soft Rock and Soil

3.2.1 Basic Hypothesis

The vibration model is shown in Fig. 4a.



(a) (b)

Fig. 4. Simplification model of rolling soft rock and soil. Left: elastoplastic body with mass; Right: elastoplastic body without mass.

In process of rolling, the soft rock and soil is similar to the hard rock, but the deformation of the former is much larger than that of the latter, thus it can not be neglected. Base on the above analyses, giving consideration to the calculation accuracy and speed, rock is simplified as elastoplastic body with mass namely Kelvin model[8] which is widely applied to simulate the deformation process under a constant stress. This model considers the mass, the stiffness and the damping coefficient of rock.

When the precision requirement is not high, elastoplastic body without mass is adopted. The vibration model is established by respectively connecting in series the stiffness and the damping coefficient of rock with that of bearings, as shown in Fig.4b.

3.2.2 Vertical Vibration Model of Rolling Soft Rock and Soil

According to Fig. 4a. The dynamic equation can be build:

$$\begin{cases} m_1\ddot{z}_1 = k_p[z_2 - z_1] + c_p(\dot{z}_2 - \dot{z}_1) + k_1[q(t) + z_3 - z_1] + c_1[\dot{q}(t) + \dot{z}_3 - \dot{z}_1] + m_1 g \\ m_2\ddot{z}_2 = -k_2 z_2 - c_2 z_2 + k_p[z_1 - z_2] + c_p(\dot{z}_1 - \dot{z}_2) + m_2 g \\ m_3\ddot{z}_3 = -k_3 z_3 - c_3\dot{z}_3 + k_1[z_1 - (q(t) + z_3)] + c_1[\dot{z}_1 - (\dot{q}(t) + \dot{z}_3)] \end{cases} \tag{5}$$

Where z_3 is the vertical displacement of deformed rock under the cutter ring; m_3 is the equivalent mass of deformed rock, is expressed by the following equation [9]:

$$m_3 = a_3 \rho (A/\pi)^{\frac{3}{2}} \tag{6}$$

Where a_3 is a parameter related to the shape of ground contact patch; A is the compacted area of ground contact path; ρ is the density of rock. k_3 is the equivalent stiffness of deformed rock;

c_3 is the equivalent damping coefficient of deformed rock, in the ranged 0.010 to 0.015($N \cdot s \cdot m^{-1}$), depended on the size of rock particles and the rock hardness.

When $q(t) = q_r(t)$, Eq.1, 5 are the dynamic equations without considering the envelope trait of disc cutter; when $q(t) = q_e(t)$, they do consider it. The disc cutter has strong trim function on the initial breaking pit, so enveloping characteristic of the cutter is too small to consider. In general case, $q_r(t)$ can replace $q_e(t)$ directly as the lower boundary of the cutter for simplicity.

4 Analysis of Example

Taking the experimental disc cutter rolling granite as the example (Fig. 1b), this paper makes a dynamic analysis on it based on the basic hypothesis mentioned in chapter 3.2.1. In addition, by neglecting the effect of clearance fit surface between the bearings and the cutter shaft, k_p , c_p is not exist, so the disc cutter can regard as a whole one with equivalent mass M , $M = m_1 + m_2$. The vibration model can be further simplified as shown in Fig. 5.

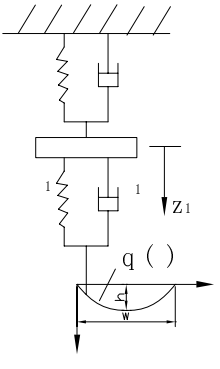


Fig. 5. Simplification model of rolling granite by the experimental disc cutter

The critical time t_0 and the initial state of the system are both set at zero. By theoretical calculation and laboratory testing, the dynamic equation can be build:

$$\begin{cases} 33.6\ddot{z}_1 + 34354\dot{z}_1 + 9.85\times10^9 z_1 = 8.2\times10^9 q_e(x) + 34354\dot{q}_e(x) + 32928 \\ z_1(0) = 0, \dot{z}_1(0) = 0 \end{cases} \tag{7}$$

Under the condition of small cutting depth, the shape of effective breaking pit can be well fitted by conic with good accuracy, so it's expressed by Eq.8.

$$q_e(x) = ax^2 + bx + c \tag{8}$$

Taking the beginning point of breaking pit as the origin o, the vertical direction as y direction and the rolling direction as x direction, the coordinate system is established(Fig. 5). The relationship is expressed by Eq.9.

$$\begin{cases} a = -h/w^2, b = 2h/w, c = 0 \\ x = vt \end{cases} \tag{9}$$

Where h is the effective cutting depth; w is the effective width of pit; v is the linear velocity of the center of disc cutter. In above coordinate system, the last form of the equation is deduced form Eq.7, 8, 9. Through numerical calculation, the dynamic response of the system can be obtained (Fig. 6).

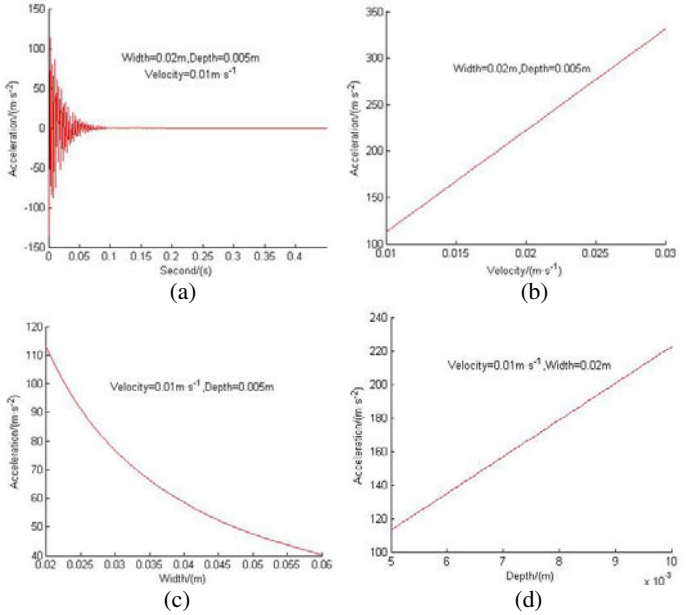


Fig. 6. Top left: Vibration acceleration vs. time; Top right: Vibration acceleration vs. linear velocity; Bottom left: Vibration acceleration vs. width; Bottom right: Vibration acceleration vs. depth

The vertical vibration is the typical damped oscillation. The amplitude of acceleration attenuates in a short time and the initial shake is large (Fig. 6a). Vibration acceleration increases linearly with the increases in depth of the effective breaking pit (Fig. 6d) and linear velocity of disc cutter (Fig. 6b) and nonlinearly decreases with the increases in width of the effective breaking pit (Fig. 6c).

Through Eq.7, the equivalent stiffness of bearing and shaft has great influence on vibration of disc cutter, so besides satisfying the strength requirement, the selection of bearings and the design of cutter shaft should to give dual attention to dynamics performance.

5 Conclusion

At the premise of small cutting depth, this paper has introduced a modeling method to establish the vertical vibration equation of disc cutter, which effectively build the relationship between rock mechanics properties and TBM tunneling performance with dynamic Study and provides a new idea to this field.

Some useful conclusions can be drawn from this paper:

1) The vertical vibration of disc cutter exacerbates with the increases in linear cutting velocity, so how to choose a suitable velocity for a certain cutter plays an important role in guaranteeing tool life. Meanwhile, the closer the disc cutter get to edge of the cutter head, the higher the demand on its dynamic performance is, relative to cutters in other installation position.
2) The shape of effective breaking pit has an influence on the vertical vibration, so the vibratory quantity varies with the different rock property. Brittle rock may cause much stronger vibration than plastic rock dose. Aiming at a certain geological condition, or say, under some complicated geological conditions, the design of the cutter should meet the dynamic requirement.
3) Besides satisfying the strength requirement, the structure of disc cutter should have good dynamic performance. This will help solve the mystery that the domestic cutters with the same strength as the imported ones have relatively shorter lifetime.

Acknowledgements

This research is supported by the National Key Basic Research and Development Program of China (973 Program, No. 2007CB714002) and Central South University Postdoctoral Sustentation Fund (743410151), which are gratefully acknowledged. The authors would like to thank Mr. XU Zhi-jun, ZHOU Xi-wen, and Ms. XUE Jing, LV Dan, for the useful suggestions during the preparation of this paper.

References

1. Yeung, A.T.: A Practical Methodology for Impact Assessment of Tunnel Boring Machine Generated Noise. Transactions Hong Kong Institution of Engineers 9(8), 26–30 (2001)
2. Polygon Media Ltd.: Sounds of Excavation. Tunnels and Tunneling International 11(36), 31–33 (2004)

3. Susumu, M., Yukio, T., Koji, I., Takahiko, O.: Influence of Vibration during TBM Excavation. In: Proceedings of the Japan Symposium on Rock Mechanics, vol. 11(2), pp. 593–598 (2001) (in Japanese)
4. Ashida, Y.: Seismic Imaging ahead of A Tunnel Face with Three-Component Geophones. Rock Mechanics and Mining Sciences 38, 823–831 (2001)
5. Flanagan, R.F.: Ground Vibration from TBMs and Shields. Tunnels and Tunnelling International 25(10), 30–33 (1993)
6. Kezhi, S., Liguang, J., Dajun, Y., Mengshu, W.: Research on Distribution Regularities of Grain Size of Rock Detritus from Discoid Cutters. Chinese Journal of Rock Mechanics and Engineering 27(1), 3016–3022 (2008) (in Chinese)
7. Wei-Ming, L.: Rigidity Calculation of Axial Position Preload Taper Roller Bearings. Bearing (5), 1–3 (2004) (in Chinese)
8. Mesquita, A.D., Coda, H.B.: A Simple Kelvin and Boltzmann Viscoelastic Analysis of Three-Dimensional Solids by the Boundary Element Method. Engineering Analysis with Boundary Elements 27(9), 885–895 (2003)
9. Jian-min, G., Ming-shu, L., Ji-hong, Y., Zhi-zhong, W.: Analysis of the Vehicle Vibration System on Soft Terrain Surface. Transactions of the CSAE 12(3), 26–31 (1996) (in Chinese)

Soft Rock Cutting Mechanics Model of TBM Cutter and Experimental Research

Jing Xue[1], Yimin Xia[1,2], Zhiyong Ji[1], and Xiwen Zhou[1]

[1] Key Laboratory of Modern Complex Equipment Design and Extreme Manufacturing, Ministry of Education, Changsha Hunan, 410083, China
[2] School of Materials Science and Engineering, Central South University, Changsha Hunan, 410083, China

Abstract. On the basis of brittle tensile fracture theory, assuming the destruction of rock was due to tensile failure, a cutting-tool loading calculation model was established, considering the compact core impact in front of cutting-tool, then the mechanism of cutter-rock, as well as the process of brittle fracture of rock were analyzed and the expressions of horizontal cutting force and vertical propulsive force for the shield cutter were obtained. At the same time, three kinds of cutters with different rake angle were designed, and the cutting experiments on a linear cutting performance testing station were carried out. By comparing, the results of experiments were basically in accordance with those of mechanical model, error rate of horizontal cutting force was below 7% and that of vertical propulsive force was in the scope of 9%. At last, the experiments proved that the force of cutter would increase along with the increasing depth, or decreasing rake angle.

Keywords: Shield machine, cutter, mechanics model, cutting experiment.

1 Introduction

Shield machines cut rock and soil in the tunnel face, during the rock-soil cutting process, with cutter which is installed on rotatory cutting head, and then the outer layer rock is stripped from the parent rock by cutter, so that it achieves the goal of excavating section[1~3]. Because the actual geological condition of tunneling tunnel is complicated and diversified, force condition of shield tools is also very complex, which easily cause abnormal wear on cutting tools, and then reduce the service life. Thus tunneling efficiency of shield machine is seriously affected, and cost of excavating tunnel is also increased. Cutter which is one of the typical shield cutting tools is mainly used to cut soft rock and soil ground. Through researched on the mechanism of cutter-rock, the law of cutters loading was analyzed, and then the mechanical model was established, which is not only benefit to analyze torque, force and penetration performance of shield machine, but also profitable to conduct the optimization design of cutter , reduce tool wear and prolong tool life[4~7].

Since 1950, many domestic and foreign scholars have carried out massive researches on the cutting mechanism which mainly focused on cutting coal, middle-hard rock and

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 383–391, 2009.

hard rock, and different rock fragmentation mechanisms were proposed then. Aiming at digging coal principle of plough, Evans established and improved the broken coal mechanism model, and studied on the cutting rock process of wedge-shaped cutter. He thought rock failure belonged to tensile failure, and the failure defective line was an arc line extended to free surface. The expressions for cutting resistance force of the shield cutter were obtained on the basis of moment balance theory. Through experiment, Roxborough F.F. confirmed that the adaptation degrees of Evans's model when cutting in the sandstone and limestone were almost the same as in coal [8~9]. Gray K.E. thought cutting force substantially was linear load which was acted on the elastic half space, and its shearing stress path with a certain angle to direction of the biggest principal stress was the logarithm spiral. When shear stress surpassed the shear strength of rock, break crack firstly developed along the path approximate to the logarithm spiral, and then turned up until free surface. The cutting broken crag mechanism model which built by Nishimatsu Y was shear model, based on the Mohr-coulomb criterion. He thought that cutting break of the rock belonged to brittle failure stave and the plastic phenomenon did not present in the cutting process[10~11]. Goodrich R et al. established the axial shear combined cutting model whose essence was that, at first, the rock was crushed into powder in front of cutter, then rock appeared fracture growth and block body break under the tensile stress. During breaking process of rock, small fragments developed into big ones. Namely, after the bulk cracks, cutting edge moved ahead, then rock in front of cutter was formed compact core after breaking and compacting[12~14]. Assumed the rock failure was tensile, the role of compact core was analyzed in this paper, and mechanical calculation model of cutter was established when the rock reached the maximum tensile stress point, based on the theory of brittle tensile failure. At last, the expressions of horizontal cutting force and vertical propulsive force for the shield cutter were concluded.

2 Mechanics Model of Cutting Soft Rock with Shield Cutter

2.1 Crushing and Exfoliation Mechanism of Rock and Force Analysis of Rock Cutting

Granular materials include three basic failure modes: flow type, shear type and tear type. Due to the complexity of rock materials, the three failure modes would occur

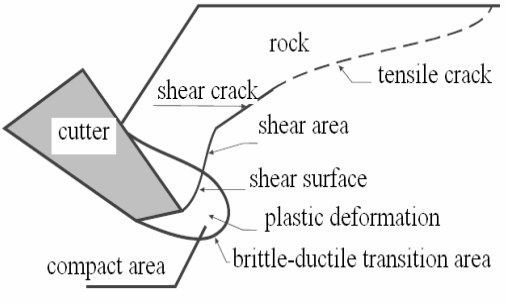


Fig. 1. Diagram of rock failure

simultaneously during the tunneling process. With the increasing cutting force, the rock in front of cutter was formed compact core. Based on the viewpoint of Fairhurst that force was transmitted by contact particles, cutting force was transmitted to rock by the discrete point load, at last it could result in the appearance of micro-crack and tensile crack. The diagram of rock failure mode by cutting is shown in Fig. 1.

According to the cutting process of shield cutter, the force analysis model of rock cutting is shown in Fig. 2. Rock shearing body was taken as research object, then front rock of cutter was considered as wedge block which the apex angle was 2ψ.Cutting force was applied to the rock through compact core.

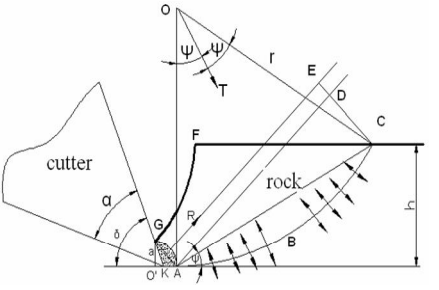


Fig. 2. Force analysis of rock cutting

2.2 Horizon Cutting Force and Vertical Propulsive Force

On the basis of cutting resistance force model of Evans wedge cutter cutting coal, when cutter of shield tunneling machine cutting soft rock, in the area of cutter compacted with soft rock, great contact stress could be generated in small volume, rock compact core was formed near front of cutter. Friction influence between cutter and rock, as well as vertical propulsive force were simultaneously considered. Assumed that:

(1) Rock fragmentation was caused by tensile stress, which complied with fragmentation theory of maximum tensile stress.
(2) The resultant force T of tensile stress in the fractured surface went through the centre of arc and bisected the camber of arc.
(3) The shape of compact core, shown as in the figure, was consisted of arc AG and straight line AK. Point O was set to the circle centre of arc AG.

According to the geometric relation (Fig. 2), resultant force which applied to center block through the compact core was:

$$R = 2\int_0^{\frac{\pi}{4}} a\sigma_c \cos\theta d\theta = \sqrt{2}a\sigma_c \tag{1}$$

Where:

σ_c soft rock uni-axial compressive strength
a radius of arc AG

When the force was large enough, soft rock was fractured along arc ABC under the tensile stress. Ultimate tension was determined by rock tensile strength, so the expression of ultimate tension could be written as:

$$T=\sigma_t\cdot r\cdot\int_{-\psi}^{\psi}\cos\beta d\beta=2\sigma_t r\sin\psi=\frac{\sigma_t h}{\sin\psi} \tag{2}$$

Where:

T ultimate tension of per width
σ_t soft rock uni-axial tensile strength
ψ crushing angle
h cutting depth

On the basis of torque balance of force R and force T to point C and the geometric relation (Fig.3), the function was obtained as follow:

$$R\left[\frac{\sqrt{2}}{2}a+\frac{h}{\sin\psi}\cos\left(\psi+\frac{\pi}{4}\right)\right]-T\frac{h}{2\sin\psi}=0 \tag{3}$$

Induced R and T into formula (3):

$$\sigma_c a^2+\frac{\sqrt{2}h\sigma_c\cos\left(\psi+\frac{\pi}{4}\right)}{\sin\psi}a-\frac{h^2\sigma_t}{2\sin^2\psi}=0 \tag{4}$$

Changed formula (4):

$$\left(\frac{a}{h}\right)^2+\frac{\sqrt{2}\cos\left(\psi+\frac{\pi}{4}\right)}{\sin\psi}\left(\frac{a}{h}\right)-\frac{1}{2\sin^2\psi}\cdot\frac{\sigma_t}{\sigma_c}=0 \tag{5}$$

Solving formula (5), because a/h was consistently more than 0, we obtained:

$$a=\frac{\sqrt{\frac{\sigma_c\cos^2\left(\frac{\pi}{4}+\psi\right)+\sigma_t}{\sigma_c}}-\cos\left(\frac{\pi}{4}+\psi\right)}{\sqrt{2}\sin\psi}h \tag{6}$$

Based on minimum energy principle, when the rock fractured, we could obtain $d\left(\frac{a}{h}\right)\Big/d\psi=0$, so the crushing angle was:

$$\psi=\frac{1}{2}\arccos\left[\frac{\sqrt{2\left(\sigma_c^4+\sigma_c^3\sigma_t\right)}-\sigma_t\sigma_c-\sigma_t^2}{\sigma_t^2+2\sigma_t\sigma_c+2\sigma_c^2}\right] \tag{7}$$

When the cutter, whose width was *b*, cut soft rock, the resultant force which applied to center block through the compact would be changed to:

$$R=\sqrt{2}ab\sigma_c \tag{8}$$

The force R was disintegrated into two forces, which were parallel and perpendicular to cutting face respectively, so we obtained:

$$R'_H = R\sin\left(\delta - 45^\circ\right) \tag{9}$$

$$R'_V = R\cos\left(\delta - 45^\circ\right) \tag{10}$$

In consideration of the practical values of R'_H and R'_V was influenced by the coefficient of friction between soft rock and cutter, based on the method of building EVANS model, we replaced $\left(\delta - 45^\circ\right)$ with $\left(\delta - 45^\circ + \phi\right)$, which could access to the practical value. So the R'_H and R'_V would be changed to:

$$R'_H = R\sin\left(\delta - 45^\circ + \phi\right) \tag{11}$$

$$R'_V = R\cos\left(\delta - 45^\circ + \phi\right) \tag{12}$$

Based on the theory of wedge-shaped body pressing into rock, the cutter must overcome the anti-stave intensity of the rock under the vertical propulsive force F_V. Therefore, when the pressed depth was h, the vertical propulsive force was:

$$F_V = 2bh\sigma_c \tan\frac{\alpha}{2} + R\cos\left(\delta - 45^\circ + \phi\right)\cos\delta + R\sin\left(\delta - 45^\circ + \phi\right)\sin\delta \tag{13}$$

Where:

α included angle

Through the analysis of the cutting tools, we obtained the expressions, according to force balance:

$$F_H - \mu N - R\cos\left(\delta - 45^\circ + \phi\right)\sin\delta + R\sin\left(\delta - 45^\circ + \phi\right)\cos\delta = 0 \tag{14}$$

$$2bh\sigma_c \tan\frac{\alpha}{2} + R\cos\left(\delta - 45^\circ + \phi\right)\cos\delta + R\sin\left(\delta - 45^\circ + \phi\right)\sin\delta - N = 0 \tag{15}$$

Where:

μ the coefficient of friction between soft rock and cutter

F_H horizontal cutting force

F_V vertical propulsive force

N support reverse force

Combined with Eq(4)~(15), then we could resolve the horizontal cutting force:

$$F_H = \frac{\sqrt{\dfrac{\sigma_c \cos^2\left(\dfrac{\pi}{4} + \psi\right) + \sigma_t}{\sigma_c}} - \cos\left(\dfrac{\pi}{4} + \psi\right)}{\sin\psi} \times bh\sigma_c \times [\cos(\delta - 45^\circ + \phi)(\mu\cos\delta + \sin\delta) + \sin(\delta - 45^\circ + \phi)(\mu\sin\delta - \cos\delta)] + 2\mu bh\sigma_c \tan\frac{\alpha}{2} \tag{16}$$

$$F_V = 2bh\sigma_c \tan\frac{\alpha}{2} + \sqrt{2}ab\sigma_c \cos\left(\delta - 45^\circ + \phi\right)\cos\delta + \sqrt{2}ab\sigma_c \sin\left(\delta - 45^\circ + \phi\right)\sin\delta \quad (17)$$

3 Experiment of Cutting Simulative Sample Rock

3.1 Experimental Equipment and Simulative Sample

This experiment was operated on the horizontal cutting test station, which was mainly composed of hydraulic pressure driving devices in the horizontal and vertical direction, installation device of cutter and control console. The experimental equipment is shown in Fig. 3. Dimensions of simulative sample were 1060×250× 144(mm), and the mechanical parameters were obtained when compressive strength was 19.14MPa, tensile strength was 2.55MPa and elastic modulus was 898.6MPa. Three cutters were designed, whose the rake angle were 10°, 15°, 20° respectively, and the width of cutters was 100mm. After each experiment, we measured the cutting depth with slide caliper.

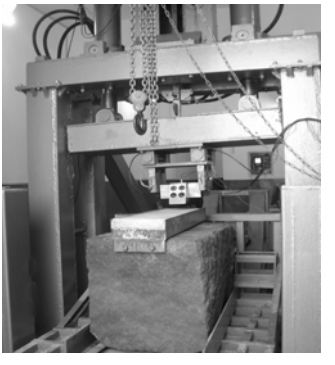


Fig. 3. Linear test bench

For testing the force directly when cutting the rock, the strain test method was used. Test system used in this experiment is shown in Fig. 4.

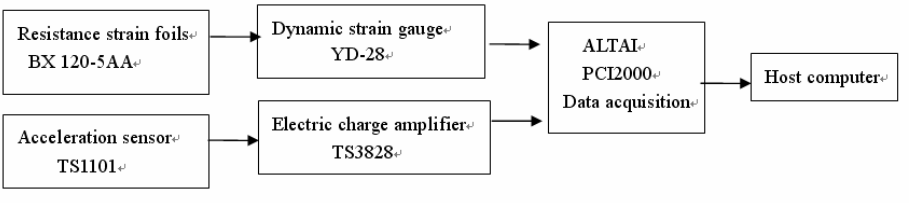
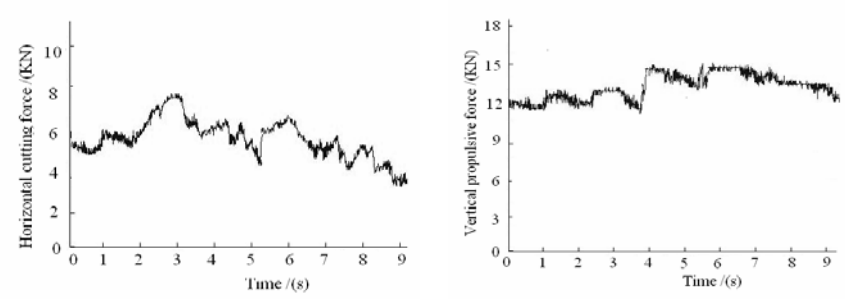


Fig. 4. Structure diagram for test system

3.2 Experimental Result and Discussion

Through the experiment, curves of cutting force along with time variation were obtained as shown in Fig. 5 when the rake angle was 15° and the cutting depth was 5.8mm.



(a) Curve of horizontal cutting force along with time variation

(b)Curve of vertical propulsive force along with time variation

Fig. 5. Curve of forces along with time variation

Experiments of three kinds of cutter with many different cutting depths, in which horizontal cutting force and vertical propulsive force were measured, were carried out. At last, results of the experiments were compared with those of mechanical model, following in Table 1.

Table 1. Comparing cutting forces of experiments with those of mechanical model in different cutting parameters and analysis of error rate

Rake angle	Cutting depth(mm)	Horizontal cutting force(KN)		Error rate	Vertical propulsive force(KN)		Error rate
		Model result	Experimental result		Model result	Experimental result	
10	4.5	4.22	3.98	6.03%	11.24	10.97	3.92%
	6.3	5.91	5.76	2.61%	17.52	16.78	4.41%
	8.6	8.44	8.11	4.19%	23.64	22.71	3.96%
	9.4	9.54	10.21	6.56%	24.71	25.46	2.95%
15	4.1	3.86	3.69	4.61%	9.04	8.45	6.98%
	5.8	5.43	5.28	2.84%	13.06	13.68	4.53%
	8.2	7.53	7.81	3.59%	16.21	17.48	7.27%
	10	9.22	9.64	4.36%	21.34	22.68	5.91%
20	2.9	2.87	2.71	5.91%	5.94	5.58	6.41%
	4.5	3.78	3.91	3.32%	9.52	9.24	3.03%
	6.6	5.76	5.94	3.04%	13.67	12.71	7.55%
	8.4	7.56	8.12	6.89%	18.25	20.04	8.93%

From the Table 1, we could see that error rate of horizontal cutting force was below 7% and that of vertical propulsive force was in the scope of 9%, through comparing the results of experiments with that of mechanical model. The results shown that the calculating results agreed well with experimental ones and correctness of mechanical model was also verified. At the same time, cutting force would increase along with increasing depth, or decreasing rake angle.

4 Conclusion

On basis of brittle tensile failure fracture theory, assumed soft rock was tensile failure, the load model of cutting soft rock was established and the expressions of horizontal cutting force and vertical propulsive force for shield cutter were obtained, after study on the cutting mechanism of shield cutter.

The cutting experiments were carried out on horizontal cutting test station. Then variation relation of cutting force with time was obtained. After comparing experimental results with calculated results of load prediction model, error rate of horizontal cutting force was below 7% and that of vertical propulsive force was in the scope of 9%. So validity and accuracy of the mechanical model were conformed.

Inference could be drawn from the mechanical model, cutting force would increase along with the increasing depth, or decreasing rake angle can be obtained.

Acknowledgments. This research is supported by the National Key Basic Research and Development Program of China (973 Program, No. 2007CB714002) and Central South University Postdoctoral Sustentation Fund (No.743410151), which are gratefully acknowledged.

References

1. Yu, Y., Xu, B., Xi, Y., et al.: Torque moment analysis of earth pressure balanced shield cutter head under soft foundation. J. Chinese Journal of Construction Machinery 2(3), 314–318 (2004) (in Chinese)
2. Guo, F., Zhao, W., Li, G.: Theoretic analysis and experimental study and soil rotary cutting resistance. J. Rock and Soil Mechanics 28(12), 2666–2670 (2007) (in Chinese)
3. Song, K., Pan, A.: Operation principle analysis of cutting tools on shield. J. Construction Machinery 2, 74–76 (2007) (in Chinese)
4. Okubo, S., Fukui, K., Chen, W.: Expert System for Applicability of Tunnel Boring Machines in Japan. J. Rock Mechanics and Rock Engineering 36(4), 305–322 (2003)
5. Sapigni, M., Berti, M., Bethaz, E., et al.: TBM Performance Estimation Using Rock Mass Classifications. J. International Journal of Rock Mechanics & Mining Sciences (39), 771–788 (2002)
6. Ahraman, S.K.: Correlation of TBM and Drilling Machine Performances with Rock Brittleness. J. Engineering Geology (65), 269–283 (2002)
7. Mroueh, H., Shahrour, I.: A Simplified 3D Model for Tunnel Construction Using Tunnel Boring Machines. J. Tunneling and Underground Space Technology (23), 38–45 (2008)
8. Evans, I.: The force required to cut coal with blunt wedges. J. International Journal of Rock Mechanics and Mining Sciences 2(1), 1–12 (1965)

9. Evans, I.: Line spacing of picks for effective cutting. J. International Journal of Rock Mechanics and Mining Sciences 9(3), 355–361 (1972)
10. Lawn, B.R., Evans, A.G.: A model for crack initiation in elastic-plastic indentation fields. J. Journal of Material Science 12(11), 2195 (1977)
11. Nishimatsu, Y.: The mechanics of rock cutting. J. International Journal of Rock Mechanics and Mining Sciences 8, 261–270 (1972)
12. Rojek, J.: Discrete Element Modeling of Rock Cutting. J. Computer Methods in Materials Science 7(2), 224–230 (2007)
13. Liu, Y., Bar-Cohen, Y., Chang, Z.: Optimal Design of Coring Bit Cutting Edge in Percussive/ Vibratory Drilling. In: J. Proceedings of International Design Engineering Technical Conferences & Computers and Information in Engineering Conference, California, vol. (56), pp. 1–9 (2005)
14. Gunes Yilmaz, N., Yurdakul, M., Goktan, R.M.: Prediction of Radial Bit Cutting Force in High-strength Rocks Using Multiple Linear Regression Analysis. J. Rock Mechanics and Mining Sciences (44), 962–970 (2007)

Neural Network Strata Identification Based on Tunneling Parameters of Shield Machine

Xiwen Zhou[1], Yimin Xia[1,2], and Jing Xue[1]

[1] Key Laboratory of Modern Complex Equipment Design and Extreme Manufacturing, Ministry of Education, Changsha Hunan, 410083, China
[2] School of Materials Science and Engineering, Central South University, Changsha Hunan, 410083, China

Abstract. A database of tunneling parameters and strata was established considering the shield tunneling practice of Guangzhou Rail Transit. Based on the data, a method of strata identification was studied by using neural network pattern recognition technology. Based on the analysis of the features of strata in shield tunneling and the data, a one-to-many mapping relation between strata and data was proposed, as well, the strata identification not only to contrast between the parameters, but also to contrast between the combined effects of each parameter mapping was pointed out. On the basis of what was mentioned above, a three-layer BP neural network model was built. What is more, some typical tunneling process parameters were input as training data, and the test results had a good agreement with practical situation. The test result could be used as strata identification. This method will enhance the scientificalness of shield tunneling control, and the timeliness and speediness of it are helpful to automatic driving for shield machine.

Keywords: Shield tunneling parameters, neural network, strata identification.

1 Introduction

Shield machines have found widespread application in tunnel construction. In the process of tunneling, a variety of geological conditions would be encountered, including mud, soft soil, soft rock and slightly weathered hard rock, which are significantly different in physical and mechanical properties, so the resulting tunneling parameters will also vary considerably. In order to tunnel with best efficiency, the tunneling parameters should always be set according to the strata with the certain mapping relation between the tunneling parameters and the strata. In turn, in accordance with the tunneling parameters, the corresponding strata can be identified by using this relationship. And also the properties of the strata are important basis to judge the abrasion of cutting tools and also to forecast tunneling distance. To identify the strata continuously in the tunneling process, the initial engineering geological investigation report can be refined, so the real-time control of the machinery tunneling can be carried on effectively. Therefore, strata identification is an important task in the mechanized tunneling process, and it is great beneficial to guide the construction.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 392–401, 2009.

Many investigators have studied the relationship between the strata and machines in the field and in the laboratory. In the 1980s, PFISTER.P., SUZUKI.Y. and GUI MW[1, 2 and 3] conducted research both on tests and theories of drilling, and this work facilitated the real-time identification of strata for rotary drilling. In domestic, the method of TBM tunneling in various types of rock was summarized by LIU Rui-qing[4], ZHAO Wen-hua[5] and LIU Hong-zhi[6], and they also explored the empirical selection criteria of tunneling parameters in different geological conditions, providing a basis for strata identification. Based on the tunneling data of Qinling Tunnel, Liu Mingyue and Zhao Weigang[7] studied the matching law of the operating parameters in different tunneling conditions by using Data Mining Techniques, providing technical support for forecasting the tunneling situation and optimizing parameters. HUANG Xiang-zhi[8] set up the extension evaluation method of tunnel surrounding rock masses stability classification according to the relation between the tunneling parameters of TBM and surrounding rock masses stability. Based on shield tunneling data at Chongqing Yangtze River tunnel, Song Kezhi[9, 10] studied the method to distinguish the surrounding rock conditions and rock cuttabilities by using field tests and fuzzy theory.

From the literature mentioned above we can see that fruitful research and exploration have been made on drilling technology, shield construction technology and data processing. All of these provide important theoretical basis for identifying the rock type in shield construction. In this paper, some typical strata were selected which are commonly encountered in Guangzhou Rail Transit, and the method of BP neural network was adopted to identify the strata. The test results were conformable with the practical situation. The timeliness and speediness of this method are helpful to real-time monitoring of strata changes.

2 Principle of Strata Identification with Neural Network

2.1 Relation between Tunneling Parameters and Strata

When tunneling in different strata, the shield machine will have different thrust, torque, driving speed, cutter-head rotation speed and penetration, due to different physical and mechanical properties of strata, such as joint sets, bedding plane and foliations. However, sometimes in order to protect cutting tools or improve tunneling efficiency, the parameters such as thrust and torque will be adjusted, which means there may be different tunneling parameters in the same strata. Therefore, the corresponding relation between the tunneling parameters and strata is not a simple causal relationship but a one-to-many mapping relation, that is to say a kind of stratum could correspond to many sets of tunneling parameters, but a set of tunneling parameters can be only corresponding to a kind of stratum, as shown in Fig. 1.

Therefore, sometimes the analysis of one or several parameters makes no sense. In order to identify the strata correctly, we must analyze the tunneling parameters comprehensively and find out the intrinsic link between the strata and the parameters.

Before starting strata identification, some priori information such as the physical parameters of the strata, as well, several typical tunneling parameters are required to be known in order to establish the corresponding database. When the tunneling

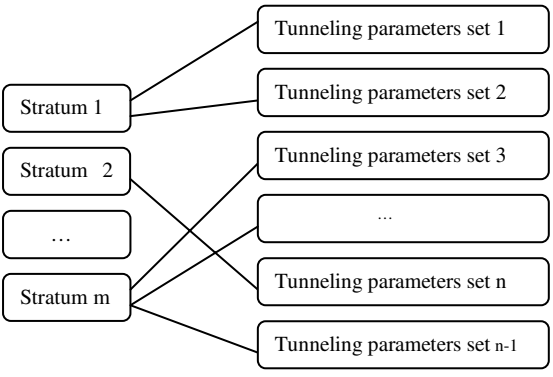


Fig. 1. Relation between tunneling parameters and strata

parameters to be identified are obtained, they should be compared with the database already established. This comparison includes not only the contrast between the parameters, but also the contrast between the combined effects of various parameters, and the latter comparison is of utmost importance. When the combined effects of test parameters and typical parameters reach a certain degree of similarity, it could be considered that this set of parameters correspond to this kind of stratum. But this identification needs to be given the experience knowledge and discriminant function before. The former is known as the database of strata and the corresponding tunneling parameters, and the discriminant function is unknown, but it is most critical in the process. Therefore, the key of strata identification based on this method is to establish such a function between the tunneling parameters and the strata.

2.2 Pattern Recognition Based on Neural Network

As mentioned above, the important part of strata identification is to establish the discriminant function which could express the relation between tunneling parameters and the strata. However, in many cases, especially for linearly inseparable and complex decision-making region, the form of discriminant function is particularly complex. Even in some case, a completely accurate function could hardly be established, and if probability model is used, the accuracy of pattern recognition would be bound to lose. In this case, we have to take other methods into consideration to realize the pattern identification.

Neural network pattern recognition is a kind of technology of self-adaption, it does not need discriminant function and knowledge experience before hand. It automatically forms the decision-making region required, mainly through the mechanism of self-learning. By training the information under different states one by one, it can obtain the mapping relation, and it also can keep learning. If the environment changes, this relation can also be adjusted adaptively[11, 12]. The structure of typical pattern recognition systems based on neural network is shown in Fig. 2.

As seen from Fig. 2**,** the diagnosis of neural network pattern recognition is mainly divided into two steps. At first, the neural network should be trained based on a certain number of sample data in order to establish inter-relation between input data

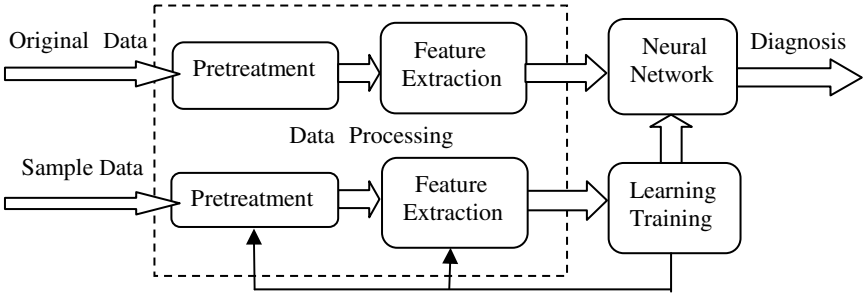


Fig. 2. The structure of typical pattern recognition systems based on neural network

and goals. After that, we can get a diagnostic network which is of a certain accuracy. And then by using this network, results can be obtained by inputting the original data for diagnosis.

Due to the complexity of shield construction process, which means geological conditions, circuit conditions and operating levels fluctuating significantly or even at random, in ambiguity, it is difficult to express it explicitly by an identified mathematical model. The neural network is ideally suited to deal with the non-linear and complex issue, with the characteristics of self-adaptive, nonlinear, and fault-tolerant capability. Therefore, using neural networks to identify the strata shield tunneling can be regarded as a viable and effective approach.

3 Analysis of Tunneling Parameters in Different Strata

The classification of shield strata mainly depends on the hardness, strength and brittleness of the rock. The greater difference in hardness between the strata is, the more types should be divided into. But at the same level, as a result of the different joint sets, bedding plane or foliations of the rock, the strata could be divided into many categories. Taking Guangzhou Rail Transit for an example, most of strata shield

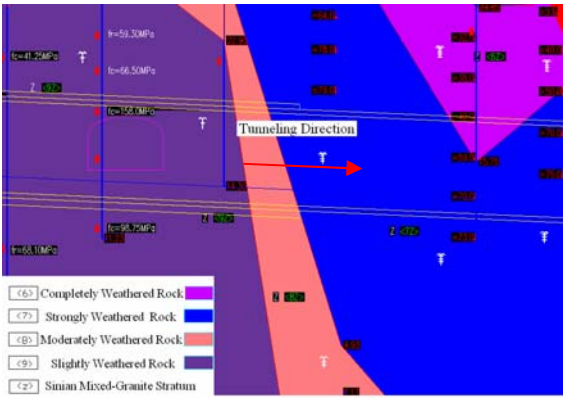


Fig. 3. The geological cross-section diagram of Guangzhou Rail Transit

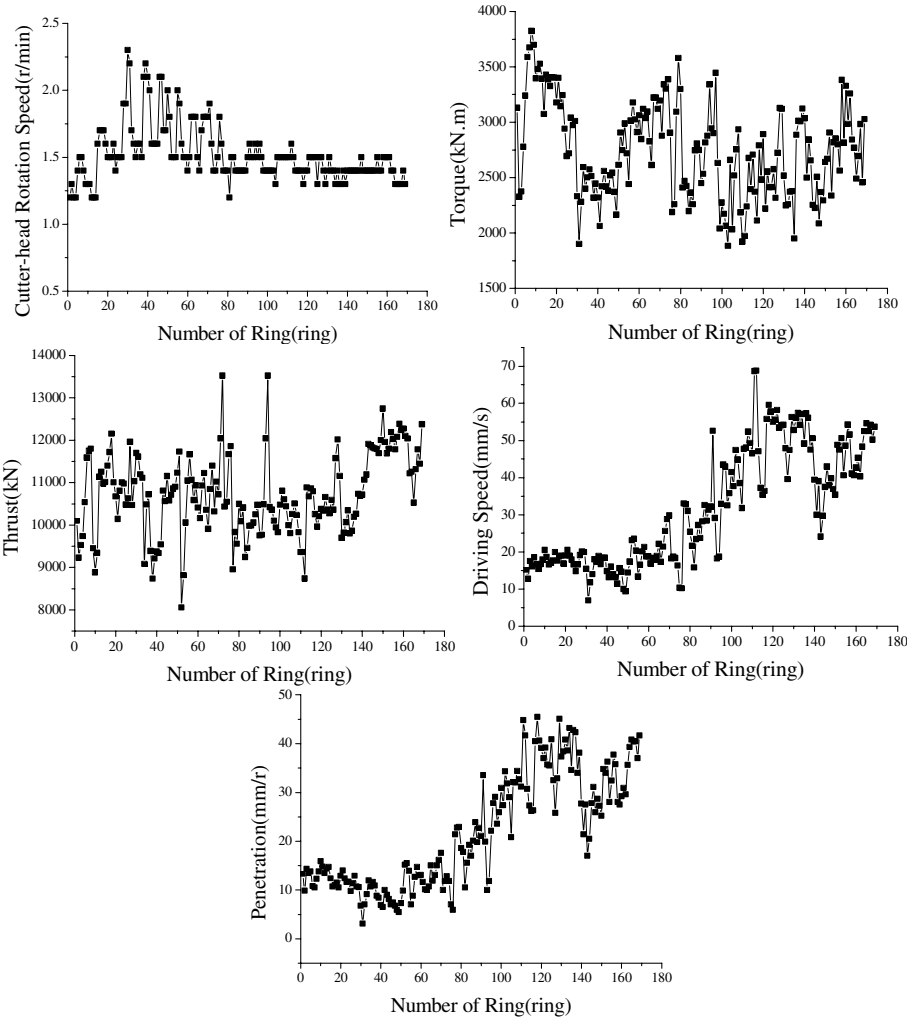


Fig. 4. The tunneling parameters in a certain period

tunneling are rocky. The geological cross-section diagram is shown in Fig. 3. Along the tunneling direction, the strata in this part are <9Z>, <8Z>, <7Z>, which are made of Sinian strata. They mainly include migmatite, mixed-granite, mixed-gneiss and quartzite, with a good integrity and high mechanical strength of fresh rock. <9Z>, <8Z>, <7Z> represent slightly weathered, moderately weathered and strongly weathered rock respectively, with the compressive strength decreasing gradually yet the weathered degree and cuttability increasing in sequence. Therefore, <9Z>, <8Z>, <7Z> can be defined as hard rock, medium hard rock and soft rock.

All of tunneling parameters are from its own data input and output system, and they can reflect the current working state of the machine directly. These parameters mainly include thrust, torque, cutter-head rotation speed, the attitude of shield machine, cabin pressure, driving speed, penetration, the oil temperature and so on. Based on the situation in this part of the strata and with the open tunneling mode of the machine, the impact of earth pressure is not taken into consideration. And also some parameters have nothing to do with the strata, so only thrust, torque, cutter-head rotation speed, driving speed and penetration are selected as the main tunneling parameters for strata identification in this paper. The tunneling parameters of some shield tunneling machine in this period are plotted in Fig. 4.

As seen from Fig. 4, the tunneling parameters change significantly in different strata and the trends in each parameter are dramatically different, but it seems that there is an obvious relation between the parameters and the strata.

4 Realization of Strata Identification Based on BP Network

4.1 Design of Input and Target Vector

In this paper, the main tunneling parameters such as the driving speed, torque, thrust, cutter-head rotation speed and penetration were selected as the input vector. As the parameters are different in unit and magnitude, the following formula was adopted for normalization before they are input into the neural network:

$$x^* = x / x_{max} \tag{1}$$

Considering differences of tunneling parameters in different strata, three typical sets of parameters were selected for each of hard rock, medium hard rock and soft rock and then the parameters were normalized. The sample of normalized tunneling parameters is shown in Table 1.

Table 1. The sample of normalized tunneling parameters

NO.	The Sample	Stratum
1	0.5217 0.8185 0.7467 0.2209 0.2923	hard rock
2	0.9565 0.4971 0.8588 0.1002 0.0681	hard rock
3	0.6956 0.6229 0.8247 0.1656 0.1538	hard rock
4	0.7826 0.8355 0.7928 0.4331 0.3868	medium hard rock
5	0.6086 0.8093 0.7271 0.4767 0.5011	medium hard rock
6	0.6086 0.5745 0.6996 0.3953 0.4219	medium hard rock
7	0.6086 0.5332 0.7268 0.5203 0.5692	soft rock
8	0.6956 0.5855 0.6461 1 0.9164	soft rock
9	0.5652 0.7298 0.7366 0.8662 1	soft rock

As there are three models in the strata, including hard rock, medium hard rock and soft rock, as the design of target vector, the form of output could be set as follows: hard rock (1,0,0); medium hard rock (0,1,0); soft rock (0,0,1).

4.2 Realization of Strata Identification Based on BP Network

BP network refers to realize the mapping function from input to output in essence. It has been proven by mathematics theory that it has function to realize any complex non-linear mapping. So it is specially suitable for solving problems with complex internal mechanism. And also it can automatically extract reasonable solving rules by learning a set of instances with correct answers. That is to say, it has a powerful self-learning ability. So it is very suitable to solve such Pattern Recognition problems due to these advantages.

So BP network is used here to identify strata, and the main steps of it include creation of network, network training and testing. In the creation of the network, because strata identification is a general pattern recognition problem, there is no need to create a very complex network, and a three layer network is enough to solve this problem. In this network, the number of neurons in hidden layer N_2 and the number in the input layer N_1 have the following relations approximately.

$$N_2=2*N_1+1 \tag{2}$$

Network training is a process of continuous weight amendment and threshold value modification, thus through continuous training, the output error of network can meet the requirements of practical application. And in this paper, the Levenberg-Marquardt algorithm is selected for training the network, and also the number of training times is set to 1000, the training goal 0.0005, and the learning rate 0.1. And then run the program, after 35 times of training and the performance of the network will reach the requirements and the result is shown in Fig. 5.

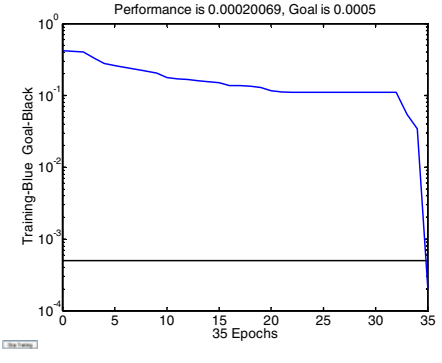


Fig. 5. The result of neural network training

Then the original data were tested by using the trained BP network, and part of the test results are shown in Table 2.

Table 2. Part of the test results

Ring NO.	Result	Stratum	Ring NO.	Result	Stratum
55	0.9958 0.0059 0.0000	hard rock	88	0.0018 0.9960 0.0001	medium hard rock
56	0.9947 0.0056 0.0000	hard rock	89	0.0002 0.9883 0.0037	medium hard rock
57	0.9749 0.0305 0.0000	hard rock	90	0.0001 0.9922 0.0080	medium hard rock
58	0.8239 0.2321 0.0000	hard rock	91	0.0000 0.0005 0.9999	soft rock
59	0.9816 0.0273 0.0000	hard rock	92	0.0010 0.9975 0.0001	medium hard rock
60	0.9865 0.0203 0.0000	hard rock	93	0.9983 0.0028 0.0000	hard rock
61	0.9881 0.0201 0.0000	hard rock	94	0.9960 0.0042 0.0000	hard rock
62	0.9844 0.0182 0.0000	hard rock	95	0.0001 0.9979 0.0009	medium hard rock
63	0.9890 0.0129 0.0000	hard rock	96	0.0000 0.0098 0.9919	soft rock
64	0.9471 0.0680 0.0000	hard rock	97	0.0000 0.1559 0.7487	soft rock
65	0.2268 0.8562 0.0000	medium hard rock	98	0.0001 0.9349 0.0454	medium hard rock
66	0.9933 0.0132 0.0000	hard rock	99	0.0000 0.0010 0.9991	soft rock
67	0.9381 0.0680 0.0000	hard rock	100	0.0000 0.0002 0.9999	soft rock
68	0.0314 0.9754 0.0000	medium hard rock	101	0.0000 0.0004 0.9997	soft rock
69	0.0054 0.9954 0.0000	medium hard rock	102	0.0000 0.0002 1.0000	soft rock
70	0.0012 0.9986 0.0000	medium hard rock	103	0.0000 0.0002 1.0000	soft rock
71	0.9879 0.0138 0.0000	hard rock	104	0.0000 0.0009 0.9978	soft rock
72	0.9960 0.0042 0.0000	hard rock	105	0.0001 0.8842 0.1770	medium hard rock
73	0.9936 0.0087 0.0000	hard rock	106	0.0000 0.0004 0.9999	soft rock
74	0.9976 0.0034 0.0000	hard rock	107	0.0000 0.0005 0.9998	soft rock
75	0.9998 0.0004 0.0000	hard rock	108	0.0000 0.0007 0.9997	soft rock
76	0.9996 0.0006 0.0000	hard rock	109	0.0000 0.0002 1.0000	soft rock
77	0.0001 0.9822 0.0500	medium hard rock	110	0.0000 0.0002 1.0000	soft rock
78	0.0001 0.9980 0.0005	medium hard rock	111	0.0000 0.0004 1.0000	soft rock
79	0.0003 0.9989 0.0000	medium hard rock	112	0.0000 0.0006 1.0000	soft rock
80	0.1287 0.8559 0.0000	medium hard rock	113	0.0000 0.0005 0.9998	soft rock
81	0.1168 0.9103 0.0000	medium hard rock	114	0.0000 0.0013 0.9988	soft rock
82	0.9922 0.0167 0.0000	medium hard rock	115	0.0000 0.0726 0.8970	soft rock
83	0.0085 0.9949 0.0000	medium hard rock	116	0.0000 0.0029 0.9970	soft rock
84	0.0007 0.9957 0.0011	medium hard rock	117	0.0000 0.0003 1.0000	soft rock
85	0.0141 0.9897 0.0000	medium hard rock	118	0.0000 0.0007 1.0000	soft rock
86	0.0006 0.9915 0.0026	medium hard rock	119	0.0000 0.0005 1.0000	soft rock
87	0.0001 0.9593 0.0254	medium hard rock	120	0.0000 0.0007 0.9999	soft rock

As seen from Table 2, the trend of change in strata is essentially from hard rock to medium hard rock and then to soft rock. Based on all the test results data, a conclusion can be drawn that ring NO.1~76 are mostly hard rock, ring NO.77~95 are mostly medium hard rock, and ring NO.96 ~ 169 are soft rock. And there is another phenomenon in the table that hard rock alternate with medium hard rock from ring 61 to ring 76, and that medium hard rock alternate with soft rock from ring 96 to ring 105. This shows that the shield machine is working at the junction of two strata during this period. The physical and mechanical properties of the strata change significantly, resulting great fluctuation of tunneling parameters. Compared with geological cross-section diagram, the result of pattern recognition based on neural network is basically in line with the geological map. That means the BP network can accurately identify the strata corresponding to the tunneling parameters.

5 Conclusion

Depending on the studies mentioned above, conclusions obtained can be listed as below:

(1) The relation between the tunneling parameters and strata is a one-to-many mapping relation, that is to say a kind of stratum could correspond to many sets of tunneling parameters, but a set of tunneling parameters can be only corresponding to a kind of stratum. Therefore, when strata are being identified according to tunneling parameters, not only the similarity between the parameters should be contrasted, but also the similarity between the combined effects of various parameters should be contrasted.

(2) Considering the shield tunneling practice in Guangzhou Rail Transit, based on the established basis database of tunneling parameters and strata, the neural network method for strata identification is described and applied. It proves that the method is feasible and effective. The key is to select comprehensive, typical and representative samples of tunneling parameters. What's more, appropriate training goals and learning rate should be set, and efficiency and speed should also be taken into account.

(3) This neural network technology for strata identification is a useful complement to geological survey. At the same time, in comparison with other methods, it has the same high accuracy and also the impressive speed of identification, with no need of any complex discriminant function. All of these provide a train of thought for real-time monitoring of strata changes. Combined with the other parameters such as the characteristics of muck, it will significantly enhance the scientificalness of shield tunneling control, and also it is of great significance to shield construction.

Acknowledgments. This research is supported by the National Key Basic Research and Development Program of China (973 Program, No. 2007CB714002) and Central South University Postdoctoral Sustentation Fund (No.743410151), which are gratefully acknowledged.

References

1. Pfister, P.: Recording drilling parameters in ground engineering. J. Journal of Ground Engineering 18(3), 16–21 (1985)
2. Suzuki, Y., Sasao, H., Nishi, K., et al.: Ground exploration system using seismic cone and rotary percussion drill. J. Journal of Architecture and Building Science (1), 180–184 (1995)
3. Gui, M.W., Soga, K., Bolton, M.D., et al.: Instrumented borehole drilling for subsurface investigation. J. Journal of Geotechnical and Geoenvironmental Engineering 128(4), 283–291 (2002)
4. Liu, R.-q.: The selection criteria and experience of tunneling parameters under different geological conditions of Tunneling. J. Construction Machinery (7), 40–42 (2000) (in Chinese)
5. Zhao, W.-h.: The selection of tunneling parameters in all types of rock of TB880E TBM. J. Railway Construction Technology (5), 16–18 (2003) (in Chinese)
6. Liu, H.-z.: TBM TB880E construction technology for weak rock tunnel. J. Construction Machinery (6), 53–54 (2003) (in Chinese)
7. Mingyue, L., Weigang, Z., Zhuping, Z.: Digging and Analysis of Implicit Relationship between Working Parameters of Tunnel Boring Machine (TBM). J. Journal Of Shijiazhuang Railway Institute 19(3), 55–59 (2006) (in Chinese)
8. Huang, X.-z., She, C.-x.: Research on methods of surrounding rock masses stability classification based on extension theory. J. Rock and Soil Mechanics 27(10), 1800–1804, 1814 (2006) (in Chinese)
9. Song, K., Sun, S., Yuan, D., et al.: Fuzzy recognition for rock cuttability based on shield driving parameters. J. Chinese Journal of Rock Mechanics and Engineering 27 (supp.1), 3196–3202 (2008) (in Chinese)
10. Song, K., Dajun, Y., Mengshu, W.: Fuzzy identification of surrounding rock conditions based on analysis of shield tunneling data. J. China Civil Engineering Journal 42(1), 107–113 (2009) (in Chinese)
11. King, D., Lyons, W.B., Flanagan, C., Lewis, E.: A multipoint optical fibre sensor system for use in process water systems based on artificial neural network pattern recognition techniques. J. Sensors and Actuators A: Physical 115(2-3), 293–302 (2004)
12. Zhao, Z., Chow, T.L., Rees, H.W., et al.: Predict soil texture distributions using an artificial neural network model. J. Computers and Electronics in Agriculture 65(1), 36–48 (2009)

Analysis of Chamber Pressure for Earth Pressure Balance Shield Machine by Discrete Numerical Model

Fuzheng Qu[1], Li Wu[2], and Wei Sun[1]

[1] School of Mechanical Engineering, Dalian University of Technology, 116024 Dalian, China
[2] School of Mechanical Engineering, Dalian Jiaotong University, 116028 Dalian, China
fzqu@dlut.edu.cn, jlwuli2004@yahoo.com.cn, sunwei@dlut.edu.cn

Abstract. Taking the shield machine tunneling in a railway section as the research object, the discrete element numerical model for analyzing the chamber pressure is established. Comparing the numerical simulation results of the observed pressure on the chamber board with the field construction data, the validity and the feasibility of the simulation method is testified. Then the pressure distribution rules on the chamber board in different opening ratios of the cutting wheel and tunnel axis depth are studied, and the mathematical model of the pressure distribution is proposed. The mapping relation between the excavation face pressure and the chamber board pressure is advanced. The results provide basis for the design and control of the earth pressure balance shield machine.

Keywords: Shield machine, Muck chamber, Pressure distribution, Discrete element method.

1 Introduction

The earth pressure balance (EPB) shield machine has been widely adopted in the soft ground condition of tunnel construction. During excavation, the earth surface settlement is controlled through the continuous support of the excavation face [1]. The chamber pressure at the excavation face should be within a field between the active and the positive soil pressure [2], or it will lead to the ground depression or heaves. The chamber supporting pressure can be controlled by the incoming and outgoing materials in the chamber, i.e., through regulation of the screw conveyor rotation and of the excavation advance rate [3].

Since the chamber pressure on the excavation face cannot be observed directly during the excavation, the pressure on the chamber board is measured instead, which is regarded as the regulation basis for the screw conveyor rotation and the advance rate. In fact, the pressure on the excavation face does not equal to the pressure on the chamber board. However, the present study results on the chamber pressure is mostly from the model test, the site test and the theoretical analysis of the shield tunneling [4]-[7], and it cannot provide the relation between the pressure on the chamber board and that on the excavation face. Therefore, the numerical method is necessary to discover the chamber pressure distribution.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 402–411, 2009.

In this paper, the discrete element method (DEM) is adopted to analyze the chamber pressure of the shield machine. The mapping relation between the excavation face pressure and the chamber board pressure is introduced, and the mathematical model of the pressure distribution on the chamber board in different opening ratios of the cutting wheel and the tunnel depth is proposed, which provide the theoretical basis for controlling the pressure balance of the muck chamber.

2 DEM Model of Chamber System of EPB

2.1 Modeling of Chamber System

According to the ground condition and structure parameters of the shield machine used in a section of railway construction, the chamber pressure analysis model is established by DEM, which is composed of the geometry model, the boundary condition and the DEM model parameters of the conditioned sands.

(1) Geometry model
The DEM model of the chamber pressure analysis is shown in Fig.1, which consists of the ground pressure system, the chamber system and the screw system. In the chamber system, the diameter is 6.28m, and the length is 0.9m; the cutting wheel is the type of board, open rate is 32%. In screw system, the screw diameter is 0.7m, the length is 5m, the pitch is 0.56m, the lean angle is 22° and the vertical distance between the chamber center and the screw inlet center is 2.5m. Due to the dimension of the cutting tools, in the ground pressure system, the distance between the excavation face and the cutting wheel is set as 0.3m.

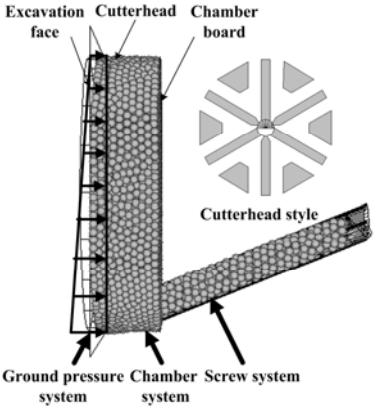


Fig. 1. DEM model of chamber pressure analysis

(2) Boundary condition
The boundary condition of the numerical model includes the ground pressure, the incoming and outgoing material volume, the cutting wheel rotary speed and the material gravity. The trapeziform ground pressure distribution at the excavation face is reproduced by the combination of the pressure servo control mechanism of the wall

element from the excavation face and the material gravity. The dynamic particles generation algorithm is adopted to simulate the incoming volume of the conditioned sands, which produces a certain numbers of particles on the excavation face randomly every a few calculation cycles. The incoming material flux is calculated by the generated particle numbers per unit time. By setting the screw rotary speed, the conditioned sands is removed continuously from the chamber. When the particles, which represent the conditioned sands, get to the screw outlet, they are deleted immediately. The outgoing material flux is calculated by the removed particle numbers from the inlet of the screw per unit time. The cutting wheel rotary speed and the material gravity can be reproduced conveniently by setting the velocity of the wall element and the material gravity acceleration.

(3) DEM model parameters of conditioned sands

The geometry model dimension should be greater than 10 times of the particle diameter [8], thus the particle diameter is selected as 0.2m. The number of particles in the analysis model is about 12000. The particle density is set to 2630 kg/m^3 according to the density of the dried sand. The constitutive model acting at a particular contact (Fig.2) consists of the elastic component (normal and tangential contact stiffness k_n, k_s), which simulates the elastic contact force in the conditioned sands; the viscous component (normal and tangential strength F_{cn}, F_{cs}), which simulates the viscous force of the foam and bentonite slurry acting to the conditioned sands; and the plastic component (friction coefficient f_u), which simulates the friction between conditioned sands particles.

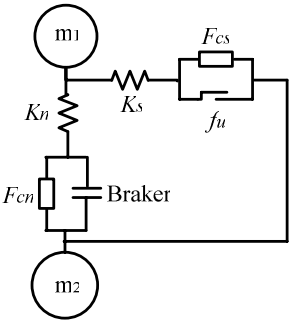


Fig. 2. Contact model of conditioned sands

For calibrating the DEM model parameters of the sample, a triaxial compression experiment is modeled (Fig.3). The diameter of the synthetic triaxial sample is 3.91 m and the height is 8 m. The top and bottom walls simulate loading platens and the lateral ones simulate the confining pressure experienced by the sides of the sample. During the simulation process, the velocity of the lateral walls are controlled automatically by a numerical servomechanism which maintains a constant confining stress within the sample.

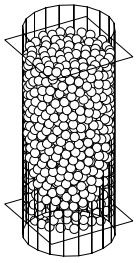


Fig. 3. DEM model of triaxial tests of conditioned sands

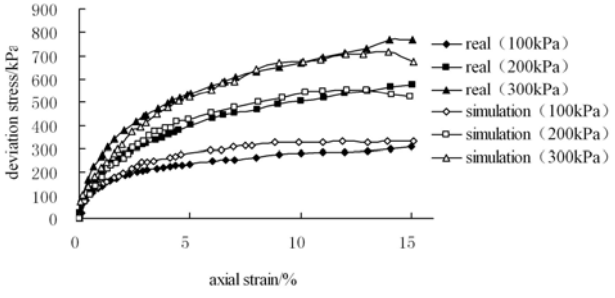


Fig. 4. Comparison between predicted curves and observed curves of triaxial test

The contact model parameters are determined by comparing the real experiment data with the numerical results, which is that the contact stiffness is 100kN/m, the bond strength is 0.3 N and the friction coefficient is 0.50. Adopting the above model parameters, the predict curves of the numerical model of the triaxial test is compared with the experiment curves (Fig.4), which shows that the simulation curves agree well with the real curves.

2.2 Validation of Simulation Results

In the analysis model, according to the field ground condition and working parameter, the excavation face pressure is set as the value of the lateral earth pressure at rest where the tunnel axis depth is 15m; the advance speed is 60mm/min; the screw rotary speed is 10 r/min; the cutting wheel rotary speed is 1 r/min. The simulation results of the pressure on the chamber board are showed in Tab.1, which is compared with the record data of construction site.

As shown in Table 1, due to the error of the adopted contact model parameters of the conditioned sands, the simulation results are smaller than the observed data on site slightly. However, the pressure simulation results of three measured points on the chamber board conform to the site record data, which illustrates that it is effective and feasible to simulate the chamber pressure distribution by DEM.

Table 1. Comparison between simulation results and site data

Measured point	I	II	III
Coordination (m)	(0.8,2.4)	(-0.23,-0.2)	(-1.46, -1.8)
Observed value (kPa)	68.0	97.0	126.0
Simulation value (kPa)	62.5	92.5	120.2
Relative error (%)	8.09	4.64	4.60

3 Analysis of Chamber Pressure Distribution

3.1 Pressure Distribution on Chamber Board

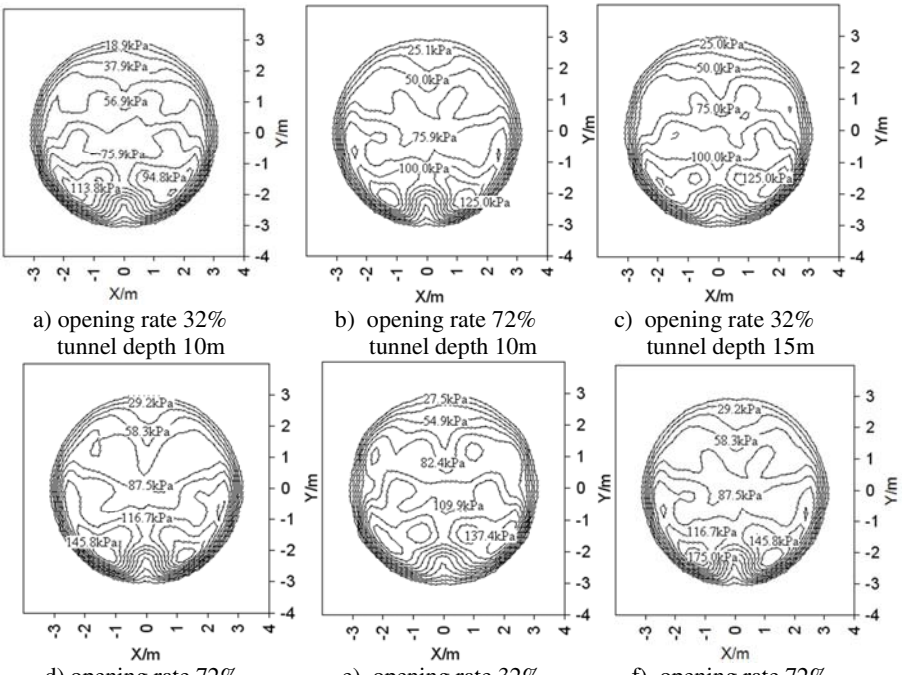
As we all know, the typical cutting wheels include the board type and the spoke type, which opening rate is often 30-50% and 60-95% respectively. The analysis model of chamber pressure is established by the board cutting wheel with an opening rate of 32% and the spoke cutting wheel with an opening rate of 72%, as well as the tunnel depth of 10m, 15m and 20m is supposed in each case. The pressure on the chamber board has been monitored during the simulation by putting a set of measurement spheres. The average pressure distribution on the chamber board is shown in Fig.5.

There are several significant features related to Fig.5.

(1) The cutting wheel structure influences the distribution of pressure contour on the chamber board. Especially when the coordinate y varies from -1m to 0m, the difference of the distribution of pressure contours between the two cutting wheel types is significant: For the board cutting wheel, the peak pressure is near to the position when the coordinate x equals 1.5m, while for the spoke cutting wheel, the peak pressure is near the position when the coordinate x equals 2.5m. In the same position of the chamber board, the pressure transfers through the cutting wheel with an opening rate of 72%, and it is slightly larger than that of 32%.

(2) In the same opening rate of cutting wheel, the tunnel depth has little influence on the shape of the pressure contour of the chamber board, but the pressure values. When the tunnel depth is 10m, 15m and 20m respectively, the average chamber board pressure transfers through the cutting wheel with the opening rate of 32% is 63.0kPa, 87.0kPa and 111.9kPa respectively, while 88.0kPa, 128.7kPa and 165.5kPa respectively with the opening rate of 72%.

(3) In all the simulation cases, the chamber board pressure reveals strong nonlinear distribution. Near to the position of the screw inlet center (0, -2.49), the average pressure gradient is 166.7kPa/m; in the area of the coordinate y between 0.8 and 1.8, the average pressure gradient is 12.5kPa/m. The area of low pressure gradient is considered as the ideal position for putting the muddlers, which mixes around the admixture of the sand, foam and bentonite et al. in the chamber with the rotation of the cutting wheel and enhances the characteristics of the plastic flow of the conditioned sands.



a) opening rate 32% tunnel depth 10m

b) opening rate 72% tunnel depth 10m

c) opening rate 32% tunnel depth 15m

d) opening rate 72% tunnel depth 15m

e) opening rate 32% tunnel depth 20m

f) opening rate 72% tunnel depth 20m

Fig. 5. Pressure distribution of chamber board

3.2 Model of Chamber Board Pressure Distribution

In order to provide basis for the chamber pressure control, the pressure distribution on the chamber board is fitted by a multinomial function. The function is described as equation (1).

$$p(x, y) = a + bx^2 + cx^4 + dy + ey^2 + fy^3 + gy^4 + hy^5 + ix^2 y + jx^2 y^2 \ . \tag{1}$$

where x, y is the coordination on the chamber board, a-j is the function coefficient. So the effective pressure on the chamber board can be described as formulation (2).

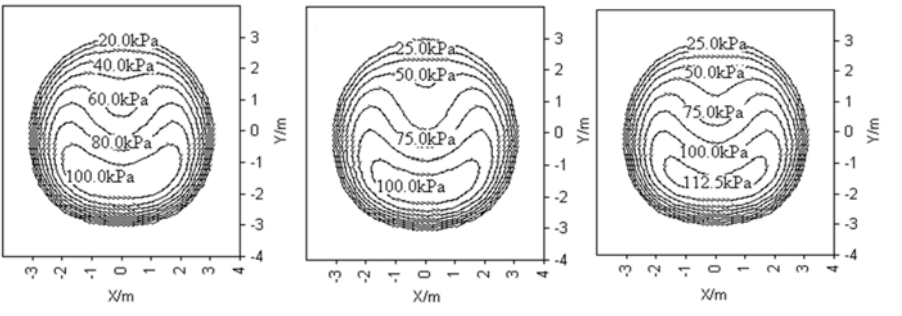
$$\overline{p}_b = \frac{1}{A}\int_A p(x, y)dA \ . \tag{2}$$

For the board and the spoke cutting wheel, the coefficient *a* - *j* and the fitting relatively coefficient R^2 of the pressure distribution function when the tunnel depth is 10m, 15m and 20m respectively, which are listed in Table 2. The contour of the fitting pressure function is shown in Fig.6.

As shown in Table 2, in the same opening rate of cutting wheel, model parameters *a*, *b*, *e*, *f*, *i* increase with the tunnel depth, while *c*, *d*, *g*, *h*, *j* decrease with the tunnel depth. The relative coefficient of the fitting equation in all cases is larger than 0.8, which illustrates that the fitting equation expression of the pressure distribution is applicable to describe the pressure distribution on the chamber board.

Table 2. Model parameters of pressure on chamber board

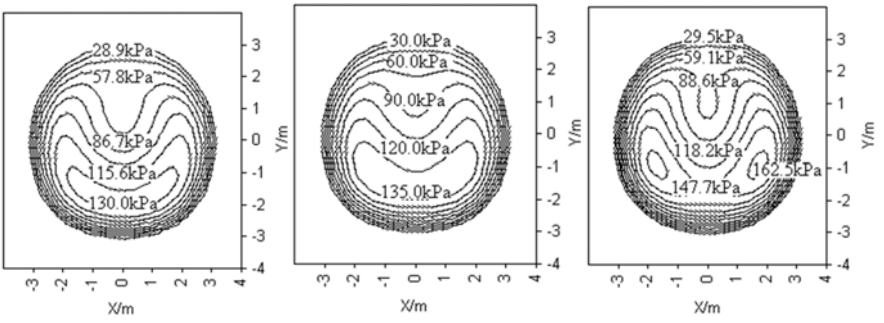
Model parameter	Opening rate 32%			Opening rate 72%		
	10m	15m	20m	10m	15m	20m
a	69.67	79.25	98.73	64.67	77.32	97.96
b	11.82	13.72	17.88	14.10	22.59	29.17
c	-1.92	-2.23	-2.86	-2.06	-3.04	-3.90
d	-18.45	-21.63	-25.11	-19.80	-24.21	-27.15
e	6.64	6.81	8.15	9.16	13.50	13.02
f	0.96	1.45	2.16	0.30	0.46	3.74
g	-1.45	-1.60	-1.96	-1.62	-2.21	-2.41
h	0.11	0.09	0.06	0.19	0.22	-0.09
i	1.15	1.34	1.68	1.56	1.96	1.50
j	-2.68	-2.94	-3.73	-3.19	-4.58	-5.47
R^2	0.87	0.88	0.89	0.83	0.84	0.86



a) opening rate 32% tunnel depth 10m

b) opening rate 72% tunnel depth 10m

c) opening rate 32% tunnel depth 15m



d) opening rate 72% tunnel depth 15m

e) opening rate 32% tunnel depth 20m

f) opening rate 72% tunnel depth 20m

Fig. 6. Fitting pressure distribution of chamber board

As shown in Fig. 5 and Fig. 6, except for the field near the screw inlet centre, the fitting equation can represent the pressure gradient distribution on the chamber board approximately.

3.3 Relation between Chamber Board Pressure and Excavation Face Pressure

For the board cutting wheel with the opening rate of 32% and 56%, and the spoke cutting wheel with the opening rate of 72% and 86%, the data of the chamber board pressure and the excavation face pressure when the tunnel depth is 10m, 15m and 20m respectively are obtained by the simulation. For each cutting wheel, the relation between the effective pressure on the chamber board and that on the excavation face is fitted respectively, which is shown in Fig.7.

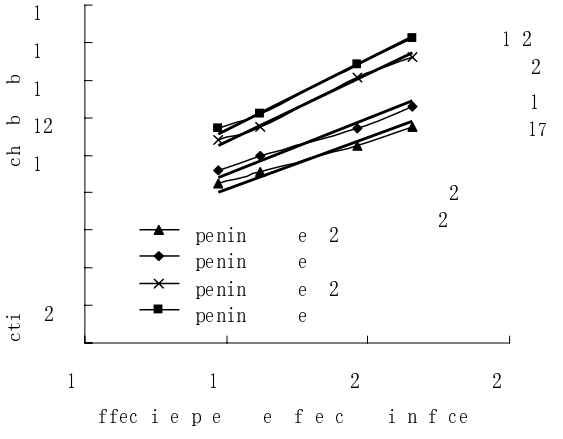


Fig. 7. Relation of chamber pressure and tunneling face pressure

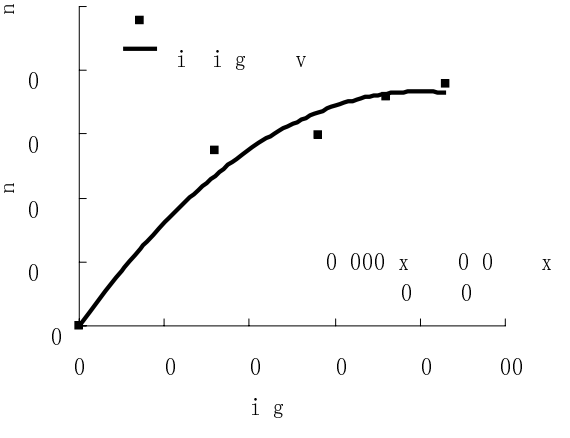


Fig. 8. Curve between pressure transfer coefficient and opening rate cutting wheel

The following conclusions can be obtained:

(1) In different opening rate of cutting wheel and tunnel depth, there are pressure differences between the excavation face and the chamber board. The effective pressure on the excavation face is larger than that on the chamber board.
(2) By regress analysis, it is discovered that for a given cutting wheel, the effective pressure on the chamber board and that on the excavation face is approximate linear proportional relation, the scale coefficient of which is the pressure transfer coefficient .
(3) The pressure transfer coefficient increases with the opening rate. Considering the limit conditions, when the opening rate equals 0, the excavation face force cannot transfer to the chamber board, so the pressure transfer coefficient is 0. The curve between the opening rate and the pressure transfer coefficient is described in Fig.8.

As shown in Fig. 8, the pressure transfer coefficient and the opening rate of cutting wheel is approximate the relation of quadratic function curve, which is presented in equation (3).

$$\alpha = 0.017 + 1.958\eta - 0.996\eta^2 \quad . \tag{3}$$

where η is the opening rate of cutting wheel.

4 Conclusions

(1) The chamber board pressure distribution shows strong nonlinear, and its distribution state is directly related to the cutting wheel structure and the position of the screw inlet. The chamber board pressure distribution can be approximately represented by a quintic polynomial function.
(2) For a given cutting wheel, the effective pressure on the chamber board and that on the excavation face is approximate linear proportional relation. The scale coefficient is the pressure transfer coefficient, which increases with the opening rate, and for a given opening rate of cutting wheel, it is independent of the tunnel depth.
(3) More simulations are required to discover the chamber pressure distribution rules under different ground earth conditions and working parameters. The present results should qualify as preliminary study for the designing and controlling of the EPB machines.

References

1. Anagnostou, G., Kovar, K.: Face stability conditions with earth-pressure-balanced shields. Tunnelling and Underground Space Technology 11(2), 165–173 (1996)
2. Yi, H.W., Sun, J.: Mechanism Analysis of Disturbance Caused by Shield Tunnelling on Soft Clays. Journal of Tongji University (Natural Science) 28(3), 277–281 (2000)
3. Hu, G.L., Gong, G.F., Yang, H.Y.: Realization of earth pressure balance for shield tunnelling machine. Journal of Zhejiang University (Engineering Science) 40(5), 874–877 (2006)

4. Adachi, T., Kimura, M., Kishida, K.: Experimental study on the distribution of earth pressure and surface settlement through three-dimensional trapdoor tests. Tunnelling and Underground Space Technology 18(2), 171–183 (2003)
5. Meguid, M.A., Saada, O., Nunes, M.A.: Physical modeling of tunnels in soft ground: A review. Tunnelling and Underground Space Technology 23(2), 185–198 (2008)
6. Xu, Q.W., Zhu, H.H., Liao, S.M.: Model experimental study on stratum adaptability of tunnel excavation with EPB shield machine in sandy stratum. Chinese Journal of Rock Mechanics and Engineering 25(z1), 2902–2909 (2006)
7. Wang, H.X., Fu, D.M.: A mathematical model and the related parameters for EPB shield tunneling. China Civil Engineering Journal 9(31), 86–90 (2006)
8. Manuel, J.M.M., Luis, E.M.R.: Discrete numerical model for analysis of earth pressure balance tunnel excavation. Journal of Geotechnical and Geoenvironmental Engineering 131(10), 1234–1242 (2005)

Dynamic Mechanism and Key Rectification Techniques of Shield Machine in the Vertical Plane*

Ming Yue, Jian Wei, Wei Sun, and Zhenggang Guo

School of Mechanical and Engineering, Dalian University of Technology,
Dalian, Liaoning, China
yueminghit@163.com

Abstract. With the development of the urban construction, the shield machine becomes more and more important as the main equipment to extend the underground space for the human being. In this paper, we investigate the dynamic mechanism of the shield machine in the vertical plane. Through this analysis, the basic requirement maintained balance of the shield machine while excavating is given. Meanwhile, the key rectification techniques about the deviation from the desired trajectory in vertical plane are discussed. This study is developed based on the engineering problems from which are frequently met in construction, and we hope to provide useful guidance to the actual excavating process.

Keywords: Shield machine, Rectification, Dynamic mechanism, Vertical plane.

1 Introduction

Rapid growth in urban development has promoted the development and the use of underground space. Tunnel is a main underground structure and is a chief method to extend the underground space for the human being. It is widely used for transportation, water passage, and other purposes such as electricity or communication cable installations. Shield machine becomes one of the most popular methods used in the construction of urban tunnels, such as rapid transit systems and large diameter underground pipelines, in order to reduce the impact on traffic [1, 2].

In literature [3, 4], the shield machine is considered as a kind of underground robot. Different from the other common robots, the shield machine suffers from the changeful and complex underground circumstance, which would lead to the complicated forces acting on the shield body. It is difficult to measure and evaluate the forces around the shield machine because of the complexity of the geological structure. Considering the different geological-tectonic conditions, two types of the tunneling equipments are mostly used, which are Tunnel Boring Machine (TBM) [5-7] and Earth Pressure Balance Shield (EPB) [8-10]. Generally, Tunnel Boring Machine is applied to the rock circumstance and Earth Pressure Balance Shield is used for clay or sandy situation [11, 12].

The construction techniques of Earth Pressure Balance Shield are very difficult for the underground constructors because of the immeasurable geological structures. In such a diverse driving environment, the deviation phenomena often take place. The

* This research was supported by grants from the Chinese National Basic Research Program (973 Program) (No. 2007CB714006).

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 412–422, 2009.

deviation between the actual trajectory and desired one would lead to unpredictable results, including the settlement of ground surface and enormous economic loss. Therefore, these deviation phenomena must be avoided as much as possible [13, 14].

The deviations can be divided into two kinds according to different planes: in horizontal plane and in vertical plane. Comparatively speaking, the deviation in vertical plane is very different to control because that the complex forces around the shield body have no symmetrical characters [15]. For an operator of the shield machine, this deviation rectification process needs abundant construction experiments. For example, when there is a depression angle, the operator should increase the force of the lower hydraulic cylinders. There is no reference for them and it just relying on the engineering experience. Hence, both the absence of experience and the inattention will result in a great failure in the process of the excavation [16]. Considering these disadvantages, in this paper, we investigate the dynamics mechanism of the shield machine, and derive the system balance equation, in order to apply a reference forces value to the operator and avoid deviation phenomena ultimately. Moreover, some techniques for rectification are also developed. We hope that these studies could give meaningful guidance to the actual construction of the shield machine.

This paper organizes as follows: Section 2 illustrates the elementary theory of the deviation rectification; Section 3 illustrates the load model of the shield machine; Section 4 illustrates the dynamic model of the shield.

2 Elementary Theory of the Deviation Rectification

The shield machine can bend with the help of the articulated device, which makes it possible to excavate in a curved path and also makes it easy for the shield machine to realize the self-rectification (as shown in Fig. 1). However, the rectification capacity of the shield machine is restricted by its structure and articulated device.

2.1 Basic Conception of the Rectification

(1)The tendency of the shield machine

The concept of the tendency used in shield machine refers to the value of the deviation from designed tunnel alignment (DTA) while the machine excavates ahead

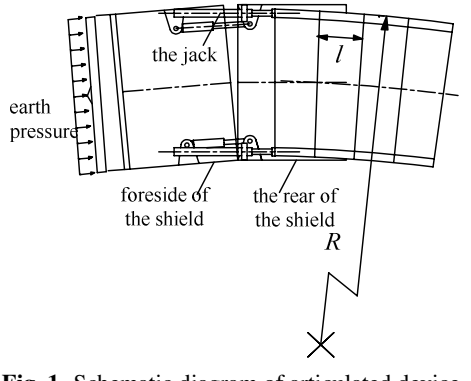


Fig. 1. Schematic diagram of articulated device

about 1m. Here, the measurable unit can be illustrated by mm/m. Let χ express the curve tendency of the shield machine, and it can be derived from:

$$\chi=(1/R)\times 1000(\text{mm/m}) \tag{1}$$

Where L is the excavating length of shield machine; R is the curve radius of shield running track.

If the length of a single segment is l and the average rotation angle of per segment is a rad ($1000a$mm/m), rectification capacity of each segment can be calculated as follows:

$$\lambda=1000a-\chi l \tag{2}$$

According to the *Eq.*(2), it can be concluded that when the shield running track is a line segment, λ is equal to $1000a$mm/m (because that $\chi \to 0$ as $R \to \infty$).

(2) Rectification capacity of the shield machine

The concept of tendency deviation refers to the difference between the tendency to realize the rectification and the tendency which is needed in the process of excavating.

If allowable tendency deviation is X(mm/m); rectification capacity of the shield machine is λ(mm/m); the excavating length of shield machine is L(m), and then, there will exist limiting conditions as follows:

$$0.5XL \leq 50 \tag{3}$$

Where the allowable excavating error is ±50mm; $0.5X$ is the average excavating tendency.

$$X=\lambda L \tag{4}$$

Where L=the excavating length of the shield machine with the tendency decreasing from X to 0.

From the two above *Eq.*(3) and *Eq.*(4), the following formula can be obtained: $X \leq 10\sqrt{\lambda}$. By substitution of the value of λ, the allowable tendency deviation X and the excavating length of shield machine L can be obtained by calculation.

If rectification capacity of each segment is λ and the needed excavating length is L, the maximum rectification angle is (X/1000) = ($10\sqrt{\lambda}$ /1000) rad. Therefore, if the deviation angle exceeds this value, it will be impossible to rectify the deviation.

2.2 The Realization of Rectification

The attitude deviation of the shield machine in vertical plane may occur during excavating because of construction error or some other reasons.

The deviation angle is assumed as θ_0. From the analysis above, it can be obtained that each segment has a given rectification capacity. The rectification capacity of the shield machine can be calculated from the proposed rectification capacity of the segment, which is denoted as λ (where $\theta_0<10\sqrt{\lambda}$ /1000).

If $\theta_0<\lambda$, rectification can be completed when the shield machine moves for a distance of one segment; if $\theta_0>\lambda$, rectification should be carried out step by step. The rectification angle of a single step is assumed as θ ($\theta<\lambda$). n ($n \approx \theta_0/\theta$) steps should be taken to realize the rectification .

3 Load Model of Shield Machine

To establish general equation of space motion for shield machine, the basic assumption is as followed:

The shield is considered as a rigid body with a weight of m, a length of L and a radius of R. The load model of shield machine is established according to requirement [1, 17], as shown in Fig. 2.

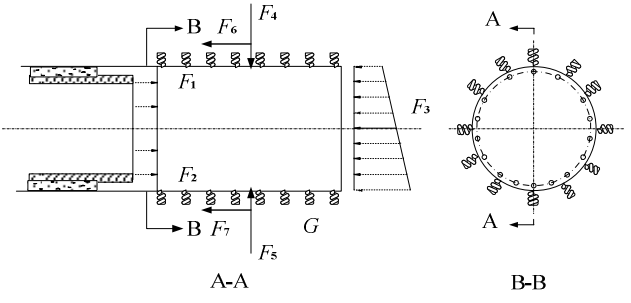


Fig. 2. Load model of shield machine

Coordinate systems of shield machine $Oxyz$ is established as shown in Fig. 3. The point O is assumed to be situated at the centre of shield tail.

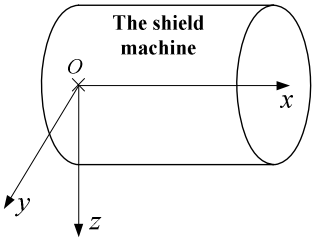


Fig. 3. Schematic diagram of coordinate systems

This paper is chiefly concerned with attitude deviation of shield machine in vertical plane. Therefore this paper mainly analyzes the forces which will affect the pitch angle of the shield, and these forces can be summarized as follows: the self weight of shield; the thrust of hydraulic cylinders (control input); the earth pressure acting on the shield face; the earth pressure acting on the upside and downside of shield; the friction force opposite to the forward direction of the shield. Since there is a close relationship between the ground displacement and the earth pressure acting on the shield periphery, so the earth pressure mentioned above can be calculated from the ground displacement. For the concrete method, please refer to reference 1.

3.1 Self Weight of Shield

The weight of cutter head and sealed cabin is large by contrast with other parts of shield. Thus, the center of gravity of the whole shield will be located on the anterior part of the body and its concrete position can be measured by the pressure sensor. The following were assumed:

The gravity of the shield is expressed as G with its direction vertically downward, and the x coordinate of the center of gravity is expressed as l_G. Side view of the shield is shown in Fig. 4.

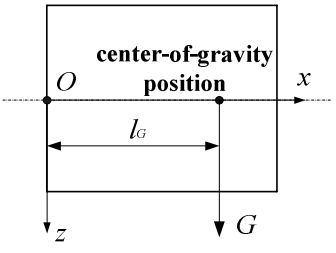


Fig. 4. Schematic diagram of shield's center of gravity

3.2 Thrust Forces of Hydraulic Cylinders

During the designing and manufacturing of shield, hydraulic cylinders are divided into four regions in order to control and correct the tunneling direction [18]. Assume that the promoting hydraulic cylinders are divided as shown in Fig. 5.

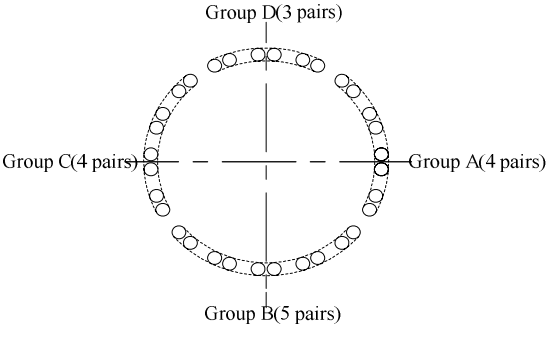


Fig. 5. Division of the promoting hydraulic cylinders

This paper is chiefly concerned with attitude deviation of shield machine in vertical plane (that is the pitch angle of the shield θ). By comparison, the forces from group A and group C have a little influence on the pitch angle. Therefore, the impact from group A and group C will not be considered in this case. What follows is the analysis of force from hydraulic cylinders of group B and group D.

The thrust forces of three pairs of hydraulic cylinders from group D can be equivalent to a resultant force F_1 (F will be used to express the equivalent force in the following explanation), which is in the x positive direction with coordinate of action point (0, 0, -h_1).

The thrust forces of hydraulic cylinders from group B can be equivalent to a resultant force F_2, which is in the x positive direction with coordinate of action point (0, 0, h_2).

3.3 Earth Pressure on the Shield Face

There is uneven distribution of earth pressure on the shield face. Earth pressure is assumed to be trapezoid distributed with the earth pressure of upper bottom p_1 and the earth pressure of lower bottom p_2(as shown in Fig. 6).The resultant total earth pressure, expressed as F_3, can be obtained by integration of earth pressure on the shield face. F_3 is in the x opposite direction and its coordinate of action point is (L, 0, h_3), as shown in Fig. 6. The force F_3 and position h_3 can be calculated as follows:

$$\begin{cases} F_3 = \dfrac{1}{2}\pi R^2 (p_1 + p_2) \\ h_3 = \dfrac{(p_2 - p_1)}{4(p_1 + p_2)} R \end{cases} \tag{5}$$

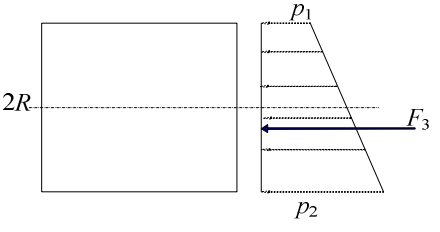


Fig. 6. The distribution of earth pressure on the shield face

3.4 Earth Pressure on the Upside and Downside

Besides on the shield face, there is earth pressure acting on the shield periphery during excavating, which can be divided into two parts: the upper and lower halves of shield periphery. The equivalent earth pressure per unit length acting on the upper half is expressed as σ_1. The earth pressure acting on the upper half can be equivalent to a resultant force F_4, which is vertical-down to the axis of shield. And the coordinate of action point is (h_4, 0, -R), as shown in Fig.7. The force F_4 and position h_4 can be represented by:

$$\begin{cases} F_4 = \sigma_1 L \\ h_4 = \dfrac{1}{2} L \end{cases} \tag{6}$$

The earth pressure per unit length acting on the lower half is expressed as σ_2. The earth pressure acting on the lower half can be equivalent to a resultant force F_5, which is vertical-up to the axis of shield with coordinate of action point (h_5, 0, R), as shown in Fig. 7. The force F_5 and position h_5 can be calculated as follows:

$$\begin{cases} F_5 = \sigma_2 L \\ h_5 = \dfrac{1}{2} L \end{cases} \tag{7}$$

3.5 The Friction Force Opposite to the Forward Direction

The friction force occurs along the opposite direction, which is due to the existence of the earth pressure on the upside and downside of shield. There is a significant linear correlation between the friction force and this earth pressure, as illustrated in Fig. 7.

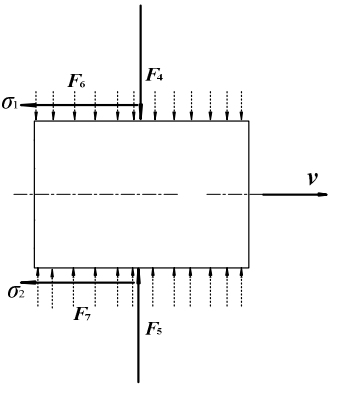


Fig. 7. The friction force due to F_4, F_5

The equivalent friction coefficient is assumed as μ. The friction force acting on the upper half can be equivalent to a resultant force F_6, which is in the opposite direction to the forward direction of the shield. The coordinate of action point is (L/2, 0, -h_6). F_6 and h_6 can be obtained from:

$$\begin{cases} F_6 = \mu F_4 \\ h_6 = R \end{cases} \tag{8}$$

The friction force acting on the lower half can be equivalent to a resultant force F_7, which has the opposite direction with the forward direction of the shield with coordinate of action point (L/2, 0, h_7). F_7 and h_7 can be obtained from:

$$\begin{cases} F_7 = \mu F_5 \\ h_7 = R \end{cases} \tag{9}$$

4 The Dynamic Model of the Shield

The dynamic model of the shield is divided into two forms as follows according to the conditions with or without a pitching angle.

4.1 The Mechanical Relationship Deduced from the Balance

This paper puts emphasis on the study of pitching angle, in order to avoid the attitude deviation of shield in vertical plane, that is, to avoid the rotation about the *y* axis.

The moment M about the *y* axis, which caused by the forces discussed above, will be introduced. It is stipulated that seeing from the *y* positive direction, clockwise direction is the positive direction. M can be obtained from:

$$\Sigma M=Gl_G+F_1h_1-F_2h_2+F_3h_3+F_4h_4-F_5h_5-F_6h_6+F_7h_7 \tag{10}$$

It is expected that there is no rotation about the *y* axis that is $\Sigma M=0$, from which the relationship of the forces can be obtained deduced from the balance. The condition that there is no attitude deviation of shield in vertical plane can be obtained from:

$$F_1h_1-F_2h_2=Gl_G+F_3h_3+F_4h_4-F_5h_5-F_6h_6+F_7h_7 \tag{11}$$

4.2 Mechanical Model of the Shield for Rectification

The relationship of the forces, when the shield is in the state of balance in vertical plane, is as shown in *Eq.*(11). However during excavating, the shield can not move forward in accordance with the design direction, which is due to many different factors. So it is most likely that there will be the attitude deviation.

Due to the structure of shield, its center of gravity is located on the anterior part of the body, which will likely lead to shield tilt, especially in soft soil area.

The following analyze the changes of forces and the mechanical model when the attitude deviation of shield occurs in vertical plane. The pitching angle of the shield is expressed as θ which can be measured by Laser Guiding System. As is known, θ is very little. So the influence on action points of the following forces is ignored: self weight of shield, the thrust of hydraulic cylinder and earth pressure on the shield face.

(1) Earth pressure on the upside and downside

It is assumed that there is attitude deviation of shield in vertical plane and the depression angle is θ.

Depression angle θ has a little influence on the earth pressure acting on the upper half which is viewed as unchanged.

Due to the depression angle θ, it is assumed that an elastic force $KL\theta$ is added to the earth pressure acting on the lower half. The earth pressure acting on the lower half can be equivalent to a resultant force F_5', which is vertical-up to the axis of shield. The distance between the action line of F_5' and *y* axis is expressed as h_5', as shown in Fig. 8. The force F_5' and position h_5' can be calculated as follows:

$$\begin{cases} F_5' = \sigma_2 L + KL\theta \\ h_5' = \dfrac{1}{2}L \end{cases} \tag{12}$$

The shield moves forward with very low driving speed and can be viewed as quasi-static model. Therefore, the viscous damping caused by the elastic force will be left out from consideration.

(2) The friction force opposite to the forward direction

The earth pressure on downside of shield changes, therefore the friction force due to this earth pressure should change accordingly, as shown in Fig. 8.

The equivalent friction coefficient is assumed as μ. The friction force acting on the lower half can be equivalent to a resultant force F_7', which is in the opposite direction to the forward direction of the shield. The distance between the action line of F_7' and y axis is expressed as h_7'. The force F_7' and position h_7' can be formulated as follows:

$$\begin{cases} F_7' = \mu F_5' \\ h_7' = R \end{cases} \tag{13}$$

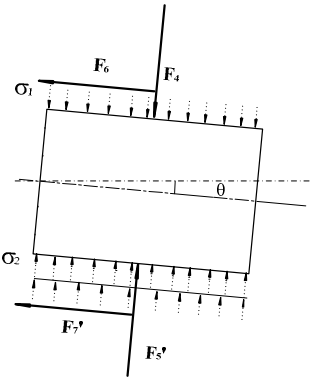


Fig. 8. The friction force due to F_4, F_5'

The inertia of the shield about the y axis is expressed as J_Y. We take clockwise direction as the positive direction. According to differential equation of rotation of a rigid body about a fixed axis (The product of the inertia of a rigid body about a fixed axis and angular acceleration is equal to the algebraic sum of the moments of main powers acting on the rigid body), the following equation can be obtained:

$$J_Y\ddot{\theta} = Gl_G + F_1h_1 - F_2h_2 + F_3h_3 + F_4h_4 - F_5'h_5' - F_6h_6 + F_7'h_7' \tag{14}$$

By substitution of values of the forces, the following equation can be obtained:

$$J_Y\ddot{\theta} + KL\theta h_5' = Gl_G + F_1h_1 - F_2h_2 + \pi R^2(p_1+p_2)h_3/2 + \sigma_1 Lh_4 - \sigma_2 Lh_5' - \mu\sigma_1 Lh_6 + \mu(\sigma_2 L + KL\theta)h_7' \tag{15}$$

The *Eq.*(15) presents the relationship between pitching angle θ, angular acceleration $\ddot{\theta}$ and the forces acting on the rigid body. According to *Eq.*(15), the attitude deviation of shield in vertical plane can be rectified in a single step by the adjustment of power output of hydraulic cylinders (f_1 f_2) or other factors.

5 Conclusions

Due to the complex rock and soil mechanism around the shield machine, the deviation from the desired trajectory will occur frequently. Because of the influence of the shield gravity, the deviation in vertical plane becomes easier to occur than the case in horizontal one. Based on these actual problems, this paper developed research work in order to propose some guidance to the practical construction.

Through the substitution method, the equation maintained the balance of the shield machine has been derived. The deviation phenomena will be avoided as much as possible by regulating the input forces of the shield machine according to the proposed equation. Meanwhile, this paper also gave the key rectification techniques. Considering the practical engineering situation, we divided the process of rectification into a series of steps. In each step, the dynamic mechanism has been analyzed. Further work will investigate the dynamic mechanism of all postures of the shield machine, and will discuss their rectification techniques in the following studies.

Acknowledgments

We thank the Chinese National Basic Research Program (973 Program) (No. 2007CB714006) for the financial support of this investigation.

References

1. Sugimoto, M., Sramoon, A.: Theoretical Model of Shield Behavior During Excavation I: Theory. Journal of Geotechnical and Geoenvironmental Engineering 128(2), 138–155 (2002)
2. Sramoon, A., Sugimoto, M., Kouji, K.: Theoretical model of shield behavior during excavation II: Application. Journal of Geotechnical and Geoenvironmental Engineering 128(2), 156–165 (2002)
3. Roozbeh, K.: Advanced robotics in civil engineering and construction. In: Fifth International Conference on Advanced Robotics, Pisa, Italy, pp. 375–378 (1991)
4. Huang, Y.A., Xiong, Y.L.: Force Planning for Underground Articulated Robot. In: Proceedings of the First International Conference on Intelligent Robotics and Applications, pp. 64–74. Springer, Heidelberg (2008)
5. Song, Y., Tang, Y., Mu, G.H., Wang, G.: Monitoring of tunnel clearance displacements under condition of TBM construction. Chinese Journal of Rock Mechanics and Engineering 28(3), 621–627 (2009)
6. Gong, Q.M., Zhao, J., Zhang, X.H.: Performance prediction of hard rock TBM tunneling. Chinese Journal of Rock Mechanics and Engineering. 23(2), 4709–4714 (2004)
7. Saffet, Y.: Utilizing rock mass properties for predicting TBM performance in hard rock condition. Tunnelling and Underground Space Technology 23(3), 326–339 (2008)
8. Suchatvee, S., Einstein, H.H.: Artificial neural networks for predicting the maximum surface settlement caused by EPB shield tunneling. Tunnelling and Underground Space Technology 21(2), 133–150 (2006)
9. Sadaghiani, M.H., Aliari, Y.: Effect of EPB tunnelling on surface settlement. In: Annual Conference of the Canadian Society for Civil Engineering 2007, Yellowknife, NT, United states, pp. 1003–1012 (2007)

10. Vinai, R., Peila, D., Oggeri, C., Pelizza, S.: Laboratory tests for EPB tunnelling soil conditioning. In: 33rd ITA-AITES World Tunnel Congress - Underground Space - The 4th Dimension of Metropolises, Prague, Czech republic, pp. 273–278 (2007)
11. Farsakh, A., Murad, Y., Voyiadjis, G.Z.: Computational model for the simulation of the shield tunneling process in cohesive soils. International Journal for Numerical and Analytical Methods in Geomechanics 23(1), 23–44 (1999)
12. Wayne, C.G., Eric, L.: EPB shield tunneling in mixed face conditions. Journal of geotechnical engineering 119(10), 1640–1656 (1993)
13. Dieter, H., Lutz, T.: North-South Urban Light Railway Cologne - A challenge for planning and construction - Technical solutions for the accomplishment of the upcoming tasks concerning the tunnels. In: 33rd ITA-AITES World Tunnel Congress - Underground Space - The 4th Dimension of Metropolises, Prague, Czech republic, pp. 1161–1167 (2007)
14. Shen, J.Q., Jin, X.L., Wang, J.Y.: Three dimensional numerical simulation of shield tunneling through flood levee based on parallel computing. Journal of Shanghai Jiaotong University 42(8), 1296–1300 (2008)
15. Mo, H.H., Chen, J.S.: Study on inner force and dislocation of segments caused by shield machine attitude. Tunnelling and Underground Space Technology 23(3), 281–291 (2008)
16. Shen, S.L., Cai, F.X., Bai, T.H., Zhu, J.M.: Direction deviation control during Double-O-Tube (DOT) tunnel construction. In: The GeoShanghai Conference on Underground Construction and Ground Movement, Shanghai, China, pp. 185–192 (2006)
17. Mitsutaka, S., Aphichat, S., Shinji, K., Yutaka, S.: Simulation of shield tunneling behavior along a curved alignment in a multilayered ground. Journal of Geotechnical and Geoenvironmental Engineering 133(6), 684–694 (2007)
18. Hu, G.L., Gong, G.F., Yang, H.: Thrust hydraulic system of shield tunnel boring machine with pressure and flow compound control. Chinese Journal of Mechanical Engineering 42(6), 124–127 (2006)

A Precise Measurement Method of Azimuth Angle for TBM

Xinbao Zhang, Shuguang Liu, and Bin Zhao

Mechanical School of Science and Engineering,
Huazhong University of Science and Technology
zhangxinbao1@mail.hust.edu.cn

Abstract. Based on the non-diffracting beam and the phase grating interferometer, a precise measurement method of azimuth angle for TBM has been developed in this paper. A high positioning precision is got on the image of the non-diffracting beam and its axis stability and a circular phase grating interferometer is used for angle measurement, so this measurement system can achieve that the measurement uncertainty of horizontal azimuth angle in absolute coordinates system is less than 0.507" and the measurement error of yaw angle, the angle between the non-diffracting beam and the axis of TBM, is less than 0.05mradial, when the distance between two conjoint work station positions is 200m. Therefore, this measurement system can be used as a guidance system of TBM in the shield tunneling and it can improve the orientation precision in the shield tunneling.

Keywords: Non-diffracting beam, phase grating interferometer, angle measurement, shield machine azimuth, yaw angle.

1 Introduction

The shield tunneling construction method is a kind of construction method that digs the tunnel underground by using the shield machine, which is such a transportable steel tube structure that can dig and move ahead in the stratum [1]. And it can not only support the pressure from the stratum, but also make the upper strata not sink or uplift. At present, the shield tunneling construction method has been introduced into the related constructions in several cities in China, such as Shanghai, Guangzhou, Beijing and Wuhan [2]. As the result of the high safety and little impact on the environment, the shield tunneling technology is more and more applied to the subway tunneling construction [3].

During the shield tunneling construction, there is always an automatic guidance system of shield machine which is able to measure and control the position and pose of shield machine and to display in the real-time. Because of the advantages of low human resource cost and high measurement speed, the automatic guidance system of shield machine has become the most important one of the shield tunneling measurement technology. According to the measuring instruments selected and the level of

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 423–432, 2009.

automation, the measurement methods can be divided into: manual measurement method (ruler method), semi-automatic measurement method (gyroscope method), auto-oriented method (prism method and ELS laser method) [4].

As related technologies can enhance the accuracy and reliability of shield tunneling construction, the laser-guide system has widely been adopted [5]. Presently, the automatic guidance systems of shield machine are mainly ZED from England, VMT from German [6] and GYRO from Japan. ZED from England and VMT from German are the laser-guided, while GYRO from Japan is guided by gyroscope.

And now, a new type of laser target was studied and developed by Huazhong University of Science and Technology and Shanghai Tunnel Engineering Co., Ltd [7].

The existing automatic guidance systems of shield machine are all based on the laser total station. To improve the measuring precision, a precise measurement method of azimuth angle for the shield machine, such as TBM, have been developed based on the non-diffracting beam technology [8,9] and the phase grating interferometer [10] in this paper. The new non-diffracting beam guidance system is achieved with independent intellectual property rights and will improve the measurement accuracy of automatic guidance system of TBM.

2 Principle of Non-diffracting Beam Guidance System

In shield tunneling construction, application sketch map of non-diffracting beam guidance system is as shown in Figure 1 from reference [6]. Non-diffracting beam guidance system consists of Two-dimensional measurement and control device, the front laser-electronic target and the behind laser-electronic target (reference target), a laser range finder, and an industrial control computer, and it's used to measure and to display the position and pose (azimuth, elevation angle and roll angle) in the city coordinate system (absolute coordinate system) and to calculate the error between the direction of TBM and the designed axis of tunnel.

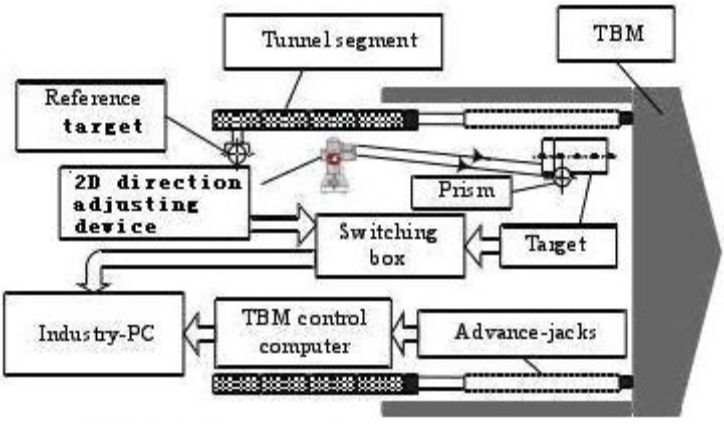


Fig. 1. Sketch map of automatically oriented system for controlling tunnel boring machine

Two-dimensional measurement and control device and the reference target are fixed on the formed tube of tunnel. The front laser-electronic target, the prism of laser range finder, and the industrial control computer are installed on TBM.

After the error is measured, it is delivered to the control computer of TBM and is changed into the displacement of the jack of TBM. Then the position and pose of TBM can be changed in time, so the trend of TBM is controlled on the designed axis of tunnel with a small margin of error.

3 Working Principle of Non-diffracting Beam Guidance System

To measure the azimuth of TBM easily, the datum axis of laser target, yaw angle measurement system, is installed to parallel with the axis of TBM, so the angle θ between non-diffracting beam and the datum axis of measurement system can be regarded as the angle between non-diffracting beam and the datum axis of TBM. As shown in figure 2 from reference [7], Coordinate system $O_1X_1Y_1Z_1$ is established and its coordinate axes parallel to the axes of the absolute coordinate system. Drawing auxiliary plane ABCD, the projection of OA and OB on horizontal plane is OD and OC.

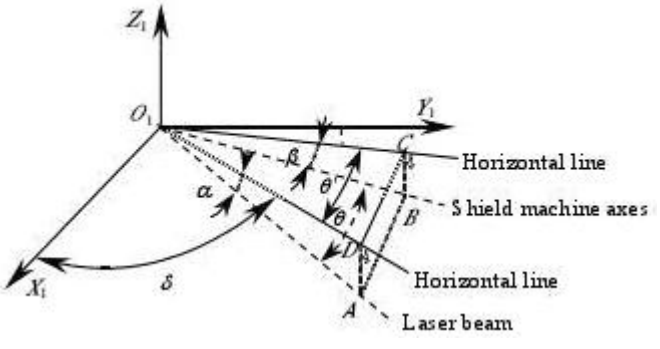


Fig. 2. Measurement of azimuth angle

The diagram of non-diffracting beam guidance system is as shown in figure 3. Two-dimensional measurement and control device consists of two-dimensional mechanism whose direction can be adjusted automatically in horizon and vertical direction and whose two rotation axes are perpendicular to each other.

The each of two rotation axes is attached with a circular phase grating interferometer which is used to measure rotation angle and, in addition, a non-diffracting beam generator and a laser range finder is fixed on the device, and the optical axis of non-diffracting beam is parallel to that of the laser range finder. In the figure 3, the laser-electronic target consists of laser yaw angle measurement system, two clinometers (one for pitch angle of TBM and another for roll angle of TBM), and a reflection prism. The distance between the reflection prism and the optical axis of non-diffracting beam is equal to the distance between the optical axis of the laser range finder and the optical axis of non-diffracting beam.

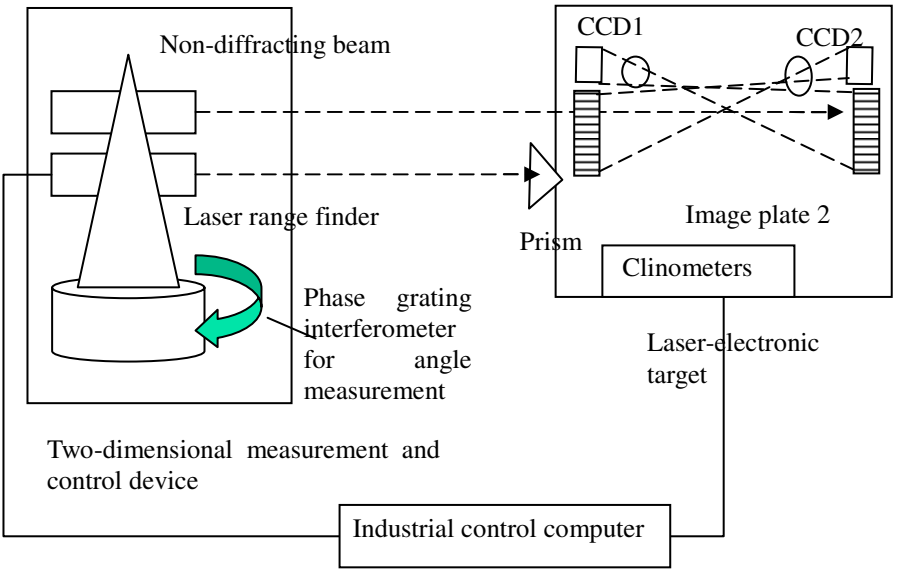


Fig. 3. The diagram of non-diffracting beam guidance system

3.1 The Principle of Horizontal Azimuth Angle Measurement for TBM in the Absolute Coordinate System

The principle of horizontal azimuth angle measurement for TBM is shown in figure 4. A facula is formed and its image is detected by CCD when non-diffracting beam illuminates the front image plate of laser-electronic target. O(0,0) is the reference point on the front image plate of laser-electronic target and, compared with O(0,0), the offset of the facula center is S(x',z'), so the offset in horizontal and vertical direction in the absolute coordinate system can be calculated when the pitch angle of TBM is supposed as 0:

$$\begin{cases} x = x' \cos\varphi - z' \sin\varphi \\ z = x' \sin\varphi + z' \cos\varphi \end{cases} \tag{1}$$

Where φ-- roll angle of TBM rotates around its own axis (Y_1 axis), which can be measured by clinometers for roll angle of TBM, $S = \sqrt{x^2 + z^2} = \sqrt{x'^2 + z'^2}$, and L0 is the distance between Two-dimensional measurement and control device and the front laser-electronic target.

The measurement principle of azimuth angle for TBM in the absolute coordinate system: horizontal azimuth angle δ of non-diffracting beam in the absolute coordinate system and the angle α between non-diffracting beam and horizontal plane can be gauged directly from the phase grating interferometer of Two-dimensional measurement and control device. As the adjusting error of the mechanics of Two-dimensional measurement and control device is far more than the measuring error of phase grating

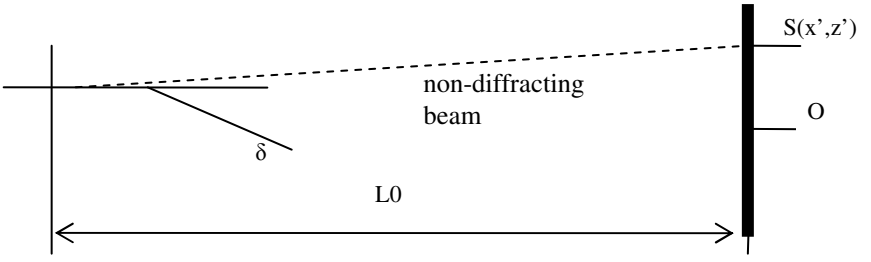


Fig. 4. The principle of horizontal azimuth angle measurement for TBM

interferometer (the ratio of limits of error can be 10-20:1), the facula image center of non-diffracting beam can not track reference point O (0,0) on the front image plate of the front laser-electronic target exactly. The horizontal and vertical azimuth angle of the reference point on the front image plate of the laser-electronic target in the absolute coordinate system can be calculated:

$$\begin{cases} \delta_x = \delta + x / L0 \\ \delta_z = \alpha + z / L0 \end{cases} \tag{2}$$

3.2 The Measurement Principle for Angle between Non-diffracting Beam and the Axis of TBM

The measurement principle for angle between non-diffracting beam and the axis of TBM is shown in Figure 3, the offset of the facula image center of non-diffracting beam to O(0,0) is S(x',z') and the offset of the facula image center of non-diffracting beam to reference point Oh(0,0) on the behind image plate of laser-electronic target is Sh(xh',zh')(considered pitch angle is 0). S(x,z) is S(x',z')in the absolute coordinates and Sh(xh,zh) is Sh(xh',zh'), so the projection of the angle between non-diffracting beam and the axis of TMB on horizontal plane is θz, and on vertical plane is θx:

$$\begin{cases} \theta_z = \arctan((x - xh) / d) \\ \theta_x = \arctan((z - zh) / d) \end{cases} \tag{3}$$

Where θz is yaw angle and d is the distance between the front image plate and the behind image plate, supposed d=300mm.

The axis through reference points of the front and behind image plates is supposed to parallel to the axis of TMB. As is shown in figure 2, according to cosine theorem, the projection of the angle between non-diffracting beam and the axis of TMB on horizontal plane and vertical plane is

$$\begin{cases} \theta_z^{'} = \arccos(\dfrac{\cos\theta_z - \sin\alpha\sin\beta}{\cos\alpha\cos\beta}) \\ \theta_x^{'} = \arccos(\dfrac{\cos\theta_x - \sin\alpha\sin\beta}{\cos\alpha\cos\beta}) \end{cases} \tag{4}$$

Where β is angle between horizontal plane and axis of TMB and α is angle between non-diffracting beam and horizontal plane.

According to Equation 4, to measure azimuth angle of TMB, yaw angle, which is the angle θz between non-diffracting beam and the datum axis of the front target on the horizontal plane, should be measured firstly. The optical axis of non-diffracting beam is very straight and its direction is stable, so it can be used as a straight line. According to the basic optical principles, the position of the facula on the front image plate and the behind will follow to change if the angle between non-diffracting beam and the datum axis of TMB changes.

The absolute measurement accuracy of offset S(x',z') of the facula on the front image plate and the behind is less than 0.015mm/20m~0.15mm/200m. As image matching of two faculae on the front image plate and the behind, the measurement accuracy of the difference between the faculae on the front image plate and the behind can be up to 0.015mm and the measurement accuracy of angle between non-diffracting beam and the axis of TMB is less than 0.05mradial if d=300mm.

3.3 The Principle of Measuring Rotation Angle of Two-Dimensional measurement and Control Device

3.3.1 The Angle Measurement Principle of Circular Phase Grating Interferometer

Based on reference [10], the angle measurement principle of phase grating interferometer is developed and shown in figure 5. There is just one circular phase grating and there is just one optical spot on the phase grating. The angle measurement principle of phase grating interferometer is as following:

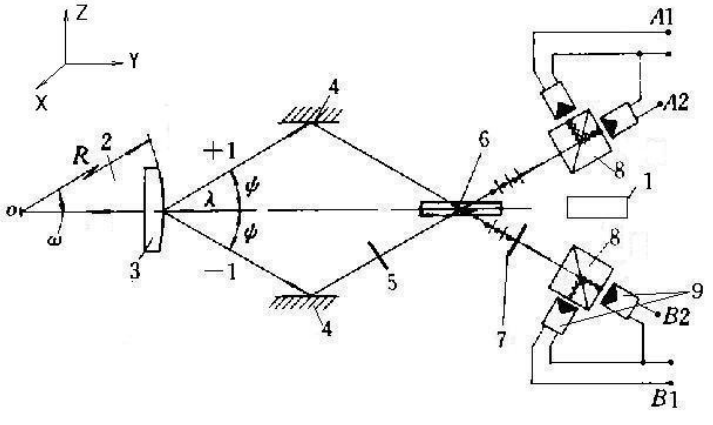


Fig. 5. The angle measurement principle of phase grating interferometer

The linearly polarized, coherent, and collimated laser beam from the laser diode1 illuminates on circular phase grating 3 along the direction shown in figure 5 and is diffracted to the positive and negative first orders, ±1,of circular diffraction. The reflected first orders of diffraction respectively reaches two plane reflectors 4, which are

paralleled, and are deflected on plane reflectors 4, namely the positive first order of diffraction is deflected into the direction of the negative first order of diffraction and the negative first order of diffraction is deflected into the direction of the positive first order of diffraction. A half-wave plate 5 in the beam path of the negative first order effects a phase shift of 180º and rotation of the polarization direction by 90º. After passing the non-polarizing beam splitter 6, the diffracted wave trains are superimposed on each other.

Since the polarization directions of both beams are perpendicular to each other, no interference is yet possible. One beam pair is phase-shifted by 90º by the quarter-wave plate 7. Both beam pairs pass through an analyzer 8, which rotates the directions of polarization by +45º and -45º such that the beams of the negative first and positive first order have the same polarization direction and interfere. The signal voltages of photoelectric sensors 9 are A1, A2, B1, and B2. When the circular phase grating 3 moves relative to the scanning unit, two modulation signals, A1 and A2, with a phase shift of 180º can be detected at the output of two analyzers. And so can B1 and B2. By connecting the photoelectric sensors 9 in difference, the DC component of light and dark current is rejected. Two symmetric 90º phase-shifted sinusoidal signals can thus be derived. The circular phase grating 3 is attached on the mechanics 2 which radius is R.

3.3.2 The Theoretical Analysis of Range and Accuracy of Rotation Angle Measurement

For the sinusoidal phase grating from references [10], it can be concluded that when the phase difference of the interference signal changes a cycle, the phase difference between diffraction waves will change $\Delta\phi = 2\pi = 4\pi z / d$. Where, z_0=d/2 is the cycle constant of the phase difference.

In this angle measurement system, the cylindrical radius of cylindrical holographic phase grating is R=112mm and on its surface, there are many sinusoidal diffraction grating stripes which is parallel to the axis of the cylindrical surface, the valid arc-length of the grating is ±10mm, and the grating pitch constant is 0.833um. By 256 subdivision of the interference signal, the angle measuring range and resolution are

$$\begin{aligned} \phi_0 &= \pm 10 / R \\ &= \pm 5.11^\circ \end{aligned} \tag{5}$$

$$\begin{aligned} d\phi &= \sin^{-1} 0.833 / R / 256 / 2 \\ &\approx 0.003'' \end{aligned} \tag{6}$$

3.3.3 Error Factors

(1) Supposed the second derivative of absolute curvature to Y coordinate <0.00035radial, it causes the measurement error of phase grating interferometer is less than 0.06um, corresponding the angle measurement error is less than 0.075";

(2) Supposed the repeat error of rotation axis positioning is less than 0.1um, the angle measurement error that it causes is less than 0.184".

According to the square principle, the total angle measurement error of phase grating interferometer is less than 0.199".

3.4 The Measurement Principle of Azimuth Angle of the Reference Line of Two-Dimensional Measurement and Control Device

When Two-dimensional measurement and control device is moved to next work station position, the azimuth angle of the reference line of Two-dimensional measurement and control device in the absolute coordinate system should be measured again. The azimuth angle δ20 of the reference point O2(0,0) on the image plate of the reference target is measured (shown in figure 6).

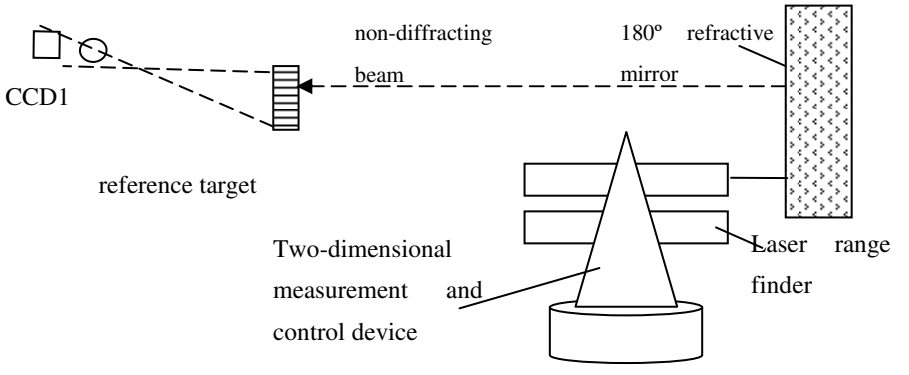


Fig. 6. The measurement principle of azimuth angle of the reference line of Two-dimensional measurement and control device

By the refractive mirror of 180 degrees which is fixed along the horizontal direction or plumb directions, non-diffracting beam is deflected 180 degrees. By the use of the behind target, the reference azimuth angle δ20 (-5.11°~+5.11°) of the reference point on the image plate of the reference target to Two-dimensional measurement and control device can be measured and calculated as is shown in figure 7.

$$\delta 20=\delta 2+ (x_2\text{-}dd)/L01 \tag{7}$$

Compared with O2(0,0), the offset of the facula center on the image plate of the reference target is S2(x_2,z_2),where S2=$\sqrt{x_2^2 + z_2^2}$ and δ2 is the angle measured by the

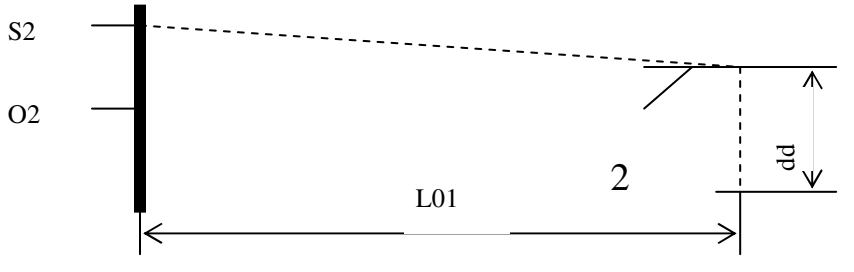


Fig. 7. The calculated diagram of azimuth angle of the reference line of Two-dimensional measurement and control device

phase grating interferometer. dd is the distance of reflected light beam to non-diffracting beam and L01 is the distance between Two-dimensional measurement and control device and the reference target.

And more, the coordinates and the azimuth angle of the reference line of Two-dimensional measurement and control device in the absolute coordinate system are calculated.

4 System Error Analysis

4.1 The Rotation Angle Measurement Error of Two-Dimensional Measurement and Control Device

The measurement error of rotation angle is less than 0.199" in one times, while measurement of the two angles in two perpendicular sections of Two-dimensional measurement and control device is inter-disturbed and supposed the measurement error is less than 3%, so the comprehensive measurement error of rotation angle is less than 0.2".

4.2 The Measurement Error of Horizontal Azimuth Angle in the Absolute Coordinate System

The comprehensive measurement error of rotation angle of Two-dimensional measurement and control device is less than 0.2".The measurement error of the facula offset is less than 0.015mm/20m~0.15/200m, so the equivalent angle error is less than 0.155". According to the square principle, the measurement error of horizontal azimuth angle, the measurement error of the front target, in the absolute coordinate system is less than 0.253".

4.3 The Angle Measurement Error of the Reference Line of Two-Dimensional Measurement and Control Device

The measurement error of the horizontal azimuth angle of the reference target in the absolute coordinate system is also less than 0.253".

Supposed the refractive angle error of 180º refractive mirror is less than 0.2"and the positioning error of 180º reflection mirror is less than 0.15mm, so when the work station position of Two-dimensional measurement and control device is advanced 200m (or 100m), the equivalent angle error of the rotation angle of the positioning error of the mirror is less than 0.155" in 200m (or 0.309" in 100m);

According to the square principle, the angle measurement error of the reference line of Two-dimensional measurement and control device is less than 0.391" /200m(0.474" /100m).

4.4 The Comprehensive Measurement Error of Horizontal Azimuth Angle in the Absolute Coordinate System

According to the square principle, the comprehensive measurement error of horizontal azimuth angle in the absolute coordinate system is 0.507" /200m (0.573" /100m).

5 Summary

The following is concluded

(1) The comprehensive measurement error of horizontal azimuth angle in the absolute coordinate system is less than 0.507”/200m;
(2) The measurement error of yaw angle, the angle between non-diffracting beam and the axis of TMB on the horizontal plane, is less than 0.05mradial;

Therefore, in this paper, a precise measurement method of azimuth angle is developed and the non-diffracting beam guidance system can be achieved with independent intellectual property rights and it will improve the measurement accuracy of automatic guidance system of TBM in shield tunneling.

Acknowledgments

This research was supported by the Major State Basic Research Development Program of China (grant no. 2007CB714000) and the Foundation of Digital machining key lab of Hubei in China (grant no. SZ0605).

References

[1] Wang, N.: Research on measurement technology in shield tunneling construction. Railway Engineering (12), 3–5 (2000)
[2] Qin, C.-l.: Main contents and methods for improving precision of TBM construction surveying. Beijing Surveying and Mapping (3), 1–3 (2003)
[3] Zhang, H.: Research on some problems in engineering design and constructing process in shield tunneling of subway. Report of Post-doctor in Shanghai Jiaotong University (August. 2004)
[4] Guoling, L., Huang, W., Pang, H.: Research on application of automatic guidance System in shield tunneling. Tunnel Construction 25(5), 11–14 (2005)
[5] Dong, W., Ren, G., Ma, L.: The principle of laser navigation system of the tunnel boring machine. Engineering of Surveying and Mapping 14(4), 61–64 (2005)
[6] Liu, D., Huang, Y.: Analysis of the VMT guiding system of shield machine. Tunnel Construction 22(4), 47–48 (2002)
[7] Pan, M., Zhu, G.: Study of measuring methods of the automatic guiding system of shield machine. Construction Technology 34(6), 43–45 (2005)
[8] Zhang, X., Zhao, B., Li, Z.: A collimating tracking and positioning laser system with non-diffracting beams and Moire fringe technology. Laser Technology 25(2), 118–121 (2001)
[9] Zhang, X., Zhao, B., Li, Z.: Study of the tolerance of the non-diffracting beam to laser beam deflection. J. Opt. A: Pure Appl. Opt. 4(1), 78–83 (2002)
[10] Jiang, X., Li, Z., Xie, T.: A theoretical investigation on topography measurement of curved surfaces using hologram grating interferometer. Journal Of Huazhong University Of Science And Technology 22(2), 60–64 (1994)

Identification of Abnormal Operating Conditions and Intelligent Decision System for Earth Pressure Balance Shield Machine*

Xiuliang Li, Junjie Jiang, Hongye Su**, and Jian Chu

National Key Laboratory of Industrial Control Technology, Institute of Cyber-systems and Control, Zhejiang University, 310027 Zhejiang, China
hysu@iipc.zju.edu.cn

Abstract. In the Earth Pressure Balance shield construction, the soil dug in the capsule is difficult to form the "plastic flow state" and will cause three abnormal operating conditions including occlusion, caking in the capsule and spewing at the outlet of the dump device. These abnormal operating conditions will then trigger failure in tunneling, cutter-devices damage and even catastrophic incidents such as ground settlement. This paper effectively integrates mechanism of abnormal operating conditions and knowledge of soil conditioning and establishes a uniform model of identification of abnormal conditions and intelligent decision support system based on belief rule-base system. The model makes full use of knowledge of improving the soil, construction experience and data to optimize the model on-line. Finally, a numerical simulation with specific construction data is presented to illustrate the effectiveness of the algorithm.

Keywords: Earth Pressure Balance Shield Machine (EPB), soil pressure cabin, abnormal operating condition, belief rule-base system.

1 Introduction

In the Earth Pressure Balance shield construction, the soil dug in the capsule is difficult to form the "plastic flow state" and will cause three abnormal operating conditions including occlusion, caking in the capsule and spewing at the outlet of the dump device. These abnormal operating conditions will then trigger failure in tunneling, cutter-devices damage and even catastrophic incidents such as ground settlement.

Liu pointed out that construction fault will arises if the soil in the pressure cabin don't maintain the state of plastic flow [1]. According to the investigation report of the Japanese tunnel association, the EPB construction fault is half caused by caking, 38.5% by occlusion, and 11.5% by spewing.

In order to prevent the occurrence of abnormal conditions, we need to add suitable materials to convert the soil into "plastic flow state". But the prediction and

* Project supported by National Basic Research Program of China (973 Program 2007CB 714000) and The National High Technology Research and Development Program of China (863 Program 2006AA040309).

** Corresponding author.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 433–442, 2009.

decision-making of the abnormal operating condition mainly depend on the experience of construction staff and often lead to construction faults by human uncertainties.

In this paper, the mechanism of abnormal operating conditions and knowledge of soil improvement are effectively integrated to establish a uniform model (see Fig.1) which can identify the abnormal conditions and make intelligent decisions based on improved belief rule-base method.

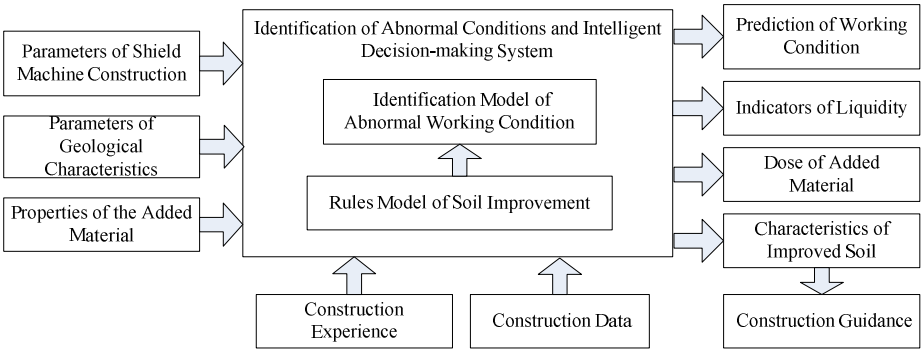


Fig. 1. Intelligent Decision-support System for Abnormal Operating Conditions

2 Establish Mechanism Model of Abnormal Conditions

2.1 Mechanism Model of Soil Caking in Capsule

By reason of the lack of liquidity, the soil under the press of sealed cabin could be compacted to form a hard "cake". In accordance with the conditions of EPB boundary, the consolidation model of the soil has been established[2]. Based on one-dimensional consolidation theory, one can define the degree of consolidation $U(t)$ at point z as follows

$$U = 1 - \frac{4}{\pi}\sum_{m=1}^{\infty}\frac{1}{m}\sin\left(\frac{m\pi z}{4B}\right)e^{-\frac{\pi^2 m^2 kt(1+e_1)}{16a_v r_w B^2}}$$

According to this formulation, the greater the permeability coefficient k is, the greater the degree of consolidation U is. The greater the compression coefficient a_v is, the smaller U is. So we can avoid caking problems in construction through improving the soil coefficients of permeability and compressibility.

2.2 Mechanism Model of Soil Occlusion in Capsule

When the soil in the capsule has a large internal friction angle, the friction between the soil and the sidewall is great. If the pressure in front of tunneling interface is big, soil adhesion phenomenon will occur at the sidewall. If the adhesive soil increases, arch is prone to occur. Soil under pressure in the Shield machine capsule arches as shown in Fig.2.

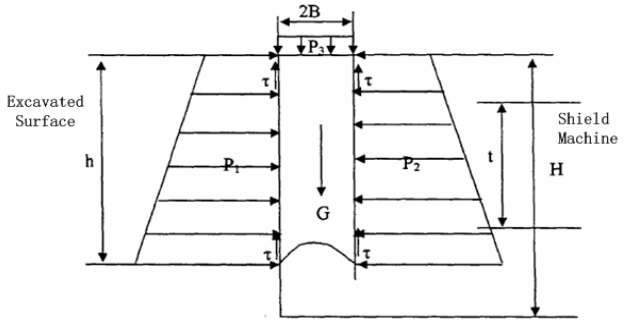


Fig. 2. Arch in Capsule

The pressure of the shield excavation interface can be calculated as[2]

$$P_1 = \mathrm{tg}^2\left(45^\circ - \varphi/2\right)(\gamma_0 - 10)(H_l + Z) + \gamma_w(H_0 + Z)$$

Where φ is the internal friction angle of soil, $K_0' P_1$ indicates the pressure of shield machine at the top, K_0' is the static side pressure coefficient of the soil. τ is the effective range of agitator in the capsule.

From the vertical force balance of the arched soil, we can get [2]

$$BK_0 P_1 + \gamma Bh = f\int_0^{\frac{H-\tau}{2}} P_1 dZ + f\int_{\frac{H-\tau}{2}}^{h} P_1 dZ$$

Where h is the height of the arched soil.

According to the analysis of the model, the height of the arched soil has main relationships with φ, f and γ. Among them, the impact of φ plays a leading role, f and γ have little impact on h. so in order to avoid "blocking", we must improve the strength of the soil in excavation.

2.3 Mechanism Model of Water Spewing

If the soil in the capsule and spiral conveyor cannot effectively resist the higher water pressure of the tunneling interface, water spraying will take place at the exit of spiral conveyor. Changes of water pressure in shield machine can be illustrated as Fig.3. Based on Darcy's law and flow balance, we can found the model [3]

$$H_2 = H_1 - \frac{QL_1}{KA_1} - \frac{QL_2}{KA_2}$$

Where L_1 is the length of the capsule, L_2 is Dump's length, A_1 is Cross-sectional area of the capsule, A_2 is Dump's cross-sectional area, H_1 is water head of excavation surface, H_2 is water head at exit of dump device, θ is obliquity of spiral Dump, K is the permeability coefficient of soil in sealed cabin, Q is cross-section seepage of capsule and Dump device.

With the equation we can see, permeability coefficient of soil in excavation is the most sensitive parameters and has a significant impact on H_2 and Q. Therefore, improving the permeability coefficient of excavation soil is the most practical and effective approach to prevent spewing.

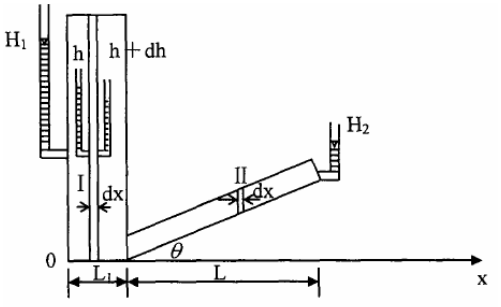


Fig. 3. Water Pressure in Shield Machine

2.4 Establish Rule Model for the Soil Improvement

In this paper, we only analysis foam material as an example, and other additives can be modeled in the similar approach.

As the bubble is a complex gas-liquid two-phase body, there is no unified test method or model to evaluate their performance. Generally the performance of the foam is given through determination of experiment curve, such as Fig.4 [2].

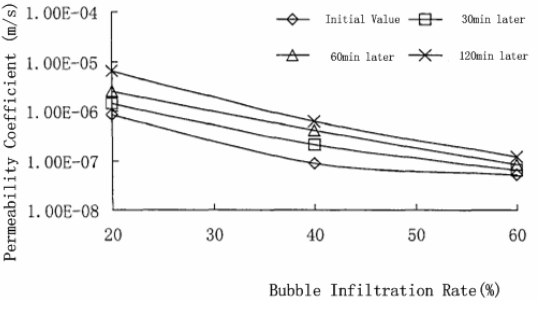


Fig. 4. Relationship between the Proportion of Mixing Bubble and Permeability Coefficient

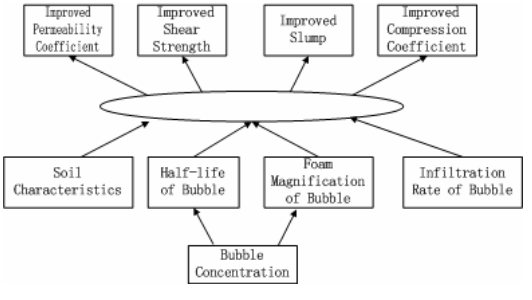


Fig. 5. The Causal Relationship between Foam Parameters and Parameters of Improved Soil

The causal relationship between foam parameters and performance parameters of improved soil can be illustrated in Fig.5. Knowledge of the rules of the cause and effect can be extracted from artificial experience, experiment data and mechanism analysis.

3 Identification of Abnormal Conditions and Intelligent Decision-Making Model Based on Belief Rule-Base Method

At present, in EPB construction, the type and quantity of adding material are determined ahead based on experience, and then, construction staff will repeatedly grope the best mixing ratio in construction, but often fail to achieve the effect, and delay the construction and increase the project cost. For this problem, this paper effectively integrates mechanism of abnormal operating conditions and knowledge of soil conditioning to establish a uniform model of abnormal conditions identification and intelligent decision support system based on belief rule-base system.

3.1 Introduction of Belief Rule-Base Algorithm

Abnormal working condition model contains a variety type of models such as quantitative models, qualitative models, rule based models, and also contains a variety type of data, such as quantitative data, qualitative data, and rule knowledge based on the experience. In order to integrate these models and data with different attributes, this paper adopts improved Belief rule base methods to establish the model of abnormal working condition prediction and intelligent decision-making system.

A belief rule base inference methodology using the evidential reasoning approach (RIMER) has been developed recently [4-10],where a new belief rule base (BRB) is proposed to extend traditional IF-THEN rules and to capture more complicated causal relationships using different types of information with uncertainties. This paper promotes the rule-base algorithm to integrate various models and data.

3.1.1 Belief Rule Base

In order to capture the dynamics of a system, a belief rule base (BRB) with a collection of belief rules is defined as follows[8]:

$$R_k : \text{IF } x_1 \text{ is } A_1^k \wedge x_2 \text{ is } A_2^k \wedge \cdots \wedge x_{T_k} \text{ is } A_{T_k}^k \text{ THEN } \left\{\left(D_1, \ \bar{\beta}_{1k}\right)\left(D_2, \ \bar{\beta}_{2k}\right), \cdots, \left(D_N, \ \bar{\beta}_{Nk}\right)\right\}$$

$$\text{With a rule weight } \theta_k \text{ and attribute weight } \delta_{1,k}, \delta_{2,k}, \cdots, \delta_{T_k,k}$$

Where $x_1, x_2, \cdots, x_{T_k}$ represents the antecedent attributes in the kth rule. A_i^k is the referential value of the ith antecedent attribute in the kth rule. $\bar{\beta}_{j,k}$ is the belief degree assessed to D_j which denotes the jth consequent.

BRB represents functional mappings between antecedents and consequents possibly with uncertainty. It provides a more informative and realistic scheme than a simple IF-THEN rule base for knowledge representation. Note that the degrees of belief $\beta_{i,k}$ and the weights could be assigned initially by experts and then trained or update using dedicated learning algorithms [4,7].

3.1.2 Belief Rule Inference Using the Evidential Reasoning Algorithm

The evidential reasoning (ER) approach mainly consists of two steps, and here will review the ER approach.

1) calculation of the activation weights

The activation weight of the kth rule at nth step, $w_k(n)$ is calculated by [8]

$$w_k(n)=\frac{\theta_k\prod_{i=1}^{T_k}\left(a^k{}_{i,j}(n)\right)^{\bar{\delta}_i}}{\sum_{l=1}^{L}\theta_l\prod_{i=1}^{T_k}\left(a^l{}_{i,j}(n)\right)^{\bar{\delta}_i}} \quad \text{and} \quad \bar{\delta}_i=\frac{\delta_i}{\max\limits_{i=1,\cdots,T_k}\{\delta_i\}}$$

Where δ_i is the relative weight of the ith antecedent attribute that is used in the kth rule. Because $w_k(n)$ will be eventually normalized so that $w_k(n)\in[0,1]$. Without loss of generality, $\theta_k(n)\in[0,1]$ and $\delta_i(n)\in[0,1]$. $a_k(n)=\prod_{i=1}^{T_k}\left(a^k{}_{i,j}(n)\right)^{\bar{\delta}_i}$ is called the normalized combined matching degree. $a^k{}_{i,j}(n)$ could be generated using various ways, depending on the nature of an antecedent attribute and data available such as a qualitative attribute using linguistic values. The input information can be one of the following types: continuous, discrete, symbolic and ordered symbolic. In order to facilitate data collection, a scheme for handling various types of input information has been summarized by Yang [4,7,8,11].

2) rule inference using the evidential reasoning approach

After activating all rules by the actual input vector $\hat{x}(n)$ at time instant n, the final conclusion $O(Y(n))$ is generated, using the ER analytical algorithms [7,8], as follows:

$$O(Y(n))=F(\hat{x}(n))=\left\{\left(D_j,\beta_j(n)\right),j=1,\cdots,T_k\right\}$$

Where $\beta_j(n)$ denotes the belief degree in D_j at time instant n.

$$\beta_j(n)=\frac{\mu(n)\times\left[\prod_{k=1}^{L}\left(w_k(n)\beta_{j,k}+1-w_k(n)\sum_{k=1}^{T_k}\beta_{i,k}\right)-\prod_{k=1}^{L}\left(1-w_k(n)\sum_{k=1}^{T_k}\beta_{i,k}\right)\right]}{1-\mu(n)\times\left[\prod_{k=1}^{L}\left(1-w_k(n)\right)\right]}$$

$$\mu(n)=\left[\sum_{j=1}^{T_k}\prod_{k=1}^{L}\left(w_k(n)\beta_{j,k}+1-w_k(n)\sum_{k=1}^{T_k}\beta_{i,k}\right)-(T_k-1)\prod_{k=1}^{L}\left(1-w_k(n)\sum_{k=1}^{T_k}\beta_{i,k}\right)\right]^{-1}$$

The logic behind the approach is that, if the consequent in the kth rule includes D_i with $\beta_{i,k}>0$ and the kth rule is active, then the overall output must be D_i to a certain degree, which is measured by both the degree to which the kth rule is important to the overall output and the degree to which the antecedents of the kth rule are activated by the actual input $\hat{x}(n)$ [4].

3.1.3 Optimization Belief Rule-Base System

Beliefs in a rule base may initially be provided by human experts based on individual experiences and personal judgments, and then optimally trained if observed input-output data are available. In other words, $\beta_{i,k}$ can be trained if appropriate data is available. In addition, a change in rule weights θ_k and attribute weights δ_i may have significant impact on the performance of a BRB system [8], so they also need to be trained for achieving desirable performance.

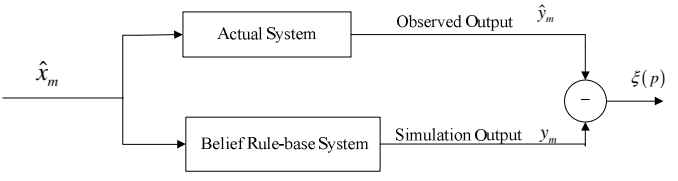


Fig. 6. diagram of training process of Belief rule-base

Fig.6 shows the process of training a BRB, where $\hat{x}_m$ is a given input, $\hat{y}_m$ is the corresponding observed output which either measured using instruments or assessed by experts, y_m is the simulated output that is generated by the BRB system and $\xi(P)$ is the difference between $\hat{y}_m$ and y_m.

The observed training data is composed of M input-output pairs $(\hat{x}_m, \hat{y}_m)$ $(m = 1,\ldots,M)$, where $\hat{y}_m$ is subjective and represented using a distributed assessment with different degree of belief as follows:

$$\hat{y}_m = \left\{\left(D_j, \hat{\beta}_j(m)\right), j = 1,\ldots,N\right\}$$

Where D_j is the reference (linguistic) term in the consequent part of a rule, and $\hat{\beta}_j(m)$ is the degree of belief to which D_j is assessed for the mth pair of observed data. A subjective conclusion y_m that is generated by aggregating the activated rules can also be represented using the same referential terms as the observed output as follows:

$$y_m = \left\{\left(D_j, \beta_j(m)\right), j = 1,\ldots,N\right\}$$

Where β_j is generated by the BRB for a given input.

It is desirable that, for a give input $\hat{x}_m$, the BRB system can generate an output, which is represented as $y_m = \left\{\left(D_j, \beta_j(m)\right), j = 1,\ldots,N\right\}$, which can be as close to $\hat{y}_m$ as possible. In other words, for the mth pair of the observed data $(\hat{x}_m, \hat{y}_m)$, the BRB is trained to minimize the difference between the observed belief $\hat{\beta}_j(m)$ and the belief $\beta_j(m)$ that is generated by the BRB system for each referential term. Such a requirement is true for all pairs of the observed data. This leads to the definition of the objectives for all referential output terms as follows:

$$\min_{p} \xi_j(p) \quad j = 1, \cdots, N$$

$$\xi_j(p) = \frac{1}{M} \sum_{m=1}^{M} \left(\beta_j(m) - \hat{\beta}_j(m) \right)^2 \quad j = 1, \cdots, N$$

The detail of solving this optimal training problem can be seen in references [4,7,9].

3.2 Integration of Abnormal Operating Conditions Based on Belief Rule-Base Model

Through the belief rule-base method, abnormal working conditions mechanism model and soil improvement model can be integrated, as illustrated in Figure 7. And the integration the model can be updated online with the new knowledge and abnormal condition data.

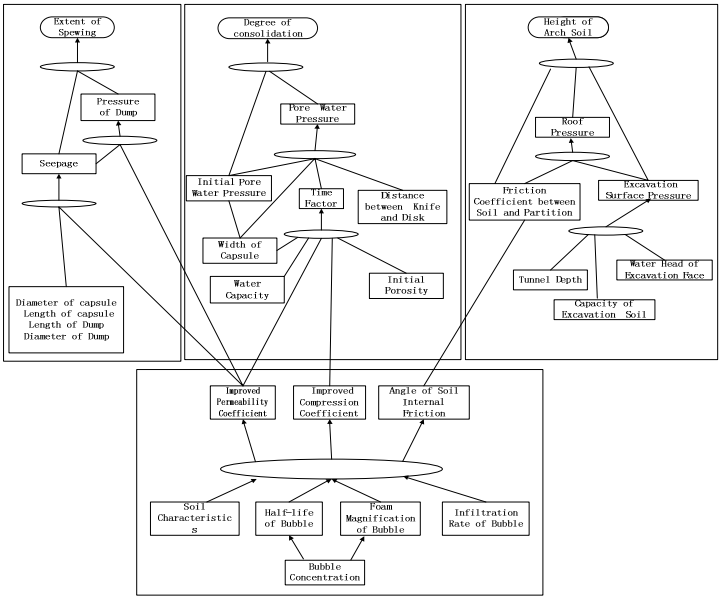


Fig. 7. Integrated Abnormal Condition Forecasting Model

4 Numerical Simulation

In order to verify the effectiveness of the algorithm, sand situation encountered in the Guangzhou EPB shield tunneling construction is selected(the detail information can be seen in [2,12]). The distribution of soil particle size is illustrated in Fig.8.

The distribution of sandy in Guangzhou is uneven with a small average particle size in a very narrow particle size range. In normal shield construction of Guangzhou, arch happen frequently, and soil arching factor increase over times. The adding material is Meyco Fix Slf30 [2,12], the permeability coefficient of sand is $1.21\times10^{-6} m/s$, digging head is $14m$. Based on belief rule base, the mechanism model of occlusion,

experience of soil improvement and data of construction are integrated to establish identification of occlusion abnormal conditions and intelligent decision-making model. The forecast effect of the model is illustrated in Fig.9, which shows that the model is good at the prediction of anomaly and the intelligent decision-making of bubble ratio.

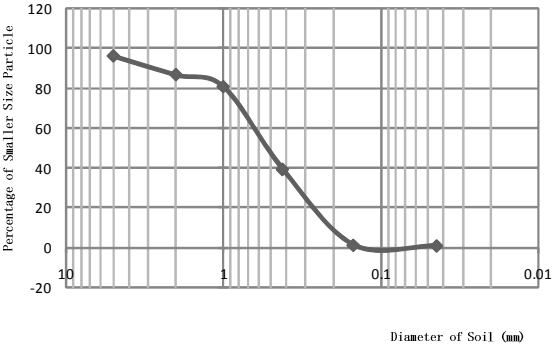


Fig. 8. Particle size curve

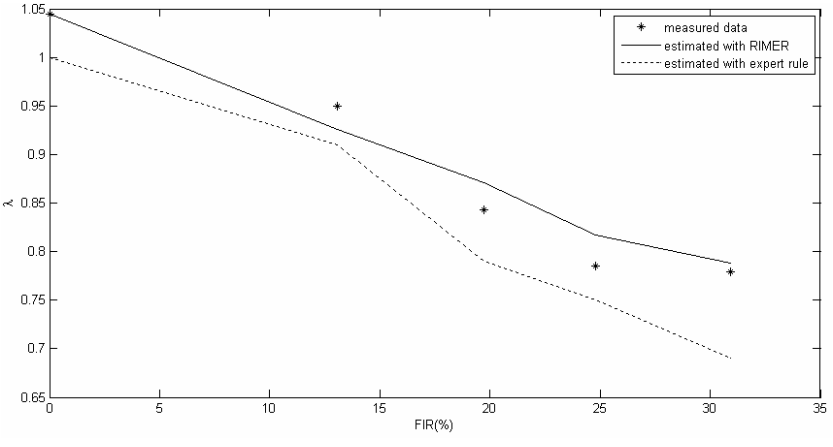


Fig. 9. Influence Prediction of bubbles ratio to arch coefficient

5 Conclusion

In this paper, mechanism of abnormal working conditions and knowledge of soil improvement are effectively integrated by improving the belief rule base, to establish a unified identification of abnormal conditions and intelligent decision-making model, which can effectively integrate a variety of attribute data and multi-attribute model, and make full use of knowledge of soil improvement, construction experience and construction data to update itself on-line. Finally, a model is established using data

and parameters given reference, to identify abnormal conditions and make intelligent decision, for problems in Guangzhou where occlusion is prone to happen. The simulation result shows that the model can forecast abnormal working conditions, decide the bubble ratio intelligently, and increase the effectiveness and efficiency of soil improvement.

Reference

1. Renpeng, L.: Review of earth pressure balance shield. Modern Tunnelling Technology (1), 1–7 (2000)
2. Mingjing, Z.: Study on the mechanism and curing method of blocking in epbs. Master thesis of HoHai University (2004)
3. Qin, J.-s., Wei, Z., Wei, K.-l.: Research on the mechanism of the spewing in the epb shield tunnelling. Chinese Jounal of Geotechnical Engineering 5, 589–593 (2004)
4. Zhou, Z.J., Hu, C.H., Yang, J.B., Xu, D.L., Zhou, D.H.: Online updating belief-rule-based system for pipeline leak detection under expert intervention. Expert Systems With Applications (2008)
5. Liu, J., Yang, J.B., Ruan, D., Martinez, L., Wang, J.: Self-tuning of fuzzy belief rule bases for engineering system safety analysis. Annals of Operations Research 163(1), 143–168 (2008)
6. Guo, M., Yang, J.B., Chin, K.S., Wang, H.W.: Evidential reasoning based preference programming for multiple attribute decision analysis under uncertainty. European Journal of Operational Research (2008)
7. Yang, J.-B., Jun, L.: Optimization models for training belief-rule-based systems. IEEE Transactions on Systems, Man and Cybernetics, Part A 37(4), 569–585 (2007)
8. Yang, J.-B., Jun, L., Jin, W.: Belief rule-base inference methodology using the evidential reasoning approach-rimer. IEEE Transactions on Systems, Man and Cybernetics, Part A 36(2), 266–285 (2006)
9. Liu, J., Ruan, D., Yang, J.B., Lopez, L.M.: Self-tuning fuzzy rule bases with belief structure Tudies in computational intelligence. In: Ruan, D., Chen, G.Q., Kerre, E.E. (eds.), Physica-Verlag, A Springer-Verlag (2005)
10. Yang, J.-B., Pratyush, S.: A general multi-level evaluation process for hybrid madm with uncertainty. IEEE Transactions on Systems, Man and Cybernetics 24(10), 1458–1473 (1994)
11. Yang, J.B.: Rule and utility based evidential reasoning approach for multiple attribute decision analysis under uncertainty. European Journal of Operational Research 131(1), 31–61 (2001)
12. Lin Jian, Z. w.: Study on reduction of cutterhead torque in epb shield with soil conditioning, Master thesis of Hohai University (2006)

Optimal Disc Cutters Plane Layout Design of the Full-Face Rock Tunnel Boring Machine (TBM) Using an Ant Colony Optimization Algorithm

Junzhou Huo[1,2], Wei Sun[1,2], Pengcheng Su[2], and Liying Deng[2]

[1] School of Mechanical Engineering, Dalian University of Technology, Dalian 116024, P.R. China
huojunzhou@gmail.com, sunwei@dlut.edu.cn
[2] Northern Heavy Shenyang Heavy Machinery Group Co., Ltd., Shenyang 110025, P.R. China
supc2005@126.com, shdly0123@163.com

Abstract. Cutter head is one of the most important components for the full-face rock tunnel boring machine (TBM), in which the layout design of the disc cutters is a key issue. Considering the complex engineering technical requirements, this paper studied an ant colony optimization (ACO) algorithm to solve the disc cutters plane layout problem. The searching space of the design variables (disc cutters' position) were dispersed by setting several different search steps, then an ACO algorithm was adopted to search the best searching step of each design variable dynamically during the whole optimization process. Finally, the disc cutters layout design instance was solved by the ACO method to demonstrate its feasibility and effectiveness. The computational results showed that that ACO method can provide various disc cutters layout schemes for engineers choosing from.

Keywords: Full-face tunnel boring machine, Disc cutters, layout, Ant Colony Optimization.

1 Introduction

The full face rock tunnel boring machine (TBM) has been widely applied to the subway, the railway, the highway and the water-electricity projects. Disc cutters layout design of the TBM is the most effective way to improve the global performance of a TBM [1], which directly affects its boring performance, its service life, its main bearing of the cutter head, the vibration of the TBM and the noise.

The layout design of the disc cutters is closely related to the process of cutter penetration into the rock and the cutting force models. Many investigators have studied the cutting force models that can be used to calculate the normal force and rolling force, these models can be divided into two categories: one is the Semi-theoretical model, it is constructed based on the Linear Cutting Machine(LCM) testing and the theoretical analysis, such as the Colorado School of Mines(CSM)

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 443–452, 2009.

model [1][2]. The other is the empirical model; it is constructed based on the historical field performance of machines, such as the Norwegian Institute of Technology (NTH) model [3]. Among all the cutting force models, the multi-factor cutting force models (e.g. CSM and NTH) are more widely accepted and used in the industry, since all the effects of ground conditions, rock properties, machine parameters, and operational and practical constraints are considered.

The disc cutters layout design correlates with the disc cutters head force, the crushed rock mobility and the manufactural process of the cutter head [4]. The disc cutters layout design includes the disc cutters spacing design and the disc cutters plane (circumferential) layout. Considersing the studying of the disc cutters' spacing design, many investigators adopt the numerical simulation method [5-7] and the Linear Cutting Machine (LCM) experiment method [1, 4, 8-10]. the cutters space design is excluded in this paper. After determine the cutting spacing among the disc cutters, the next step is to place all the disc cutters circumferentially on the face of the cutter head. CSM computing model can be used to design the disc cutters layout and calculate individual loads. Rostami [11] surveys the methods of the hard rock TBM cutter head modeling that has been a successful tool used by the industry at various levels of sophistications relative to the end use. Zhang [12] studies the spiral layout rule of the disc cutters and gives the computational equations of the simplified cutter head forces distribution.

The layout design of the disc cutters should meet the geometry constraints and other performance constraints. It belongs to the complex engineering design problem and involves with multi-disciplines knowledge, a practical computational model for the disc cutters layout problem should be established, and then the advanced computational methods should be studied to solve the problem efficiently.

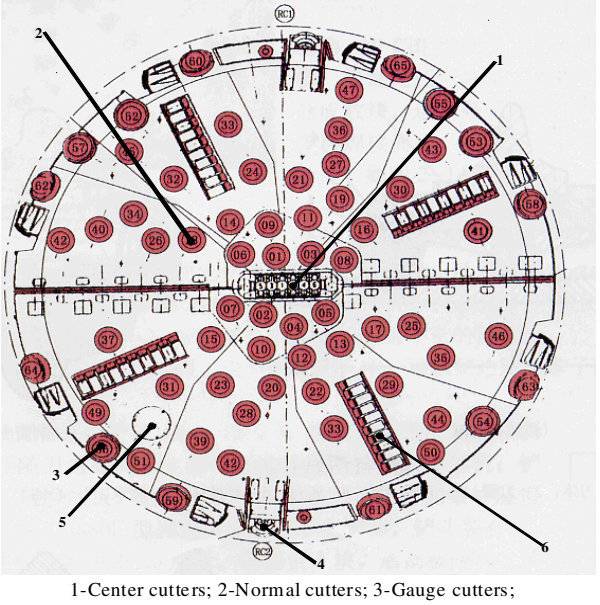
It can be seen that researches on the process of cutter penetration into the rock and the cutting force models have been widely studied, but there are only a few researches on the disc cutters plane layout design. In this study, according to the complex engineering technical requirements and the corresponding cutter head structure design requirements, a nonlinear multi-objective disc cutters layout mathematical model with complex constraints was given, and then a ACO method was used to solve the disc cutters layout design problem. Finally, the disc cutters layout design instance of the TBM was solved by the ACO method.

2 Problem Statements

As shown in Fig. 1, the discs are so arranged that they contact the entire cutting face in concentric tracks when the cutter head turns. The separation of the cutting tracks and the discs are chosen depending on the rock type and the ease of cutting. The size of the broken pieces of rock results from this. The rotating cutter head presses the discs with high pressure against the face. The discs therefore make a slicing movement across the face. The pressure at the cutting edge of the disc cutters exceeds the compressive strength of the rock and locally grinds it. So the cutting edge of the disc pushes rolling into the rock, until the advance force and the hardness of the rock are in balance. Through this displacement, described as net penetration, the cutter disc creates a high stress locally, which leads to long flat pieces of rock (chips) breaking

off. According to the relative literatures [11] [12] and the practical engineers' experiences, the technical requirements of disc cutters plane layout design are summarized as following:

the amount of eccentric forces should be as small as possible;
the amount of eccentric moments should be as small as possible;
Two adjacent disc cutters should crush the rock successively to keep the high cutting efficiency;
All the disc cutters should be contained within the cutter head, with no overlap among the disc cutters;
The position error of the centroid of the whole system should not exceed an allowable value, and the smaller the better;
all the disc cutters should not interference with manholes and buckets respectively.



1-Center cutters; 2-Normal cutters; 3-Gauge cutters;
4-Reaming cutters; 5-manhole; 6- muck buckets

Fig. 1. A disc cutter layout scheme

3 Mathematical Model

Suppose that the set of disc cutters to be located on the cutter head with R radius is $\boldsymbol{CUTs}=\{Cut_1, Cut_2,\cdots, Cut_n\}$ where n=total number of disc cutters. $\boldsymbol{U}=\{U_1,\cdots,U_q\}$ indicates the dynamic rock properties. As shown in

Fig. 1, all the disc cutters are simplified as circles in this work and regarded as rigid bodies with uniform mass distribution. So the ith cutter can be denoted by $Cut_i\ (p_i, r_i)$, where $\boldsymbol{p}_i=(\rho_i,\theta_i,\gamma_i)^T \in R^3$ = position of a reference point (the

centroid of the object in this paper) of Cut_i in the coordinate system *oxyz*, $\rho_i \in (0, R)$ =the radius of *i*th cutter from the center, $\theta_i \in [0, 2\pi)$ = the position angle of the *i*th cutter, $\gamma_i \in [0, \frac{\pi}{2})$ =the tilt angle of the *i*th cutter, r_i = radius of the cutter. Generally, the tilt angle of the normal cutter is zero, the masses and dimensions of all the disc cutters are given in advance, so $\boldsymbol{p}_i$ are the variables to be manipulated in the following procedure. Thus, a general disc cutters layout scheme of a cutter head can be formulated as:

$$\boldsymbol{X} = \{x_1, x_2, \cdots, x_n\}, x_i = \{\rho_i, \theta_i, \gamma_i\}, i = 1, 2, \cdots, n \tag{1}$$

Then based on the technical requirements of disc cutters layout design, the mathematical model of whole disc cutters layout problem can be formulated as follows:

Find a layout scheme $\boldsymbol{X}$ R^{3n}, such that

$$\min_{\boldsymbol{X} \in \boldsymbol{D}} \boldsymbol{y} = f(\boldsymbol{X}, U^I) = (f_1(\boldsymbol{X}, U^I), f_2(\boldsymbol{X}, U^I)) \tag{2}$$

s.t.

Overlapping constraints:

$$g_1(\boldsymbol{X}, \boldsymbol{U}^I) = \sum_{i=0}^{n-1} \sum_{j=i+1}^{n} \Delta V_{ij} \le 0 \tag{3}$$

Two adjacent disc cutters successive rolling constraints:

$$g_3(\boldsymbol{X}, \boldsymbol{U}^I) = \sum_{i=0}^{n-1} (\theta_{i+1} - \theta_i) \ge \Delta\theta \tag{4}$$

Static balance constraints:

$$\begin{aligned} g_4(\boldsymbol{X}, \boldsymbol{U}^I) &= |x_m - x_e| - \delta x_e \le 0 \\ g_5(\boldsymbol{X}, \boldsymbol{U}^I) &= |y_m - y_e| - \delta y_e \le 0 \end{aligned} \tag{5}$$

Manholes and buckets constraints:

$$g_8(\boldsymbol{X}, \boldsymbol{U}^I) = \{\forall i \in \{1, \cdots, n\} : cut_i \cap OP \in \varnothing\} \tag{6}$$

where D denotes the feasible region of variable $\boldsymbol{X}$, $f_1(\boldsymbol{X})$ denotes the side force F_s of the cutter head. $f_2(\boldsymbol{X})$ denotes the eccentric moments of the cutter head. In this paper, a semi-empirical cutting forces model proposed by Rostami [11] was adopted to calculate the normal force. ΔV_{ij} denotes the overlapping area between the Cut_i and the Cut_j. $\Delta\theta$ denotes the allowable angle difference between the two adjacent disc cutters. $O_m(x_m, y_m)$=real centroid of the whole system and $O_e(x_e, y_{ee})$=expected position of O_m, $cut_i \cap OP \in \varnothing$ denotes the *i*th cutter is not to overlap with the manholes and the buckets respectively.

4 A Quasi Combinatorial Model of the Cutters Plane Layout

Disc cutters plane layout problem belongs to a multi objective optimization problem with nonlinear constraints. The swarm-based computing methods (such as the ant colony algorithm, etc.) had been proved to be an effective way to solve the complex optimization problems. Ant Colony Optimization (ACO), originated by Colorni [13], is a recently developed, population based approach that has been successfully applied to several complex optimization problems, such as the TSP [13, 14], the Assignment problems [15] and the job-shop scheduling problems [16]. ACO meta-heuristic has been proposed to provide a unifying framework for most applications of ant algorithms to combinatorial optimization problems. To avoid the search getting stuck, Dorigo [13] proposes an Ant-Q system, in which a random searching method is set to determine the searching situation between the "known information" and "unknown information" before ants select the next city. Based on the Ant-Q system, Stützle and Hoos [17] proposed a Max–Min Ant System (MMAS).

According to the TSP model characters, this paper dispersed the searching space of the design variables (Disc cutters position) by setting several different search steps. Supposed that $x_i(t) \quad R$ is the ith design variable of the disc cutters at the t moment, then the next feasible value of $x_i(t+1)$ at the $(t+1)$ moment can be formulated as:

$$x_i(t+1) \in \{x \left| \begin{array}{l} x = x_i(t) + s_i(j), \\ s_i(j) \in (sp_{i1}, sp_{i2}, \cdots, sp_{iq})^T, \} \\ j = 1, 2, \cdots, q \end{array} \right. \tag{7}$$

That is, each ant is limited to move along q discrete searching steps evenly distributed around its current position. The q searching steps are named step 1, step 2…, and step q, respectively. It can be seen that there are q possible searching steps for each design variable in the next searching step. The definitions of the discrete variable feasible searching step transform the cutters plane layout problem into a series of combinatorial optimization problems. Based on the above mentioned process, this paper constructed a quasi combinatorial cutter plane layout model, which was illustrated in Fig. 2.

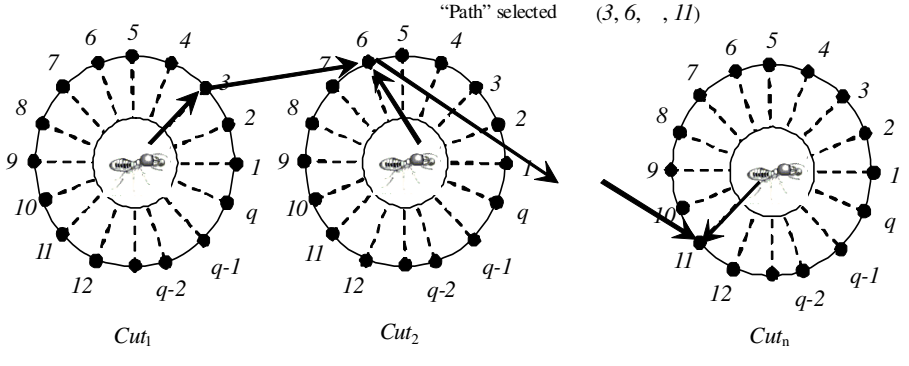


Fig. 2. Quasi combinatorial Model of the disc cutters layout

In constructing the cutters layout design scheme, each variable is asked to select one of the q feasible search steps and is regarded as a decision point. All decision points form a set of decision points *D*. There are totally *q* feasible directions (called the solution components hereafter) for the *i*th variable to choose. As shown in Fig. 2, each circle denotes the *q* feasible search steps of a variable, which forms a set of search steps $E_i = \{1 \ldots q\}$ of the *i*th variable. When the construction of a LST is completed, a path linking all the decision points comes into being, as shown by the arrows in Fig. 2.

5 ACO-Based Cutters Plane Layout Design Algorithm

Based on the above-mentioned quasi combinatorial model and the MMAS by Stützle and Hoos [13], a local search transformation ACO algorithm for the cutters plane

```
Begin
    Initialize the population size, the initial search steps
    of all variables and the number of the ants  t =0   count
    =0;
    For (count=0; count< Count_max; Count_max++)
       Set all elements of the trial intensity matrix as
       τ_max;
       For each ant_i  i=1,2, …,n
            Select the jth solution component with a
            probability P_ij;
       End For
       Produced all new candidate solutions;
       Evaluate the candidate solutions;
       Get the best individual ant_opt and the path L_opt;
       Update the trail matrix;
       If no improvement occurs and at least one element
       in the "best" path is non zero
           Adjust the current searching step scheme
           according to the path.
       End If
    End For
    Output the best searching step scheme and save the best
    cutters plane layout design scheme.
End
```

Fig. 3. Procedure of the ACO algorithm

layout problem was developed. Its main idea is: iteratively the MMAS was used to solve the cutters plane layout problem with dynamic searching steps until the termination criterion is satisfied. All variables synchronously search for the optimal search step around their current positions, which improves the design objectives as much as possible. The procedure of the ACO algorithm can be described as Fig. 3.

6 Application Instance

Taking the disc cutters layout design of full face rock TBM of a water tunnel project as a background, Forty one disc cutters need to be located on the cutter head surfaces shown in

Fig. 1. The relative parameters are as following: ①Rock physical properties the rock is mainly in granite-based geology the punch shear strength of rock τ =7~13(MPa) the Uniaxial compressive strength of rock σ_c =50~93.6 (MPa) the Brazilian tensile strength σ_t =2.14~4 (MPa). ②Cutter head geometry the cutter head radius R=4.015m the rotational speed of cutter head ω =6(r/min)= 0.6283 rad/s the mass of each cutter M=200kg the diameter of each cutter D=483mm the cutter tip width *T*=10mm, the cutter penetration *P*=7mm the cutter edge angle α = 1.5708 rad. The center cutter number n_1=8 the gage cutter number n_2=10 the normal cutter number n_3=33. The whole technical requirements of the disc cutters layout design are as following the expected centroid position of a cutter head x_e=0mm, y_e=0mm, the allowable centroid error of the whole system x_e=5mm, y_e=5mm. According to the engineering requirements, four manholes are located on the cutter head, the radius of the manholes is 200mm, and the locations of the manholes are listed in Table 1. Eight buckets are located in the cutter head symmetrically; their locations are listed in Table 2. The basic problem is to optimize the objective function formulated by Eq.(2) while satisfying the technological constraints given by Eqs.(3)–Eqs.(6).

Table 1. Locations of the manholes

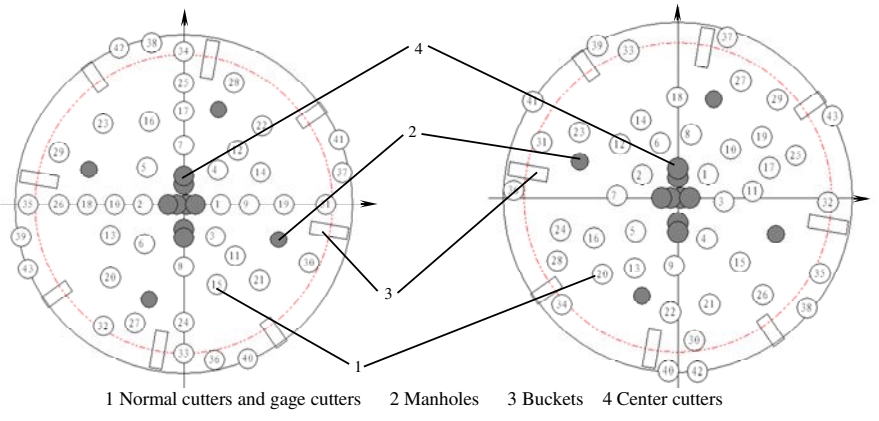
No.	1	2	3	4
ρ /mm	2700.000	2700.000	2700.000	2700.000
θ /rad	1.2217	2.793	4.363	5.934

Table 2. Dimensions and locations of the buckets

No.	1	2	3	4	5	6	7	8
ρ /mm	3700	3500	3700	3500	3700	3500	3700	3500
θ /rad	0.611	1.396	2.182	2.967	3.753	4.538	5.323	6.109
Length/mm	700	900	700	900	700	900	700	900
Width/mm	300	300	300	300	300	300	300	300

Table 3. The performance indices of the optimal layout scheme and the original layout scheme

Methods		M_v / KN.m	F_s /KN	x_m /mm	y_m /mm	Overlapping area	Unsuccessive disc cutters' number	time/*S*
Original layout scheme		154.840	11.558	-2.135	-0.221	0.000	4	Unknown
ACO	Best	3.528	8.36	0.5	-0.58	0.000	2	512
	Average	16.623	9.52	-0.532	0.262	0.000	3	543



(a) The original disc cutters Layout scheme (b) The obtained optimal disc cutters layout scheme

Fig. 4. The layout schemes of the disc cutters

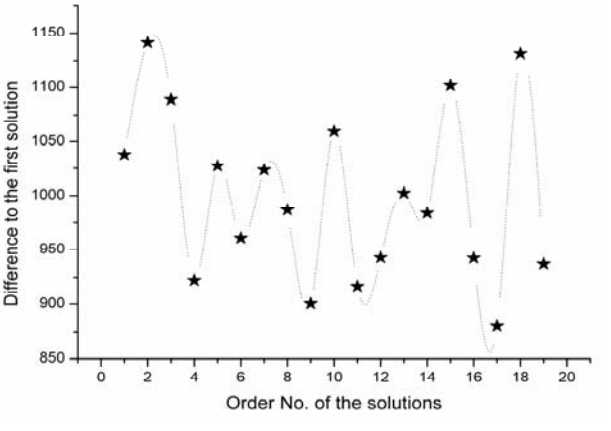


Fig. 5. Discrepancies of the solutions solved by ACO

The numerical experiments were run on an AT-compatible PC, with 1.6GHz Intel processor and 2.00 GB Memory. An ACO algorithm was used to solve the disc cutters plane layout problem. The total running times of ACO is 20; the performance indexes of the optimal layout scheme solved by two methods were listed in Table 3. The disc cutters Layout scheme was shown in

Fig. 4. Discrepancies of the solutions solved by ACO were shown in Fig. 5.

The data in Table 3 showed that, compared with the original disc cutters layout scheme, ① ACO obtained the lower value of the side force and the eccentric moments of the cutter head. And the side force and the eccentric moments of the cutter head of the original scheme are higher. as shown in

Fig. 4, the optimal disc cutters layout scheme solved by ACO can make all adjacent disc cutters cut the rock successively with a relative larger position angle difference; the static balance value of the cutter head solved by ACO is lower than that of the original disc cutters layout scheme; as shown in Fig. 5, after running 20 times, ACO can provide various disc cutters layout schemes for the engineers choosing from; ACO can obtain a optimal disc cutters layout scheme within a shorter running times.

7 Conclusions

Based on the technical requirements of the cutters plane layout design, this paper constructed a quasi TSP-based cutters plane layout model, in which the searching space of the design variables were dispersed by setting several different search steps. An ACO algorithm was adopted to search the best searching step of each design variable dynamically during the whole optimization process. Computational results showed that the proposed method can not only produce various cutters plane layout design scheme with better technical indexes, but also can improve the computational accuracy of the solutions.

Acknowledgements

This work was supported by the Major State Basic Research Development Program of China (973 Program) (Granted No.2007CB714006), Notional Key Technology R&D Program (Granted No.2007BAF09B01), Liaoning key scientific and technological project (Granted No.2008220017), and China Postdoctoral Science Foundation (Granted No.20080431136).

References

[1] Gertsch, R., Gertsch, L., Rostami, J.: Disc cutting tests in Colorado Red Granite: implications for TBM performance prediction. International Journal of Rock Mechanics and Mining Sciences 44, 238–246 (2007)

[2] Rostami, J.: Development of a force estimation model for rock fragmentation with disc cutters through theoretical modeling and physical measurement of crushed zone pressure. Golden, Colorado, USA: Dept of Mining Engineering. Colorado School of Mines (1997)

[3] Rostami, J., Ozdemir, L., Nilson, B.: Comparison between CSM and NTH hard rock TBM performance prediction models. In: Proceedings of Institute of Shaft Drilling Technology (ISDT) annual Technical Conference 1996, Las Vegas (1996)
[4] Rostami, J., Ozdemir, L.: Computer modeling of mechanical excavators cutterhead. In: Proceeding of the World Rock Boring Association Conference, Ontario (1996)
[5] Gong, Q.M., Zhao, J., Hefny, A.M.: Numerical Simulation of Rock Fragmentation Process Induced by Two TBM Cutters and Cutter Spacing Optimization. In: AITES-ITA 2006 congress, Seoul, South Korea (2006)
[6] Gong, Q.M., Zhao, J., Jiao, Y.Y.: Numerical modeling of the effects of joint orientation on rock fragmentation by TBM cutters. Tunnelling and Underground Spacing Technology 20, 183–191 (2005)
[7] Gong, Q.M., Jiao, Y.Y., Zhao, J.: Numerical modeling of the effects of joint spacing on rock fragmentation by TBM cutters. Tunnelling and Underground Spacing Technology 21(1), 46–55 (2006)
[8] Ozdemir, L., Wang, F.D.: Mechanical tunnel boring prediction and machine design, Washington (1979)
[9] Snowdon, R.A., Ryley, M.D., Temporal, J.: A study of disc cutting in selected British rocks. Journal of Rock Mechanics and Mining Sciences and Geomechanics Abstracts 19, 107–121 (1982)
[10] CSM: computer model for IBM performance prediction. (2003), http://www.mines.edu.emipaperscomputer_modeling_for_mechanical_excavatorstbm_performance_prediction.pdf
[11] Rostami, J.: Hard Rock TBM Cutterhead Modeling for Design and Performance Prediction (2008) DOI:10.1002/geot.200800002
[12] Zhang, Z.H.: An Investigation in the Cutter Arrangement Rules for Tunneler Disk. Construction Machinery and Equipment 7, 24–25 (1996)
[13] Dorigo, M., Maniezzo, V., Colorni, A.: Ant system: optimization by colony of cooperating agents. IEEE Transactions on System, Man and Cybernetics (Part B) 26(1), 29–41 (1996)
[14] Colorni, A., Dorigo, M., Maniezzo, V.: Distributed optimization by ant colonies. In: Proceeding of the First European Congress on Artificial Life. Pans, France, pp. 134–142 (1991)
[15] Talbi, E.G., Roux, O., Fonlupt, C., et al.: Parallel ant colonies for the quadratic assignment. Future Generation Computer Systems 17(4), 441–449 (2001)
[16] Colorni, A.: Ant system for the job-shop scheduling. JORBEL 34(1), 39–53 (1994)
[17] Stuetzle, T., Hoos, H.H.: Max - Min Ant System. Future Generation Computer Systems 16(9), 889–914 (2000)

Rotational Moment Analysis and Posture Rectification Strategy of Shield Machine*

Zhenggang Guo, Wei Sun, Ming Yue, and Jian Wei

School of Mechanical and Engineering, Dalian University of Technology,
Dalian, Liaoning, China
gzg1700w@126.com

Abstract. The rolling phenomenon of the shield body occurs frequently in the process of practice construction, which could lead to the derivation of the shield machine and increase the difficulties of the excavation control. It is mainly caused by the rotation of the cutter head. Based on this practical problem, we investigate the forces around the shield body and the cutter head. Meanwhile, the positive and negative rotary controlling techniques are developed to regulate the deviated rolling angle. These studies would provide theoretical basis to the further research about the whole postures rectification of the shield machine.

Keywords: Shield machine, Rotation moment analysis, Rectification strategy.

1 Introduction

1.1 Shield Tunneling Method

Shield machine is an important technique equipment in basic construction and resource development, which is widely applied in tunnel, railroad, highway, mineral mountain etc underground construction. The shield tunneling method is a method to construct tunnel with shield machine. During shield excavation, the jack at the shield tail push the tunnel segment which had been assembled where a reaction force occurred to make the shield going forward, at the same time the excavation plane must keep stabile.

Comparing with cut and cover method, shield tunneling method have some merits as follows:

1, Less occupied ground surface, less land expropriation cost.
2, The influence of weather condition is small, so there will be longer effective construction time comparatively.
3, Soft soil, sand egg soil, soft rock and the rock strata are all applicable.
4, The construction speed is quick.
5, The influence to the environment is small.

According to the above merits, the shield tunneling method get extensive application.

* We thank the Chinese National Basic Research Program (973 Program) (No.2007CB714006) for the financial support of this investigation.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 453–462, 2009.

1.2 Rotation Posture Rectification

Position and posture of shield machine is important to precision of tunnel construction during excavation. But there must have deviation from actual position and posture to desired position and posture. So the rectification of the deviation is necessary.

In this paper, we do not discuss position rectification. In posture rectification, only rotation rectification is considered.

Posture of shield machine is described by three parameters: the pitching angle θ_p, the yawing angle θ_y and the rotation angle θ_r. The pitching angle θ_p is the angle between shield machine axis and horizontal plane. The yawing angle is the angle between shield machine and vertical plane through the shield machine axis. The rotation angle is the angle that the shield machine rotates on the shield machine axis. In the following description, replace θ_p with θ. To illustrate θ, the following coordinate system is established.

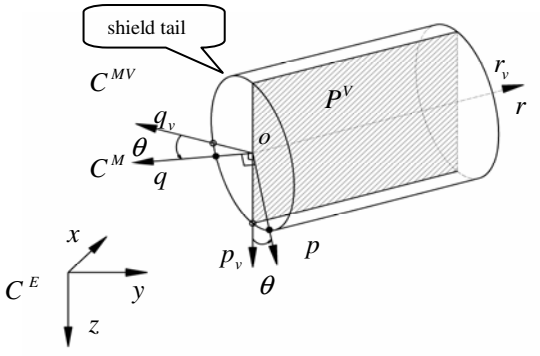


Fig. 1. Coordinate systems are established

The global coordinate system C^E is selected so that the z axis is vertically downwards and the x and y axes are on a horizontal plane. The x, y and z axes are perpendicular to each other and follow the right-hand rule.

A machine coordinate system C^M is selected so that the p axis is vertically downward when the shield being not rotate and the r axis is in the direction of the machine axis. The C^M consolidate with the shield. The origin of the machine coordinate system is selected at the center of the shield tail plane. The p, q and r axes are perpendicular to each other and follow the right-hand rule.

A reference coordinate system C^{MV} is selected so that the p_v axis is through the cross-line of P^V and shield tail plane and the r_v axis is in the direction of the machine axis. The p_v and q_v axes are on the shield tail plane. The P^V is the vertical plane through shield axis. The origin of the reference coordinate system is selected at the center of the shield tail plane. The p_v, q_v and r_v axes are perpendicular to each other and follow the right-hand rule.

The angle between p axis and p_v axis is the rotation angle θ. The θ is positive for counterclockwise rotation of shield (view from shield tail).

Shield rotation posture rectification means zero θ when θ is not zero.

During excavation, if θ is not zero, the actual jack thrust distribution will be different from theoretical jack thrust distribution. That will lead to algorithm deviation. More serious condition is that θ is over certain threshold which will lead to propulsion failure and segment assembly failure.

1.3 Previous Research

In literature [1], the silt sandy ground in Shanghai area is taken as prototype, then different parameters of shield machine and ground properties be combined for testing. Based on the test results, studies are carried out to discover the variation law of thrust force, the mechanism of friction between soil and shield and its influence factors. In literature [2], soil cutting force was investigated and a soil cutting force model was proposed. In literature [3,4,5], the reasonable calculation method of thrust force and its influencing factors were studied by applying mathematics and mechanics based on the working mechanism of EPB shield machine.

In order to research on shield rotation rectification, a shield rotation moment model must be established. The previous research is helpful to establish the model, but no specialized research was carried on.

2 Rotation Rectification Strategy

The rotation rectification strategy includes four steps: rotation angle measurement, cause analysis, need to rotation rectification or not, rotation rectification. Rotation

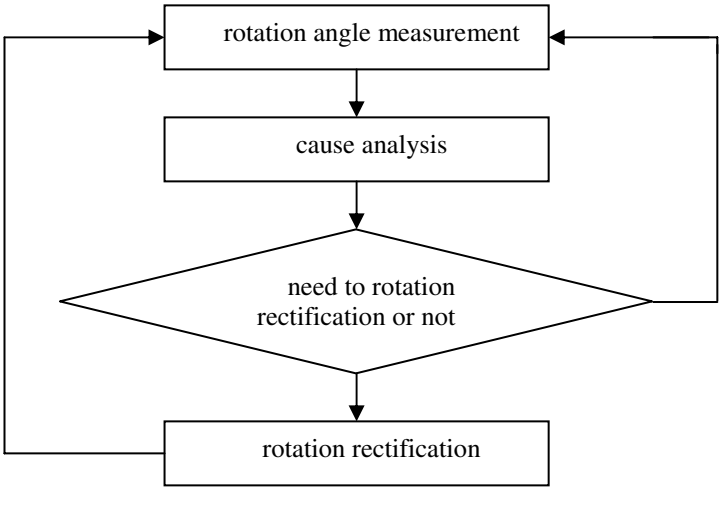


Fig. 2. Rotation rectification steps

angle measurement can be measured by an inclinometer in the shield. Based on the shield rotation moment model, the reasons that cause the current rotation situation could be found out. If θ>1.5°, rotation rectification by counter rotation of cutter is needed.

When θ_p and θ_y rectification is proceeding, rotation rectification is not carried on. When the shield machine is in curve alignments, rotation rectification is also not carried on.

To investigate the reasons that cause the rotation situation, the shield rotation moment model must be established.

3 Modeling of the Shield Rotation Moment

The loads acting on the shield are composed of four forces: force due to the self-weight of the shield F_1, force due to the jack thrust F_2, force acting at the face F_3, and force acting on the shield periphery F_4, as illustrated in Fig. 3. Corresponding to the forces, the moment around the shield axis are T_1, T_2, T_3 and T_4 respectively.

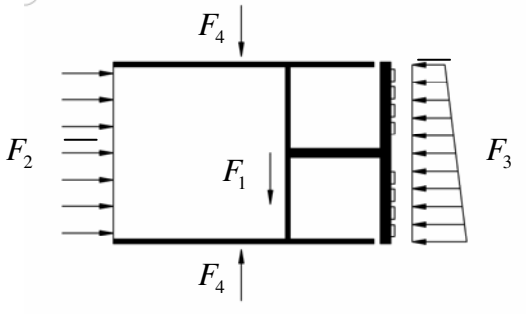


Fig. 3. Shield loads

3.1 T_1 Corresponding to the Shield Self-weight F_1

The point of the self-weight is on the shield axis, so there is not any moment around shield axis.

That means $T_1 = 0$.

3.2 T_2 Corresponding to the Jack Thrust F_2

The force F_2 is composed of jack thrust F_{21} and friction on the interface between the jacking plate and the segment F_{22}. The direction of F_{21} is parallel to shield axis, so there is not any moment around shield axis. The force F_{22} produces moment around shield axis.

$$F_{21}{}^{M} = \begin{bmatrix} 0 \\ 0 \\ \sum_{i=1}^{n} f_{ji} \end{bmatrix} \tag{1}$$

$$F_{22}{}^{M} = \begin{bmatrix} \mu_{js} sign(CF) \sum_{i=1}^{n} \alpha_{jsi} f_{ji} \cos\alpha_i \\ \mu_{js} sign(CF) \sum_{i=1}^{n} \alpha_{jsi} f_{ji} \sin\alpha_i \\ 0 \end{bmatrix} \tag{2}$$

$$F_{2}{}^{M} = F_{21}{}^{M} + F_{22}{}^{M} = \begin{bmatrix} \mu_{js} sign(CF) \sum_{i=1}^{n} \alpha_{jsi} f_{ji} \cos\alpha_i \\ \mu_{js} sign(CF) \sum_{i=1}^{n} \alpha_{jsi} f_{ji} \sin\alpha_i \\ \sum_{i=1}^{n} f_{ji} \end{bmatrix} \tag{3}$$

where α_{js} is the ratio of resistance to friction between jack and segment[6,7]; μ_{js} is the coefficient of mobilized friction between jack and segment; f_{ji} is the *i*th jack thrust force; α_i is the angle between the *i*th jack and q axis; $sign(CF)$ a positive sign for counterclockwise rotation of cutter face (viewed from shield tail).
Then

$$T_2 = r\alpha_{js}\mu_{js} sign(CF) \sum_{i=1}^{n} f_{ji} \tag{4}$$

Where r is the radius of the jack from the *r* axis.

3.3 T_3 Corresponding to the Force Acting at the Shield Face F_3

$$F_3 = F_{31} + F_{32} + F_{33} + F_{34} \tag{5}$$

Where F_{31} is the force due to the earth pressure acting on the panel of cutter face; F_{32} is the force due to the muck in the chamber; F_{33} is the force due to the earth pressure acting on the edge of the cutter face and F_{34} is the cutting force.

Fig. 4. The force acting at the shield face

3.3.1 The Force Due to the Earth Pressure Acting on the Panel of Cutter Face F_{31}

According to soil mechanics[8,9]:

$$\sigma_{CP} = K_0 \gamma h \tag{6}$$

Where σ_{cp} is the earth pressure on the cutter face panel; K_0 is the horizontal earth pressure coefficient; γ is soil bulk density; h is the vertical depth of the integrated area of cutter face from the ground surface.

$$h = H - r\sin\alpha \tag{7}$$

Where r is the radius of cutter face; H is the vertical depth of the center of cutter face from the ground surface.

Then the normal earth pressure F_{31}^r

$$F_{31}^r = -(1-\eta)\int_0^{2\pi}\int_0^{D/2} K_0\gamma(H - r\sin\alpha)drd\alpha = -(1-\eta)\frac{\pi D^2}{4}K_0\gamma H \tag{8}$$

Where η is the cutter face open rate; μ_{ms} is the coefficient of mobilized friction between muck and shield; D is the cutter face diameter.

Then

$$T_{31} = -\mu_{ms}(1-\eta)\frac{\pi D^2}{4}K_0\gamma H sign(CF) \tag{9}$$

3.3.2 The Force Due to the Soil in the Chamber F_{32}

The forces due to the soil in the chamber concerning to the shield rotation include friction forces between the soil with the bulk, with shield cylindrical inner face and

with the rear of cutter face. In this paper, only the friction force between the muck with the rear of cutter face is taken into account.

Then

$$T_{32} = -\mu_{ms}(1-\eta)\frac{\pi D^2}{4}K_0\gamma H sign(CF) \tag{10}$$

3.3.3 The Force Due to the Earth Pressure Acting on the Edge of the Cutter Face F_{33}

$$K_\alpha = (1-K_0)\left|\sin\alpha\right| + K_0 \tag{11}$$

Then

$$T_{33} = -rl_{ce}\mu_{ms}sign(CF)\int_0^{2\pi}((1-K_0)\left|\sin\alpha\right| + K_0)(H - r\sin\alpha)d\alpha \tag{12}$$

Where l_{ce} is the cutter face thickness.

3.3.4 The Cutting Force F_{34}

$\sum f_{cyi} = 0$, that is

$$N_0\cos\alpha - \mu_0 N_0\sin\alpha + N_1\cos\theta_c - (cS_c + \mu_1 N_1)\sin\theta_c = 0 \tag{13}$$

$\sum f_{cci} = 0$, that is

$$N_0\sin\alpha + \mu_0 N_0\cos\alpha - N_1\sin\theta_c - (cS_c + \mu_1 N_1)\cos\theta_c = 0 \tag{14}$$

Thereby N_0 N_1 are obtained.

Then the cutting force of one knife F_t is,

$$F_t = N_0\cos\alpha - \mu_0 N_0\sin\alpha = \frac{cS_c}{\dfrac{\cos\alpha - \mu_0\sin\alpha}{\sin\alpha + \mu_0\cos\alpha} + \dfrac{\cos\theta_c - \mu_1\sin\theta_c}{\sin\theta_c + \mu_1\cos\theta_c}} \tag{15}$$

Where S_c is the soil rupture area.

$$S_c = \frac{P}{\sin\theta_c}l_e \tag{16}$$

Where P is the cutting depth of specific cutter, l_e is the width of cutter.

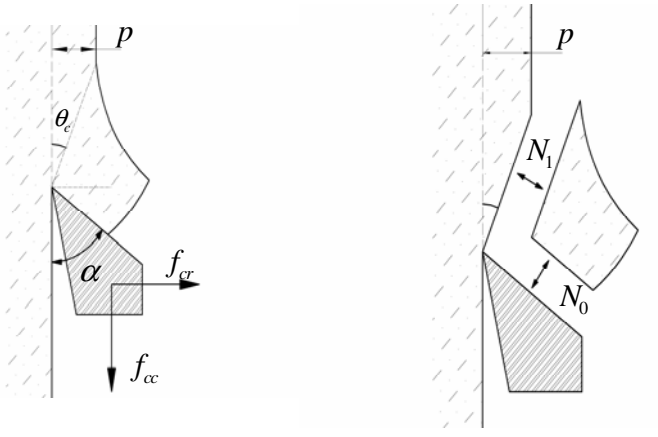


Fig. 5. The cutting force model

$$\theta_c = \frac{1}{2}(90 - \varphi) \tag{17}$$

Where φ is the internal friction angle of the soil.

$$p = \frac{v}{n \times m} \tag{18}$$

Where v is the shield propulsion speed, n is the cutter face rotation velocity, m is the number of cutter at the same assembly radius.

$$T_{34} = \sum a_i F_{ti} r_i \tag{19}$$

Where r_i is the ith cutter assembly radius. a_i =1 for active cutter and 0 for inactive cutter.

3.4 T_4 Corresponding to the Force Acting on the Shield Periphery F_4

As to F_4, a simplified method is introduced.

$$F_4 = \mu_{ms} DLf \tag{20}$$

Where D is the diameter of the shield, L is the length of the shield, f is the fiction on unit shield periphery area.

$$f = \mu_{ms}\left(\frac{u}{\pi D} + \frac{\sigma_v + \sigma_h}{2}\right) \tag{21}$$

$$\sigma_v = \gamma H \tag{22}$$

$$\sigma_h = \gamma(H + D/2)\tan^2(45^o - \varphi/2) \tag{23}$$

Equilibrium that keep the Rotation balance

$$T_4 = r\alpha_{ms}\mu_{ms}sign(CF)DLf \tag{24}$$

where α_{ms} is the ratio of resistance to friction between shield and soil around shield.

4 Conclusion

Concerning $T_1 = 0$, the equilibrium that keep the Rotation balance is

$$T_2 + T_3 + T_4 = 0 \tag{25}$$

The hindering moments of shield rotation are T_2 and T_4. The moment that causing shield rotation is T_3. T_2 and T_4 are passive moment, and T_3 is positive moment.

Under normal circumstance, the moments that provided by T_2 and T_4 are much larger than T_3, so the shield keep un-rotation. T_2 is proportional to jack thrust force. T_4 is relative to characteristic of soil around the shield. T_3 is relative to characteristic of soil front of the shield, shield machine propulsion speed and cutter face rotation velocity. T_3, especially T_{34}, changes acutely with the actual working conditions. With the increasing of the T_{34}, the possibility of shield rotation goes high. For instance, cutting large stones, excavating from sand to clay and increasing the propulsion speed will increase T_{34}.

5 Rotation Rectification Practice

After the measurement of inclinometer, if $\theta > 0.5^o$ and $\theta < 1.5^o$, then rotation rectification could not be carried on. In this situation, the reason causing the rotation deviation must be concerned. If the rotation is due to cutting the large stones in front of the shield, then the excavation parameters should remain unchanged. If the rotation is due to that the shield is in the process from sand to clay, then the ratio of cutter face rotation velocity to shield propulsion speed should be increased to prevent further rotation.

If $\theta > 1.5^o$, then the rotation rectification should be carried on. If the rotation is due to cutting the large stones in front of the shield, then the excavation parameters should remain unchanged and rotation rectification by counter rotation of cutter is needed. If

the rotation is due to that the shield is in the process from sand to clay, then the ratio of cutter face rotation velocity to shield propulsion speed should be decreased and rotation rectification by counter rotation of cutter is needed. After the rectification is done, the ratio of cutter face rotation velocity to shield propulsion speed should be adjusted to prevent over rectification.

In curve alignments, the friction between shield and the soil around the shield increase, so the possibility of rotation gets smaller. But when the shield machine is in curve alignments, rotation rectification is not carried on, so the inclinometer values should be observed closely the inclinometer values, and the excavation parameters should be kept well, to prevent over rotation.

Acknowledgments. We thank the Chinese National Basic Research Program (973 Program) (No.2007CB714006) for the financial support of this investigation.

References

1. Zhu, H.H., Xu, Q.W.: Experimental study on thrust force during tunnel excavation with earth pressure balance shield machine. Rock and Soil mechanics 28, 1587–1594 (2007)
2. Song, K.Z., Pan, A.G.: Operation principle analysis of cutting tools on shield. Construction Machinery, 74–76 (2007)
3. Xu, Q.W., Zhu, H.H.: Analysis of reasonable thrust force during tunnel excavation in homogeneous soft ground by use of earth pressure balance shield machine. Chinese Journal of Geotechnical Engineering 30, 79–85 (2008)
4. Su, J.X., Gong, G.F., Yang, H.Y.: Total Thrust Calculation and Test Research for Soil Pressure Balancing Shield Tunneling. Construction Machinery and Equipment 39, 13–16 (2008)
5. Xu, D.K., He, W.H.: The Experiment and Research of Shield Force and Lining friction under Soft Foundation, vol. 17, pp. 30–33 (1995)
6. Sugimoto, M., Sramoon, A.: Theoretical Model of Shield Behavior During Excavation I: Theory. Journal of Geotechnical and Geoenvironmental Engineering 128(2), 138–155 (2002)
7. Sramoon, A., Sugimoto, M., Kouji, K.: Theoretical model of shield behavior during excavation II: Application. Journal of Geotechnical and Geoenvironmental Engineering 128(2), 156–165 (2002)
8. Xie, X.Y., Liu, F.C.: Soil Mechanics (2005)
9. Gao, D.Z., Jia, J.Y.: Soil Mechanics and Soil Science (2000)

Compensation of Measurement Error for Inclinometer Based on Neural Network

Xiangwen Wen, Haiyang Cai, Minghua Pan, and Guoli Zhu

State Key Laboratory of Digital Manufacturing Equipment and Technology,
Huazhong University of Science & Technology,
Wuhan, 430074, China
glzhu@mail.hust.edu.cn

Abstract. In the shield tunneling construction, the inclinometer is usually utilized to measure the pitching angle and rolling angle of the shield machine in real-time. However, the nonlinearity and the temperature characteristic of the inclinometer always result in large measurement error. In order to improve its measurement accuracy, the research on compensating the nonlinear error and the temperature drift error based on BP neural network is presented in this paper. The characteristic of the inclinometer is studied by experiment at first and then its inverse model is built using BP neural network and trained with an amount of experimental data; finally the trained model is used to compensate the measurement error. The experimental results verify that the proposed compensation method can improve the measurement accuracy of the inclinometer greatly by correcting the nonlinearity and eliminating the influence of temperature.

Keywords: Inclinometer, Shield Machine, Nonlinear Error, Temperature Drift Error, BP Neural Network.

1 Introduction

Shield tunneling method has been applied to the subway tunnel construction increasingly in recent years due to its many advantages, such as rapid construction speed, security, good quality, slight impact on the environment and so on. In order to guarantee the construction quality, it is necessary to determine the position and posture of the shield machine accurately in real-time so that the shield machine can be driven ahead along the DTA (Designed Tunnel Axis). Therefore, it puts forward a high requirement on the measurement accuracy of the attitude angle in engineering. The results of the simulation experiment in MATLAB showed that the measurement error of the instrument must be within 0.8mrad in order to ensure the measurement error of each attitude angle is less than 1mrad.

As an important component of attitude measurement system for the shield machine, the inclinometer is of high resolution, high performance-price ratio, low power consumption and etc. It is mainly used to measure the pitching angle and the rolling angle of the shield machine. Therefore, the measurement accuracy of the inclinometer

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 463–473, 2009.

affects that of the two attitude angles directly. However, it is discovered by experiments that the available inclinometer can satisfy the accuracy requirement only when the measuring angle is smaller and its the measurement error will become larger obviously with the increasing of the measurement range. Besides, its measurement accuracy is susceptible to the variation of operating temperature. Thus, the measurement error of the inclinometer must be compensated effectively so as to meet the required measurement accuracy of the attitude angle in engineering. Nowadays, there are mainly two kinds of methods of error compensation for sensors: hardware improvement method [1] and software compensation method. The former can decrease the measurement error of sensors with additional compensation circuit. However, it also brings about many problems, such as complex circuit, poor universality, low precision but high cost and so on. Hence, it is seldom adopted in the practical applications. There are many methods about software compensation, such as correction method of nonlinear inverse function [2], polynomial curve fitting approach based on least square method [3-4] and etc. There is no theoretical error to the correction method of nonlinear inverse function, but it is rather difficult to obtain the nonlinear inverse function accurately in practical applications. Moreover, the nonlinear factor is the only consideration for this method, so it can't get high compensation accuracy. As a result, its application is also restricted. Generally speaking, the polynomial curve fitting approach based on least square method can acquire higher compensation accuracy, but the nonlinear factor and the temperature effect must be considered separately in order to find all coefficients of the fitting polynomial. If other factors are needed to consider, such as vibration, it will be extremely inconvenient to get the coefficients. In addition, when there are too many calibration experiment data, it is easy to emerge the phenomena of oscillatory and unable to obtain the coefficients of the fitting polynomial. In a word, it isn't the best way to compensate the measurement error of the inclinometer, either.

ANN (Artificial Neural Network) is a kind of new information processing system, which is built on the basis of imitating the structure and function of human brain. Actually, it is a complex network with numerous simple elements interconnecting. The inherent features of ANN are as follows: distributed storing information; parallel and collaborative processing information; nonlinear mapping; strong abilities for self-organizing, self-adapting and self-learning. Therefore, the neural network can show its superiority when dealing with the complex problems, especially when the internal law of the problem is difficult to know and can't be described with an exact mathematical expression, or many factors and conditions are needed to consider simultaneously, or the task requires having fault tolerance, it is always the most appropriate treatment means. In fact, the neural network has shown its incomparable advantages in the aspect of error compensation [5-9].

Therefore, a novel method is proposed in this paper to compensate the measurement error by setting up an inverse model of the inclinometer and then training it with a mass of calibration experiment data based on BP neural network. The experimental results show that this method not only can correct the nonlinear error effectively, but also can compensate the temperature drift error well.

2 Experiment of Measuring Angle with the Inclinometer and the Analysis of Measurement Error

2.1 Introduction to the Experiment

In order to study the nonlinearity and the temperature characteristic of the existing inclinometer (model: SST260, resolution: 0.001°, measurement accuracy: ±0.05°, temperature drift: ±0.1°) and examine the compensation effect of measurement error, an experiment of measuring angle with the inclinometer is carried out in this paper. As shown in Fig.1, the experimental device mainly includes a dual-axes inclinometer with a temperature sensor inside and a high precision DD (direct drive) servo motor (model: SGMCS-05B3B11, absolute accuracy: ±15 second, repeatable accuracy: ±1.3 second) with an absolute pulse encoder in the rear. In addition, a heat source is adopted to change the operating temperature in a wide range by heating the inclinometer during the experiment. When the rotating control pulse is sent out by the computer, the inclinometer will measure its inclination relative to the horizontal plane after rotating a certain angle together with the motor shaft under the action of the drive unit and the temperature sensor will measure the operating temperature of the inclinometer at the same time. The inclination, the operating temperature and the digital code corresponding to each position of the motor shaft are transmitted to the computer through the data lines. At different inclinations and temperatures, it is requisite to measure the inclination of dual-axes separately relative to the horizontal plane so as to correct the measurement error respectively. Besides, in order to avoid the installation error, it is necessary to make sure Y axis of the inclinometer is parallel to the motor shaft when measuring the inclination of X axis relative to the horizontal plane, vice versa.

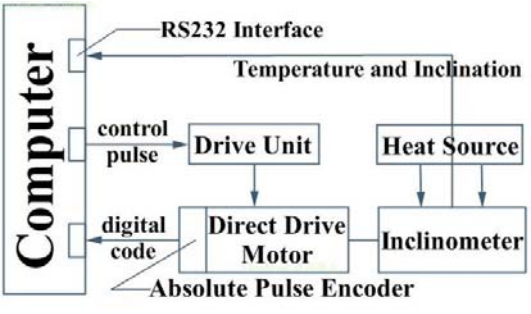


Fig. 1. Schematic diagram of the experimental device

2.2 Analysis of Measurement Error for the Inclinometer

Take the experiment of measuring the inclination of Y axis relative to the horizontal plane for example, the inclination is set to change in the range of ±45° for the purpose of analyzing the nonlinearity of the inclinometer fully and it is nearly selected by the

same interval of 5°. In addition, the calibration operating temperature of the inclinometer increases gradually at intervals of 0.92°C approximately. For each measured value of the inclinometer, the true inclination needs to be determined in real-time so as to obtain the measurement error. In this paper, the true inclination can be obtained by calibrating the zero point (It means to determine the digital code of the absolute pulse encoder when the inclinometer is in level condition in the rotary process) and then performing the unit conversion. After eliminating the gross error and decreasing the random error of the experimental data, the nonlinear error curves of the inclinometer at different calibration operating temperatures are illustrated in Fig.2. The nethermost curve represents the nonlinear error of the inclinometer when the operating temperature is 27.70°C and the uppermost one corresponds to 46.19°C.

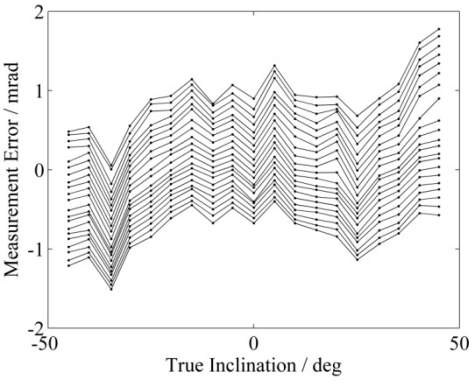


Fig. 2. Nonlinear error of the inclinometer at different calibration operating temperatures

It can be seen from the Fig.2 that the measurement error of the inclinometer isn't large generally when the true inclination is smaller. However, the nonlinear characteristic of the inclinometer itself will lead to large errors once the measurement range extends. Besides, the operating temperature also has a great impact on the measurement accuracy of the inclinometer so that the nonlinear error curves are of obvious drift with the rising of the operating temperature. Considering the complexity of the nonlinear error curves for the inclinometer, if they are expressed with a fitting polynomial as follows:

$$\varepsilon = C_0 + C_1 \times y + C_2 \times y^2 + \cdots\cdots C_n \times y^n \tag{1}$$

where ε denotes the nonlinear error, y represents the output of the inclinometer and C_i is the coefficient of the fitting polynomial. The order of the fitting polynomial not only is needed to select appropriately, the relationship between each coefficient and the temperature but also should be considered, that's to say, each coefficient C_i is a function of the temperature, too.

$$C_i = D_0 + D_1 \times t + D_2 \times t^2 + \cdots\cdots D_m \times t^m \tag{2}$$

where t implies the operating temperature. In other words, there is an interaction between the nonlinear error and the temperature drift error. If the polynomial curve fitting approach based on least square method is utilized to find a fitting function between the error and the output of the inclinometer together with the temperature, it must fit the test data for several times to obtain all coefficients. Besides, the compensation accuracy of this method is not satisfying.

Hence, in order to improve the measurement accuracy of the inclinometer as high as possible, it is essential to find a new and convenient method to compensate the nonlinear error and the temperature drift error effectively.

3 Error Compensation Based on BP Neural Network

3.1 Principle of Error Compensation

As stated above, the measurement model of the inclinometer when applied in practice can be expressed as

$$p = f(y,t) \tag{3}$$

where t is the operating temperature of the inclinometer, y and p are the input (true inclination) and the output of the inclinometer separately and the function $f(\cdot)$ is a unknown nonlinear function with two variables. The goal of error compensation is hoping to obtain the unknown input of the inclinometer according to its output at any temperature, thus the inverse function of the function $f(\cdot)$ needs to be acquired generally. However, the exact mathematical expression of the function $f(\cdot)$ is extremely difficult to get, let alone its inverse function.

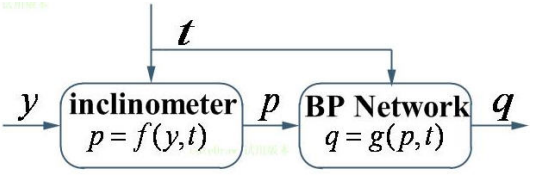


Fig. 3. Schematic diagram of principle of error compensation

It is well-known that BP neural network can realize any nonlinear mapping from the input space to the output space as long as providing sufficient input and output data for its training even though without prior understanding about the mathematical equation of the mapping relationship. Therefore, an inverse model of the inclinometer is established based on BP neural network, which can be simply described as

$$q = g(p,t) \tag{4}$$

where q represents the actual output of the neural network, the operating temperature t and the output p of the inclinometer are two inputs of the neural network, while the input y of the inclinometer is taken as the expected output of the neural network during the training. In order to achieve the goal of error compensation, the mathematical expression of the function $g(\cdot)$ is unnecessary to know and it is only needed to design an appropriate BP network model as the inverse model of the inclinometer and train it with enough test data until the error between the actual output and the expected output of the neural network is acceptable. Furthermore, the trained BP neural network is of strong generalization capability and it can generate appropriate output for new inputs. (The schematic diagram of principle of error compensation based on neural network is illustrated in Fig.3 above).

3.2 Design of BP Network Model

The design of BP network model mainly contains five aspects as follows:

1) Network layers: BP network can have many hidden layers, but it has been proved that BP network with a hidden layer can realize any nonlinear mapping with any degree of accuracy when the number of nodes in hidden layer is no restriction, so 2 layers is chosen in this paper;

2) The number of nodes in input layer and output layer: as illustrated in Fig.3, it is obvious that the number of nodes in input layer and output layer are 2 and 1 respectively;

3) Transfer function: the transfer function in hidden layer is hyperbolic tangent sigmoid transfer function usually, which expression is as follows

$$f(x) = \frac{1-e^{-x}}{1+e^{-x}} \tag{5}$$

In order to enable the network to output any value, the pure linear function is adopted in output layer and it is

$$h(x) = x \tag{6}$$

4) The feed-forward BP network and the trainable cascade-forward BP network are usually used in practical applications. The difference in their structures is that each layer of the feed-forward BP network only has a weight coming from the previous layer, but the other has weights coming from the input layer and all previous layers. It is discovered by experiments that the feed-forward BP network is more sensitive to the initial value and easy to get into local minimum when there is a large difference between the inputs and the output. In the process of training, the feed-forward BP network often leads to a long training time, slow convergence speed, difficult to achieve the target of error performance, etc. However, the trainable cascade-forward BP network doesn't have above defects and its convergence speed is much faster despite the difference is large or small, so it is adopted in this paper.

5) The number of nodes in hidden layer: there is no theoretical guidance to its chosen but it is rather crucial. If the number is too small, it will result in a poor fault tolerance and generalization capability; conversely, if there are too many nodes in hidden layer, it will lead to a complicated network structure, a long time to train and even emerge the problem of over-fitting (It means the error between the actual output and the expected output of the neural network is very small to training sets but large to new samples), which will reduce the generalization capability instead. So the trial and error method is adopted in this paper. To the same training sets, the network is trained with 2 hidden nodes at first and then increased gradually. Moreover, the training and testing are undertaken alternately. For each case, the network is trained and tested for 200 times and the minimum value of the mean absolute testing error is shown in Fig.4. Obviously, the choice of 12 nodes in hidden layer is the best.

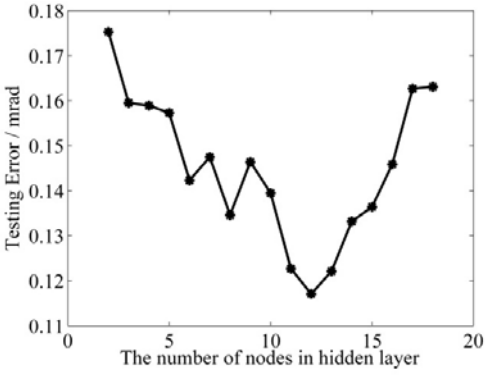


Fig. 4. The minimum value of mean absolute testing error for different hidden nodes in 200 times

To sum up, the designed network model is a trainable cascade-forward BP network with 2 nodes in input layer, 12 nodes in hidden layer and 1 node in output layer, which structure is illustrated in Fig.5.

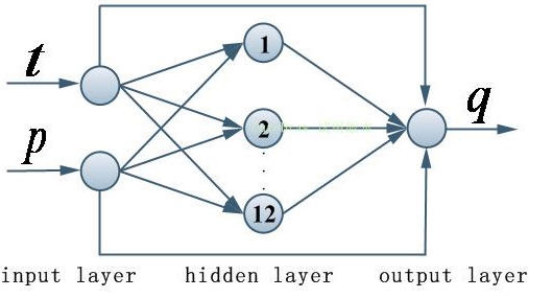


Fig. 5. The designed BP network model

3.3 Training of BP Network Model

The working mode of neural network involves two phases: training and testing. In the training phase, the weights are adjusted by the learning rule so as to minimize the error target function, which can improve the performance of the network. In the testing phase, the weights remain unchanged and the trained network can generate appropriate output according to the inputs.

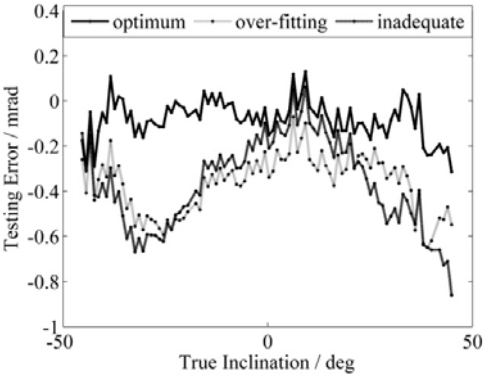


Fig. 6. Testing error at different training times

In order to guarantee the network has a good generalization capability, there are two aspects required to consider:

1) The definition of the training sets: as we know, the neural network is only of the characteristic of internal interpolation instead of external interpolation. In other words, the generalization capability of the neural network is merely effective to these new data which is in the range of training sets. If the new data exceeds the extent of training sets, the neural network will produce large output error probably. In order to avoid the problem, the experimental scope should be as wide as possible. In this paper, the inclination of training sets can vary from 45° to 45° and the range of the temperature is from 20.31°C to 48.94°C, which variation intervals are as stated in chapter 2.2.

2) The determination of training times: the performance of a network model is evaluated by its generalization capability, not its training error, so it is essential to test the trained network with new samples. In fact, even though the number of nodes in hidden layer is appropriate, the network may also emerge the problem of over-fitting due to overtraining, which will lead to a poor generalization capability. Generally speaking, with the training times increased both the training error and the testing error will reduce at first. However, after surpassing a certain extent, the training error will reduce continually but the testing error will increase instead. In other words, the generalization capability of a neural network has a trend that it will strengthen gradually at first but weaken later with the training error reducing constantly, which can be illustrated in Fig.6. Therefore, it can't be thought simply that the smaller the training

error is, the better the generalization capability will be. It is necessary to ascertain the training times when the testing error is minimum through experiments so as to obtain a good generalization capability.

4 Results of Error Compensation

In order to verify the validity and feasibility of the proposed method, three sets of new experimental data which are different from the training sets but within its extent are separately used to compensate the measurement error of the inclinometer under different circumstances: (1) Compensation for the nonlinear error; (2) Compensation for the temperature drift error; (3) Compensation for the composite error.

Fig.7 shows the measurement error curves of the inclinometer before and after compensation when the calibrated operating temperature remains unchanged at

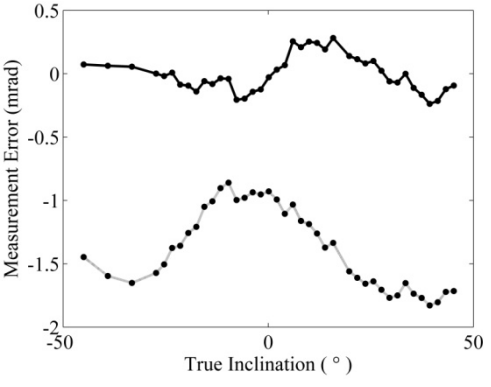


Fig. 7. Compensation result of the nonlinear error

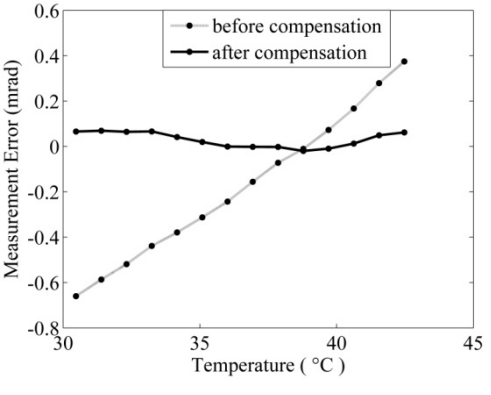


Fig. 8. Compensation result of the temperature drift error

21.24°C and the true inclination changes from 45° to 45° at interval of 2° approximately. (The solid curve denotes the measurement error after compensation by BP neural network) Compared with the measurement error before compensation, the compensation effect of BP neural network is pretty good apparently and the largest error reduces from 1.8mrad to 0.3mrad.

In the experiment of the inclinometer, when the true inclination is 10.06° constantly and the operating temperature increases gradually from 30.47°C to 42.47°C at intervals of 0.92°C, its temperature drift error is illustrated in Fig.8. Obviously, BP neural network can eliminate the impact of temperature variation on measurement effectively by error compensation.

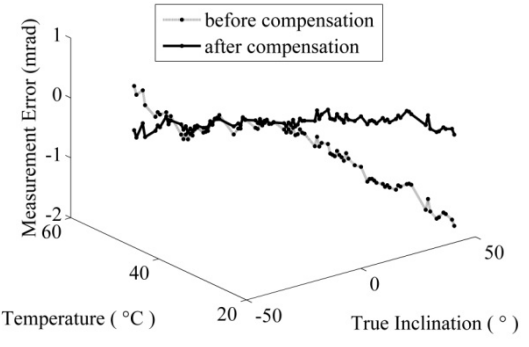


Fig. 9. Compensation result of the composite error

In the shield tunneling construction, both the attitude angle of the shield machine and the operating temperature of the inclinometer change all the time, so there exist the nonlinear error and the temperature drift error in the output of the inclinometer inevitably. In order to simulate the true process of measurement, a set of test data in which the inclination varies gradually from 44.92° to 45.33° and the temperature also rises gradually from 22.16°C to 48.01°C is utilized to compensate the composite error of the inclinometer. Compared with the true inclination, the composite error curves before and after compensation are shown in Fig.9. It can be found that the composite error of the inclinometer is very large before compensation, but it becomes much smaller after compensation by BP neural network and the largest one declines from 1.8mrad to 0.3mrad.

5 Conclusions

(1) Compared with other methods, the error compensation method based on BP neural network can compensate the nonlinear error and the temperature drift error of the inclinometer simultaneously. That's to say, the nonlinear factor and the temperature influence needn't be considered respectively. What's more, it can obtain the highest compensation accuracy and enable all measurement error to be under 0.3mrad, thus it

can be used to improve the measurement accuracy of the inclinometer in the shield tunneling construction.

(2) The range of the training sets and the determination of the training times have great impacts on the neural network's generalization capability.

(3) The trainable cascade-forward BP network is superior to the feed-forward BP network as a whole, especially when there is a large difference between the inputs and the output of the neutral network.

Acknowledgment

This work is supported by the National Basic Research Program of China (No. 2007CB714000).

References

1. Zheng, Y., Peng, S., Jing, Y., Huang, X., Zhang, F.: Method of eliminating influence of environmental temperature on tilt sensor property. J. Piezoelectrics & Acoustooptics. 30(1), 33–35 (2008)
2. Li, D., Zhao, X.: An inverse function correction method for nonlinearity of sensors. J. Chinese Journal of Scientific Instrument 12(2), 215–218 (1991)
3. Sun, H., Guo, Z.: Technology of error compensation on sensors. J. Chinese Journal of Sensors and Actuators 1, 90–92 (2004)
4. Qu, G.: On-line comprehensive compensation of piezoresistive pressure sensor. J. Journal of Transducer Technology 22(8), 45–46 (2003)
5. Liu, J., Yang, L., Cheng, T.: Nonlinear errors correction of ultrasonic wave measure distance system based on neural network. J. Microcomputer Information 1, 187–188 (2007)
6. Zhu, Q.: Sensor's characteristics linearization using artificial neural networks and best-fitting. J. Control Theory &Applications 20(1), 66–69 (2003)
7. Tian, S., Zhao, Y., Wei, H., Wang, Z.: Nonlinear correction of sensors based on neural network model. J. Optics and Precision Engineering 14(5), 896–902 (2006)
8. Wang, J.-H., Gao, Y.: A new magnetic compass calibration algorithm using neural networks. J. Measurement Science and Technology 17(1), 153–160 (2006)
9. Zhang, F., Li, T., Zhang, L.: BP Neural Network Modeling of Infrared Methane Detector for Temperature Compensation. In: 2007 8th International Conference on Electronic Measurement and Instruments, pp. 4123–4126. Inst. of Elec. and Elec. Eng. Computer Society, Xian (2007)

Redundantly Actuated PRPRP Radial Mechanism in Segment Erector of Shield Machine for Synchronization Control

Wanghui Bu, Zhenyu Liu*, and Jianrong Tan

State Key Laboratory of CAD&CG, ZheJiang University, HangZhou, 310027, China
Tel.:+86-571-87951273; Fax: +86-571-87951899
{buwanghui,liuzy,egi}@zju.edu.cn

Abstract. The shield machine is a heavy construction machine for tunnel excavation. The segment erector is an important subsystem of the shield machine. The 2DOF five-bar radial mechanism in the 6DOF erector is hard to implement the precise synchronization control. Hence, this paper proposes a redundantly actuated PRPRP radial mechanism for the segment erector. When the redundant actuator is locked or produces enough pretightening stretching force, the PRPRP mechanism can ensure the synchronization of the two driving hydraulic cylinders based on the mechanical structure. The redundant actuator can also make the two flexural torques applied at the hydraulic cylinders be the same, which avoids the overload of the single cylinder.

Keywords: Redundant actuation, synchronization control, five-bar mechanism, segment erector, shield machine.

1 Introduction

The shield machine is a heavy construction machine for tunnel excavation. It contains the cutter subsystem, impelling subsystem, segment erector subsystem, etc. The segment erector is used to safely quickly and precisely put the precast concrete segments to the required locations of the tunnel inwall, when the shield machine is cutting and impelling. The classical segment erector has 3DOF, and can implement the necessary erecting actions, but cannot adjust the pose of the segment [1]. To improve the quality and efficiency of the segment erection, the modern erector has 6DOF [2].

The radial mechanism in the classical erector only has 1DOF, and the two driving hydraulic cylinders move synchronously. There are 2DOF in the radial mechanism of the 6DOF erector, and the two hydraulic cylinders can move both synchronously and differentially. It is noted that, because of the errors and disturbances in the control system, it is hard to implement the precise synchronization control for the 2DOF mechanism. Hence, this paper proposes an improved 2DOF radial mechanism for the erectors. The redundant actuation is utilized to ensure the precise synchronization of the two driving hydraulic cylinders [3][4][5]. The redundant actuator can also make

* Corresponding author.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 474–481, 2009.

the two flexural torques applied at the hydraulic cylinders be the same, which avoids the overload of the single cylinder.

2 Radial Mechanisms in Segment Erectors

The segment erector is a multi-DOF mechanism. To guarantee the safety of the manipulation, the motion of each DOF acts one by one, and there is no coupling motion between the DOF. The typical motion process contains segment gripping, axial translation, circumferential rotation, radial lifting, etc. As for the 6DOF erectors, the motion process also contains the pose adjustment of the segments.

2.1 PRRP Mechanism in the Classical 3DOF Erectors

The classical 3DOF erector cannot adjust the segment pose. Its radial mechanism is shown in Fig. 1. There is only 1DOF, but two driving hydraulic cylinders in the PRRP mechanism. Obviously, these two cylinders should move synchronously. Note that the synchronization of two cylinders is not only implemented by the hydraulic control system, but also ensured by the mechanical structure like *link C* in Fig. 1.

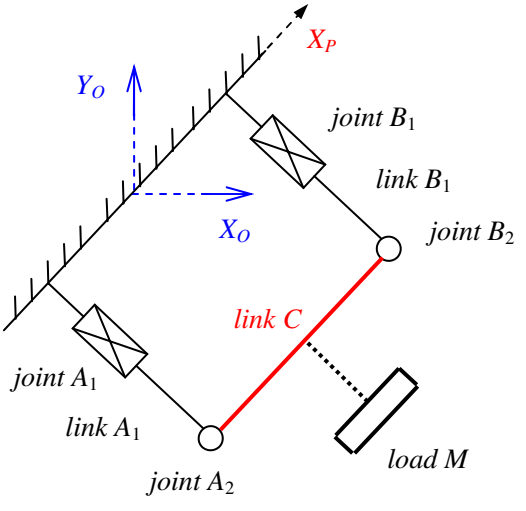


Fig. 1. The PRRP mechanism in a 3DOF erector. The *joint* A_1 and B_1 denote the two driving hydraulic cylinders. The *joint* A_2 and B_2 are two revolute joints. The *load M* denotes the segment.

2.2 PRRRP Mechanism in the Existing 6DOF Erectors

The PRRRP mechanism is adopted in the existing 6DOF erectors to adjust the segment pose, as shown in Fig. 2. There are 2DOF in this mechanism. The two hydraulic cylinders can implement synchronous lifting when the cylinders move synchronously, and can implement pose adjustment when the cylinders move differentially. Note that the synchronization of two cylinders totally depends on the hydraulic control system.

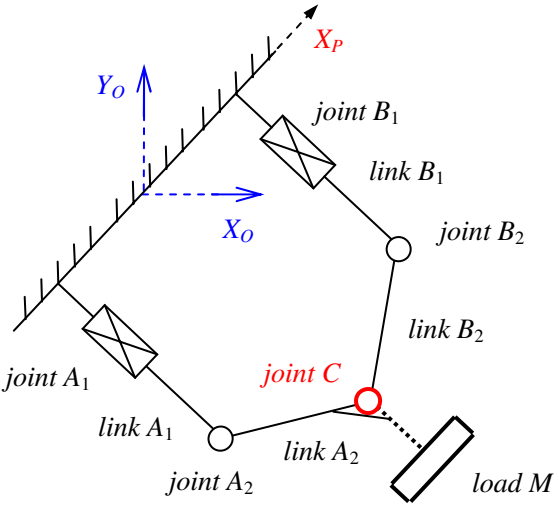


Fig. 2. The PRRRP mechanism in an existing 6DOF erector. The *joint* A_1 and B_1 denote the two driving hydraulic cylinders. The *joint* A_2 and B_2 and C are three revolute joints. The *load* M denotes the segment.

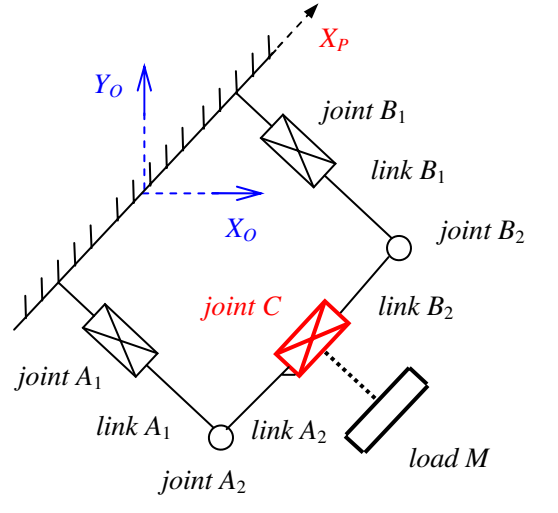


Fig. 3. The redundantly actuated PRPRP mechanism in the improved 6DOF erector. The *joint* A_1 and B_1 denote the two driving hydraulic cylinders. The *joint* A_2 and B_2 are two revolute joints. The *load* M denotes the segment. The *joint* C is the redundant actuator. The length of link A_2 is constant, while the length of link B_2 is changeable.

In the practical control system, no matter what the hydraulic elements such as synchronal valves or advanced controlling strategies are adopted, the precise synchronization of the two cylinders is hard to achieve.

2.3 Redundantly Actuated PRPRP Mechanism in Improved 6DOF Erectors

This paper proposes the redundantly actuated PRPRP mechanism to implement the radial precise synchronization of the two hydraulic cylinders in the 6DOF erectors, as shown in Fig. 3. There are 2DOF in this mechanism, which satisfies the synchronization and pose adjustment in theory. When the redundant actuator (*joint C* in Fig. 3) is locked or produces enough pretightening stretching force, the PRPRP mechanism is equivalent to the PRRP mechanism as shown in Fig. 1, so it can ensure the synchronization of the two driving hydraulic cylinders based on the mechanical structure.

3 Radial Synchronization Control Using Redundant Actuator

When the redundant driving joint C in Fig. 3 is locked or produces enough pretightening stretching force, the distance between joint A_2 and B_2 keeps constantly. If there are some errors or disturbance in the control system, for example if the velocity of joint A_1 exceeds the velocity of joint B_1, then joint A_1 will immediately drag joint B_1 through the rigid links A_2 and B_2, so the synchronization of two hydraulic cylinders is guaranteed.

After the pose adjustment of the segment, the PRPRP mechanism can easily restore its synchronous status. As link B_2 restores its initial length, the radial mechanism will restore the synchronization.

4 Effects of Redundant Actuation on Forces and Torques

The redundant actuator can not only ensure the synchronization, but also make the two flexural torques applied at the driving hydraulic cylinders be the same, which avoids the overload of the single cylinder.

Suppose that m_{A1}, m_{A2}, m_{B1}, m_{B2}, and m_d denote the masses of link A_1, A_2, B_1, B_2, and the segment, respectively. L_{A1}, L_{A2}, L_{B1}, L_{B2}, and L_d denote the lengths of link A_1, A_2, B_1, B_2, and the distance between joint C and the segment, respectively. The segment is combined with link A_2. The angle between axis X_P and X_O in Fig. 3 is θ. g denotes the gravity acceleration. f denotes the redundant driving force of joint C.

Suppose the two hydraulic cylinders perform the synchronous uniform motion. The flexural torque of joint C is

$$M_C = \frac{L_{B2} g \left[L_{A2} \left(2m_d + m_{A2} + m_{B2} \right) \cos\theta + 2L_d m_d \sin\theta \right]}{2\left(L_{A2} + L_{B2} \right)} \tag{1}$$

The flexural torque of joint A_1 is

$$M_{A1} = \frac{L_{A1}}{2} \left[\left(2m_d + m_{A1} + 2m_{A2} \right) g \sin\theta - 2f \right] \tag{2}$$

The driving force of joint A_1 is

$$F_{A1d} = \frac{g\left\{\left[L_{A2}\left(2m_{A1}+m_{A2}\right)+L_{B2}\left(2m_d+2m_{A1}+2m_{A2}+m_{B2}\right)\right]\cos\theta-2L_d m_d\sin\theta\right\}}{2\left(L_{A2}+L_{B2}\right)} \tag{3}$$

The flexural torque of joint B_1 is

$$M_{B1} = \frac{L_{B1}}{2}\left[\left(m_{B1}+2m_{B2}\right)g\sin\theta+2f\right] \tag{4}$$

The driving force of joint B_1 is

$$F_{B1d} = \frac{g\left\{\left[L_{B2}\left(2m_{B1}+m_{B2}\right)+L_{A2}\left(2m_d+2m_{B1}+2m_{B2}+m_{A2}\right)\right]\cos\theta+2L_d m_d\sin\theta\right\}}{2\left(L_{A2}+L_{B2}\right)} \tag{5}$$

If M_{A1}=M_{B1} is required, then

$$f = \frac{L_{A1}\left(2m_d+m_{A1}+2m_{A2}\right)-L_{B1}\left(m_{B1}+2m_{B2}\right)}{2\left(L_{A1}+L_{B1}\right)}g\sin\theta \tag{6}$$

According to the circumferential rotating angle θ and the lengths of link L_{A1} and L_{B1}, the redundant driving force f can be controlled in real time to make the two flexural torques applied at the driving hydraulic cylinders be the same.

Let L_{A2}=L_{B2}=1.2m, L_d=0.3m, m_{A1}=m_{B1}=500kg, m_{A2}=m_{B2}=280kg, m_d=5000kg, g=9.8m/s^2. When L_{A1}=L_{B1}=2m, the variations of torques with θ are shown in Fig. 4, where the blue real line denotes the flexural torques of joint A_1 and B_1, and the red dashed line denotes the flexural torques of joint C. The variations of forces with θ are

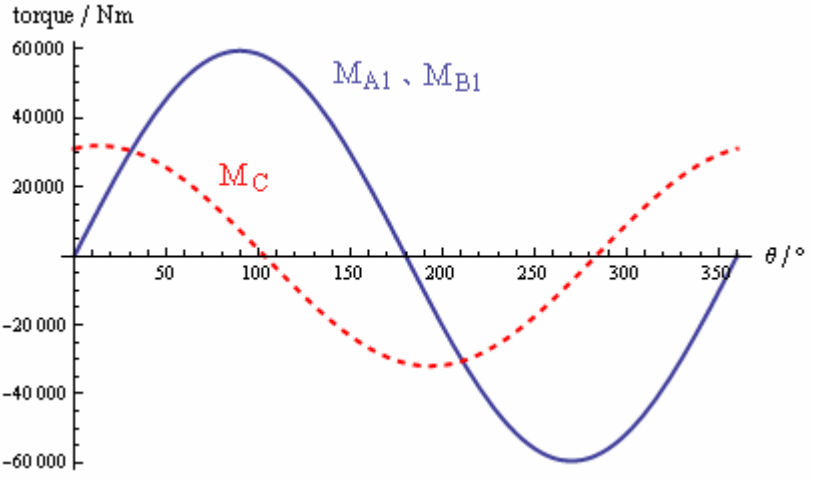


Fig. 4. The variations of torques with θ. The blue real line denotes the flexural torques of joint A_1 and B_1, and the red dashed line denotes the flexural torques of joint C.

shown in Fig. 5, where the red real line denotes driving force of joint A_1; the green dashed line denotes driving force of joint B_1; the blue dotted line denotes the redundant driving force.

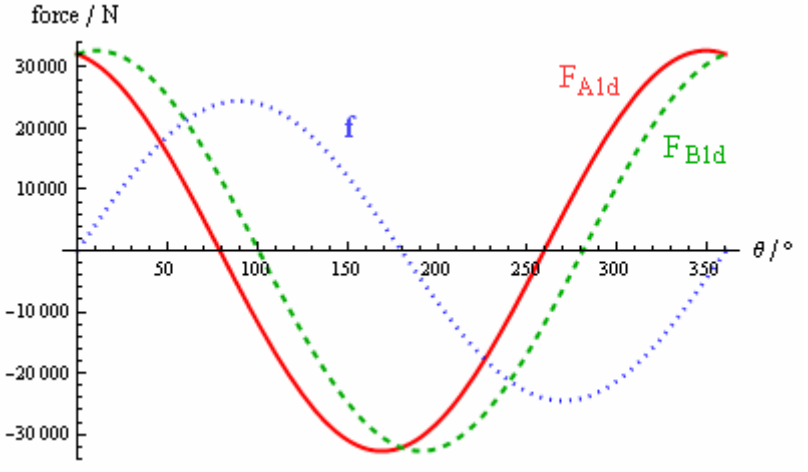


Fig. 5. The variations of forces with θ. The red real line denotes driving force of joint A_1; the green dashed line denotes driving force of joint B_1; the blue dotted line denotes the redundant driving force.

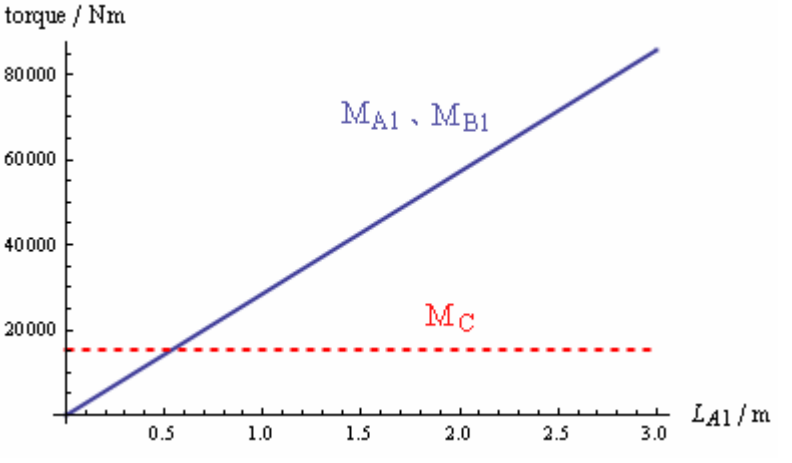


Fig. 6. The variations of torques with L_{A1} (L_{B1}). The blue real line denotes the flexural torques of joint A_1 and B_1, and the red dashed line denotes the constant flexural torques 15379.5Nm of joint C.

When θ= 74.5°, the variations of torques with L_{A1} (L_{B1}) are shown in Fig. 6, where the blue real line denotes the flexural torques of joint A_1 and B_1, and the red dashed line denotes the constant flexural torques 15379.5Nm of joint C. The driving forces of joint A_1, B_1, and C are not variable with L_{A1} (L_{B1}). The driving force of joint A_1; the driving force of joint B_1 equals to 14492.3N; the redundant driving force equals to 23608.9N.

When the acceleration a of radial motion of the erector is considered, the flexural torques of joint A_1 and B_1 are the same with formula (1) and (4), respectively. Hence, if M_{A1}=M_{B1} is required, then the redundant driving force f is the same with formula (6). The flexural torques of joint C is

$$M_{Ca} = M_C - \frac{L_{A2}L_{B2}\left(m_{A2}+m_{B2}\right)}{2\left(L_{A2}+L_{B2}\right)}a \tag{7}$$

The driving force of joint A_1 is

$$F_{A1da} = F_{A1d} - \frac{L_{A2}\left(2m_{A1}+m_{A2}\right)+L_{B2}\left(2m_{A1}+2m_{A2}+m_{B2}\right)}{2\left(L_{A2}+L_{B2}\right)}a \tag{8}$$

The driving force of joint B_1 is

$$F_{B1da} = F_{B1d} - \frac{L_{B2}\left(2m_{B1}+m_{B2}\right)+L_{A2}\left(2m_{B1}+2m_{B2}+m_{A2}\right)}{2\left(L_{A2}+L_{B2}\right)}a \tag{9}$$

According to formula (7)-(9), it is found that, no matter what the circumferential rotating angle θ is, if the two hydraulic cylinders move synchronously (L_{A1}= L_{B1}), the effects of acceleration a on the driving forces of joint A_1 and B_1 and the flexural torque of joint C are only inversely proportional to a.

5 Conclusions

The redundantly actuated PRPRP mechanism can ensure the radial synchronization of the two driving hydraulic cylinders in the segment erector of shield machine based on the mechanical structure. The redundant actuation can also make the two flexural torques applied at the driving hydraulic cylinders be the same, which avoids the overload of the single cylinder. The research indicates that:

(1) An optimum redundant driving force can be obtained to make the two flexural torques applied at the driving hydraulic cylinders be the same. This optimum redundant driving force is only related with the circumferential rotating angle of the erector, and independent of the controlling parameters such as radial displacement or acceleration.

(2) When the circumferential rotating angle is fixed, the two flexural torques applied at the driving hydraulic cylinders are directly proportional to the radial displacement. The driving forces of two cylinders and the flexural torque of the redundant actuated joint are independent of the radial displacement.

(3) When the acceleration of the radial motion of the erector is considered, if the two driving hydraulic cylinders move synchronously, the two flexural torques applied at the cylinders and the optimum redundant driving force are independent of the acceleration. The effects of the acceleration on the driving forces of the cylinders and the flexural torque applied at the redundant actuated joint are only inversely proportional to the acceleration, and independent of the circumferential rotating angle.

Acknowledgements

The support of Project of National Basic Research 973 Program of China (2007CB714007), Project of NSFC (50875239) in carrying out this research is gratefully acknowledged.

References

1. Takashi, W., Toshiharub, T., Tsuyoshic, A., Katsuhiko, M.: Development of a new composite structure segment for large diameter shield tunnel. Tunnelling and underground space technology 19, 449–450 (2004)
2. Qian, X., Gao, F., Guo, W.: Design and analysis of 6-DOF segment erector mechanism for shield machine. Machine design and research 24, 17–20 (2008)
3. Nokleby, S.B., Fisher, R., Podhorodeski, R.P., Firmani, F.: Force capabilities of redundantly-actuated parallel manipulators. Mechanism and machine theory 40, 578–599 (2005)
4. Wu, J., Wang, J., Wang, L.: Optimal kinematic design and application of a redundantly actuated 3DOF plannar parallel manipulator. Journal of mechanical design 130, 054503-1–054503-5 (2008)
5. Garg, V., Nokleby, S.B., Carretero, J.A.: Wrench capability analysis of redundantly actuated spatial parallel manipulators. Mechanism and machine theory 44, 1070–1081 (2009)

Electrohydraulic Control of Thrust Hydraulic System for Shield Tunneling Machine

Beidou Zhu, Guofang Gong, and Hu Shi

Institute of Mechatronics and Control Engineering
Zhejiang University
Hangzhou, China
zhubeidou@hotmail.com

Abstract. A thrust hydraulic system of shield tunneling machine with pressure and flow control is introduced. The electrohydraulic control model is presented. Based on conventional PID control, corresponding experimental analyses are carried out on the simulator test rig. Considering the uncertain and variable loads during shield tunneling, simulation study on fuzzy control of the thrust parameters is realized by applying fuzzy-PID compound control and fuzzy adaptive PID control strategies. Dynamic simulations are finished in AMESim and MATLAB/Simulink softwares. The simulation results show that the control with fuzzy logic algorithm can improve the performances of the thrust system evidently compared with pure PID control, and the fuzzy adaptive PID control system can achieve better performances.

Keywords: Thrust hydraulic system, shield tunneling machine, pressure and flow control, PID, fuzzy.

1 Introduction

With the rapid urban development and the continuous extension of transportation systems, demands of underground space exploration and tunnel construction have been considerably growing. In response to the increasing needs, it is necessary to develop the tunneling technique characterized by reliable security, low cost and high efficiency.

Compared with the old and conventional excavation methods, shield tunneling is a well established method at present. It allows tunneling in a wide range of geological environments with complicated conditions such as high ground water pressure, soft soils with low cover depths, etc. This method features relatively complex construction processes and interactions among the ground, the cutter, the jacks, the tunnel lining and the backfill grouting.

The thrust system, driven by hydraulic power, is a key part of shield tunneling machine. The actuators are composed of several cylinders distributed circularly in the section of shield. The thrust system usually performs the basic task of jacking shield ahead while tunneling and is also responsible for controlling the attitude and position of shield ensuring that the shield can advance along the expected route. Thus the hydraulic jacking control system plays a very important role in shield tunneling [1,2].

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 482–492, 2009.

As for the thrust control of the hydraulic system, it has been investigated in the term of speed and pressure control [3,4]. Allowing for variation and nonlinearity of load, the objective of this paper is to investigate the application of fuzzy control towards thrust hydraulic control system of shield tunneling machine. The experimental test station and the thrust hydraulic system are firstly given. Then, the control design and corresponding simulation study are presented, making comparisons between the conventional PID control and fuzzy-PID compound control.

2 Experimental System

Fig. 1 shows the experimental system of the tunneling shield machine. The test station consists of a simulator box, the shield tunneling machine, load system, hydraulic system and the data acquiring and monitoring system. The simulator box, with an inner diameter of 4m and an axial length of 6m, can be stuffed with a variety of soils which will be pressurized by the bag filled with high pressure water. The water bag loading system can assure the soils of the pressure up to 0.4MPa so as to provide an artificial geo-environment which the tunneling machine will be able to go through.

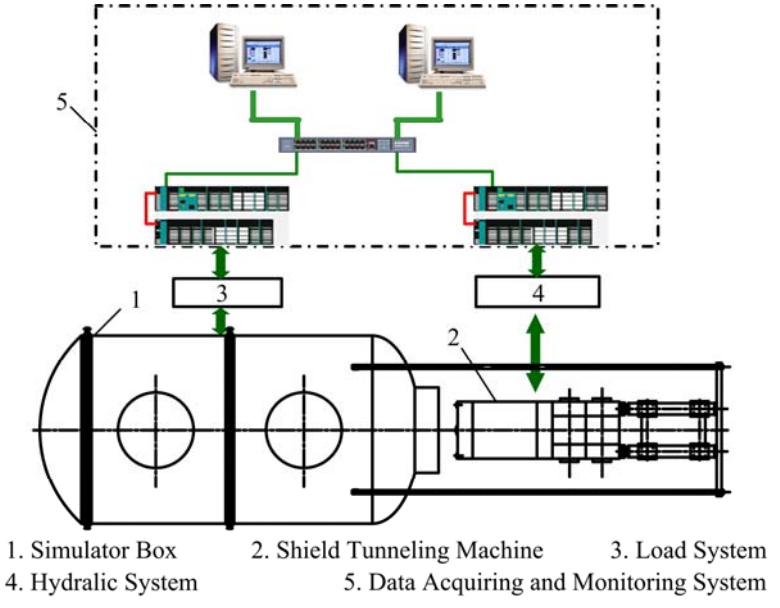


Fig. 1. Experimental System

Fig. 2 shows the schematic of the tunneling machine. It is an EPB (earth pressure balance) shield machine with a screw conveyor discharging the muck cut down by rotating cutterhead with a diameter of 1.8m. The thrust system is composed of six hydraulic cylinders which have the same stroke of 1.5m. The cylinder body is fixed with the backrest while the piston rod is moving to push the tunneling machine

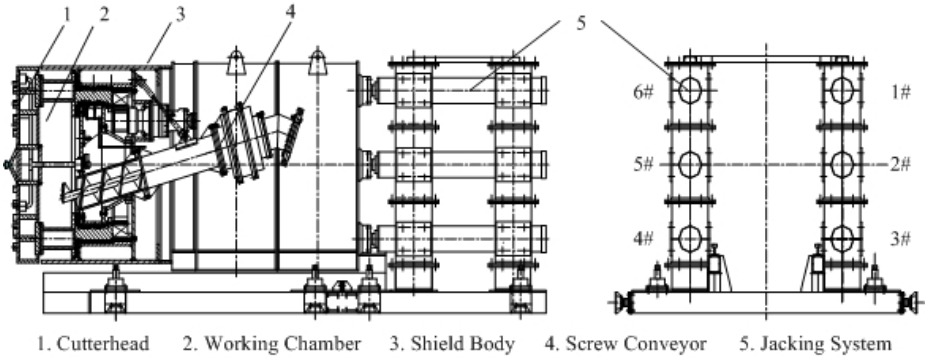


Fig. 2. Experimental Shield Tunneling Machine

forward as shown in Fig. 1. Once the driving distance is beyond the stroke of the cylinder, the tunnel segments are set to relay the jacking process.

2.1 Thrust Hydraulic System

The hydraulic system employed in the experiment for thrusting works is designed according to the proportional pressure and flow regulation principle. The thrust hydraulic cylinders are divided into six identical groups to implement control task, and Fig. 3 shows one group. The fluid flowing into each group is controlled by appropriate input signal of the flow control valve 2 while the working pressure is set by the pressure relief valve 3.

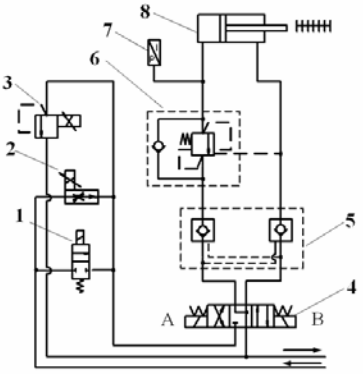


Fig. 3. Thrust Hydraulic System of Shield Tunneling Machine

As shown in Fig. 3, each group comprises a flow control proportional valve 2 and a pressure relief proportional vale 3 to achieve the flow and pressure compound control required in tunneling. The flow rate through flow control valve remains almost invariable as a combination hydrostat maintains a constant level of pressure difference across

the proportional valve, irrespective of system or load pressure changes. Besides, the distributed fluid flow also partly passes through the pressure valve to ensure that the system pressure stays at a constant level. By adjusting the electric current through the coils of the valves, the pressure and flow rate of the system can be regulated to meet the thrust requirements. When tunneling, the solenoid B of 4-way directional valve 4 is energized, shifting the valve to its right position thus making cylinder piston rod move forward. Pressure sensor and the displacement sensor real-timely measure the pressure and displacement of hydraulic cylinder. Measured signals are subjected to online data transmission to the central control system so as to be compared with reference input signals to implement pressure and flow control respectively.

When the shield machine stops for laying the tunnel segments, the hydraulic cylinders must be able to perform the retraction action correspondingly with the 4-way valve 4 working at the left position. Meanwhile, the flow valve is shorted by 2-way directional valve 1 to engender a high flow rate running condition thereby high speed of the rod. There also exist a counterbalance valve 6 and a hydraulic lock 5 for each thrust group, the former assures a stable return movement while the latter locks the circuit to prevent leakage for safety protection when the cylinder thrust is released.

3 PID Control of Thrust Hydraulic System

The control block diagram of one group in the thrust system is shown in Fig.4, which represents the thrust motion in forward direction. The thrust distance is measured by a displacement sensor equipped inside the cylinder, which is simultaneously converted into velocity parameter to be fed back to input signal of the flow control valve. Similarly, the pressure closed-loop control is completed by pressure sensor. By comparing the reference signals and the feedback ones, the errors are produced then transferred to the controllers. The adjustments of pressure and flow rate are realized synchronously.

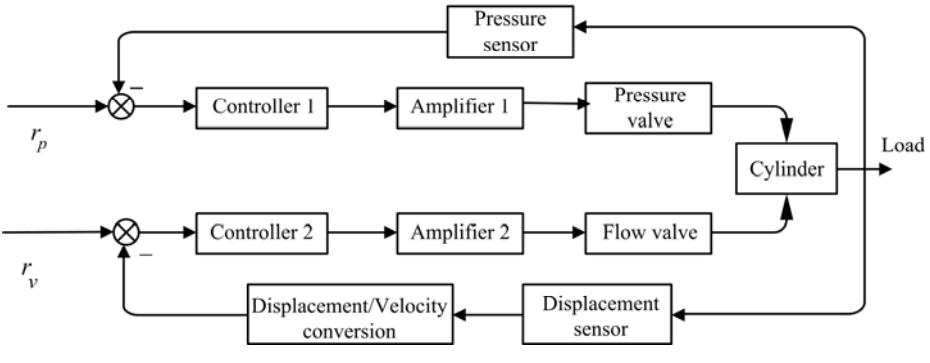


Fig. 4. Block Diagram of the Pressure and Speed Control System

3.1 Experiments

In this project, experiments are conducted with a test rig mentioned in the earlier part of this paper. In the experiments, two typical kinds of soils are employed, and their

properties are listed in Table I. The following experiment results are obtained with the thrust hydraulic control system using conventional PID controller. Fig. 5 and Fig. 6 show the thrust speed and pressure in a certain clay soil excavation section during the experiments. In the figures, the simulation results are also added to make a comparison. As shown in Fig. 5, the thrusting velocity which overshoots up to about 38mm/min in beginning, stabilizes at 30mm/min after 10 seconds or so. When adjusted to 42mm/min after 200 seconds, there exists another sharp rise lasting for a very short period, then the velocity comes back to set value level. During the course of speed adjustment, it can be seen from Fig. 6 that the thrust pressure maintains its normal value of 6MPa approximately. Moreover, the experiment results match the simulation results well.

Table 1. Parameters of Two Kinds of Soils

Parameters	Clay	Sandy gravel
Moisture content (%)	47.2	18.0
Unit weight (kN/m^3)	17.0	20.2
Angle of internal friction (°)	10.0	37.8
Cohesion (kPa)	10.0	0.1
Void ratio	1.26	0.559

In Fig.7 and Fig. 8, red and black curves are given together, because the thrust is conducted with two cylinders 3# and 4# mounted symmetrically shown in Fig. 2. This experiment is finished in the sandy gravel tunneling section. Tuning the control knob at 100s, thrust pressure follows the changes gradually. Due to the unequal load acted on two cylinders, the pressure levels are not consistent with each other. However, the driving speed curves always keep around 30mm/min with good agreement with each other. It also can be seen that the working pressure and speed in sandy gravel goes less steadily than in clay.

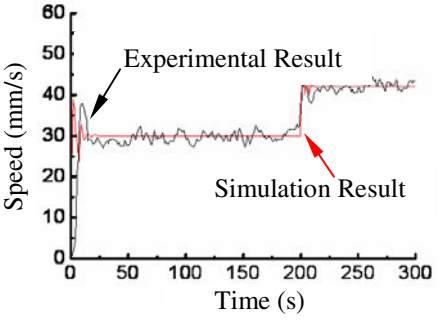


Fig. 5. Speed of Thrust Hydraulic Cylinder in Clay

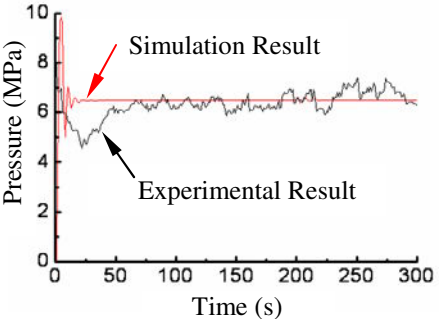


Fig. 6. Pressure of Thrust Hydraulic Cylinder in Clay

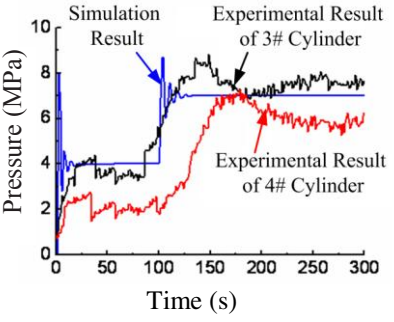


Fig. 7. Thrust Pressure in Sandy Gravel

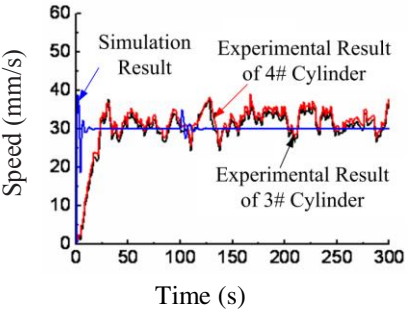


Fig. 8. Thrust Speed in Sandy Gravel

Because of the nonlinearity and uncertainty of soil and the large inertia of the tunneling machine, the experimental curves are fluctuating all the way compared with the simulation ones. In all, the thrust hydraulic control system with conventional PID controller can perform a relatively good task in tunneling. But in practice, the geological conditions are very complex and highly uncertain. To get a better performance for tunneling machine, it is preferable to employ the fuzzy control as described below.

4 Fuzzy Control of Hydraulic System

Tunneling is a very complicated process involving such unknown and unpredictable factors as nonlinear and variable load. The conventional PID control shows disadvantage in parameter adjusting for varying condition, whereas the fuzzy control can provide a better solution for the process with uncertainty and variation [5]. Taking both advantages of conventional PID and fuzzy control, this paper proposes two different control strategies as follows.

4.1 Fuzzy-PID Compound Control Design

The principle of combining PID and fuzzy control to improve the performance of the control system is shown in Fig. 9, in which the former is to eliminate the error of steady state while the latter functions for fast response of the system [6]. The PID controller and the fuzzy controller work simultaneously, and their outputs are summed up and presented to the controlled object *y*.

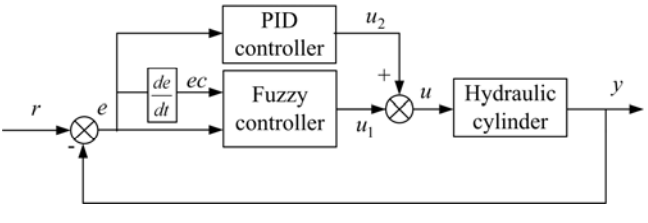


Fig. 9. Principle of Fuzzy PID Compound Control

Fig.10 is the structure of two dimensions Mamdani fuzzy controller [7] which is adopted as the fuzzy controller of fuzzy-PID compound control. The two dimensions Mamdani fuzzy controller has two inputs e and ec which represent error and error variation rate respectively, and it has one output namely u_1. Both of E and EC are divided into 7 linguistic values as {PB, PM, PS, ZO, NS, NM, NB}, which express positive big, positive medium, positive small, zero, negative small, negative medium, negative big. Trigonometric function is chosen for inputs and outputs membership function. According to the properties of the thrust hydraulic system of shield tunneling machine the fuzzy control rules are listed in Table 2.

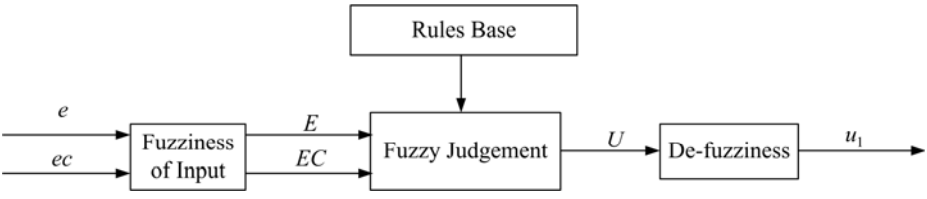


Fig. 10. Structure of Two Dimensions Mamdani Fuzzy Controller

Table 2. Fuzzy Control Rules of The Variable-U

E\EC	NB	NM	NS	ZO	PS	PM	PB
NB	NB	NB	NB	NB	NM	ZO	ZO
NM	NB	NB	NB	NB	NM	ZO	ZO
NS	NM	NM	NM	NM	ZO	PS	PS
ZO	NM	NM	NS	ZO	PS	PM	PM
PS	NS	NS	ZO	PM	PM	PM	PM
PM	ZO	ZO	PM	PB	PB	PB	PB
PB	PB	ZO	PM	PB	PB	PB	PB

4.2 Fuzzy Adaptive PID Control Design

Tuning the parameters of PID controller is a troublesome task usually, and the fixed values of the parameters do not always function satisfactorily under changing working conditions. So the PID controller with the parameters suffered real time adjustment is preferred. As shown in Fig. 11, the fuzzy inference system in the fuzzy control can tune the three PID parameters K_p, K_i, K_d online by the aid of error and its variation rate. Thereby, the controller can adapt itself to meet the requirements of control system with variable parameters.

Like the fuzzy-PID compound control, the two variables E and EC are also divided into 7 linguistic values as {PB, PM, PS, ZO, NS, NM, NB} in the fuzzy adaptive PID control. Trigonometric function is chosen for inputs and outputs membership function, and PID parameters can be tuned as:

$$\begin{cases} K_p = K_{p0} + \Delta K_p \\ K_i = K_{i0} + \Delta K_i \\ K_d = K_{d0} + \Delta K_d \end{cases} . \tag{1}$$

where, K_p, K_i and K_d refer to the PID online parameters, K_{p0}, K_{i0}, K_{d0} are the PID initial parameters.

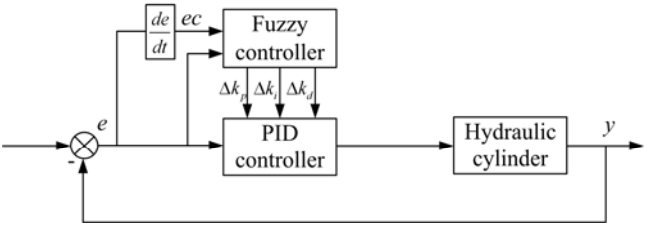


Fig. 11. Principle of Fuzzy Adaptive PID Control

On the basis of operating experiences, the fuzzy control rules of the fuzzy adaptive PID control are listed in Table 3.

Table 3. Fuzzy Control Rules of The PARAMETERS' MODIFICATION

K_p/ K_i/ K_d *EC* / *E*	NB	NM	NS	ZO	PS	PM	PB
NB	PB/NB/ PS	PB/NB/ NS	PM/NM/ NB	PM/NM/ NB	PS/NS/ NB	ZO/ZO/ NM	ZO/ZO/ PS
NM	PB/NB/ PS	PB/NB/ NS	PM/NM/ NB	PS/NS/ NM	PS/NS/ NM	ZO/ZO/ NS	NS/ZO/ ZO
NS	PM/NB/ ZO	PM/NM/ NS	PM/NS/ NM	PS/NS/ NM	ZO/ZO/ NS	NS/PS/ NS	NS/PS/ ZO
ZO	PM/NM/ ZO	PM/NM/ NS	PS/NS/ NS	ZO/ZO/ NS	NS/PS/ NS	NM/PM/ NS	NM/PM/ ZO
PS	PS/NM/ ZO	PS/NS/ ZO	ZO/ZO/ ZO	NS/PS/ ZO	NS/PS/ ZO	NM/PM/ ZO	NM/PB/ ZO
PM	PS/ZO/ PB	ZO/ZO/ NS	NS/PS/ NS	NM/PS/ NM	NM/PM/ PS	NM/PB/ PS	NB/PB/ PB
PB	ZO/ZO/ PB	ZO/ZO/ PM	NM/PS/ NM	NM/PM/ NM	NM/PM/ PS	NB/PB/ PS	NB/PB/ PB

4.3 Simulation Model

The investigation of fuzzy control applied in the thrust hydraulic system of the tunneling machine is carried out by simulations in AMESim and Matlab/Simulink environment simultaneously that is called co–simulation, and the first software is for hydraulic

modeling while the second is in the charge of control model creating. It can be seen from Fig. 12 that the simulation model is simply composed of proportional valves, cylinders, load models, measuring components, control loops. The block with two inputs from sensor signals and two output ports to the valves is an interface connecting the two softwares together. By this way, co-simulation is achieved.

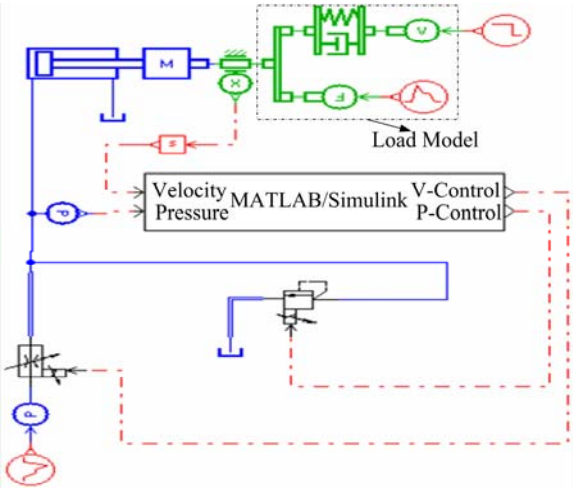


Fig. 12. Co-Simulation Model in AMEsim

In the simulation, based on the thrust speed and pressure control in actual tunneling, different speed values are set to verify the adaptability of the control strategies.

4.4 Simulation Results and Comparison

Fig. 13 and Fig. 14 give the simulation results of thrust hydraulic system with fuzzy-PID compound control. Fig. 15 and Fig. 16 present the simulation results with fuzzy adaptive PID control. Here, comparisons are drawn between the results by different control actions.

According to the plots, in the fuzzy-PID compound control system, the speed output has good performance and fast response, reaching the set value 0.8mm/s in about 3s. The working pressure rises rapidly to steady state. As for different speed shown in the figures, the lower the speed is, the faster the response is. That is the same to the influence on pressure.

In addition, the speed response is faster and the overshoot is smaller in fuzzy adaptive PID control system than that in fuzzy-PID compound control system, and so is the pressure response. Especially when the thrust speed is low, the difference is much more obvious. In a word, fuzzy adaptive PID control is superior in both speed and pressure characteristics to fuzzy-PID compound control.

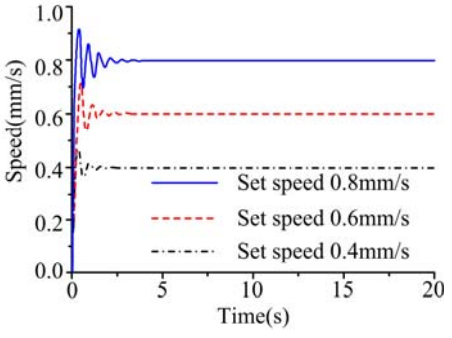


Fig. 13. Speed Response with Fuzzy-PID Compound Control

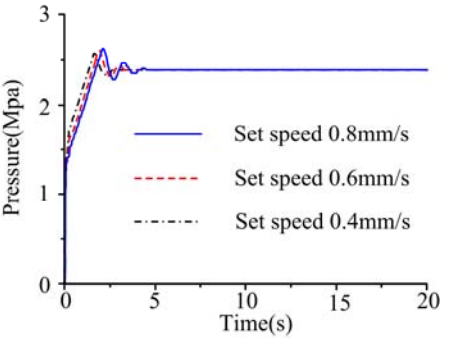


Fig. 14. Pressure Response with Fuzzy-PID Compound Control

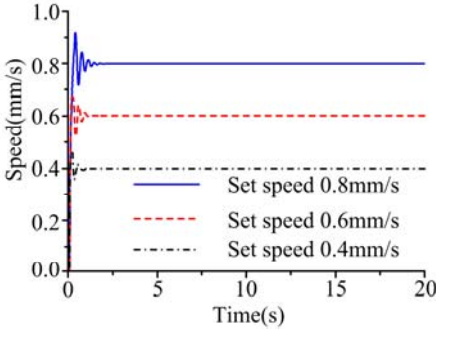


Fig. 15. Speed Response with Fuzzy Adaptive PID Control

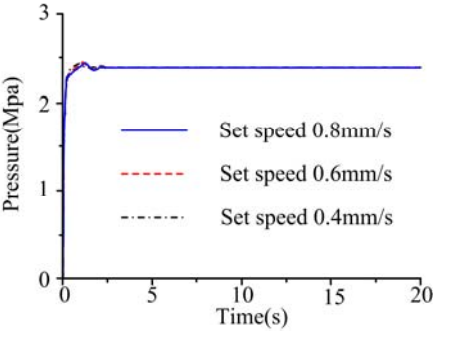


Fig. 16. Pressure Response with Fuzzy Adaptive PID Control

5 Conclusion

This paper dealt with the control of thrust hydraulic system of shield tunneling machine. By simulations as well as the experiments, some comparisons were drawn with different control strategies. According to the simulation results, the thrust hydraulic system employed the fuzzy inference can achieve better thrust performance than pure PID control. Adaptive PID control with on-line parameter tuning is favorable for tunneling process control. Experiments of fuzzy control will be focused in the next work to verify the simulation and the comparison.

Acknowledgments. The authors wish to thank National Basic Research Program (973 Program) of China under Grant No.2007CB714004 and National High Technology Research and Development Program of China (863 Program) under Grant No.2007AA041806 for their supporting.

References

1. Walter., W., Herbert, H.: Hydraulic control system for and methods of controlling the operation of tunneling apparatus. U.S. Patent 4391553 (1983)
2. Maidl, B., Herrenknecht, M., Anheuser, L.: Mechanized Shield Tunneling. Ernst & Sohn, Berlin (1996)
3. Hu., G.L., Gong., G.F., Yang, H.Y.: Electro-hydraulic control system of shield tunnel boring machine for simulator stand. In: 6th International Conference on Fluid Power Transmission and Control, Hangzhou, China, pp. 94–99 (2005)
4. Hu, G.L., Gong, G.F., Yang, H.Y.: Thrust hydraulic system of shield tunnel boring machine with pressure and flow compound control. Chinese Journal of Mechanical Engineering 42(6), 124–127 (2006)
5. Liu, J.K.: Advanced PID Control and MATLAB Simulation. Electronic Engineering Press, Beijing (2003)
6. Hu, G.L.: Research into electro-hydraulic control system for a simulator test rig of shield tunneling machine. Ph.D. dissertation, Dept. Mech. Eng., Zhejiang Univ., Hangzhou, China (2006)
7. Zhang, Y.H., Shao, C., Cui, B.: Fuzzy-P ID hybrid control of melted aluminum in aluminum powder atomization furnace. Journal of Dalian University of Technology 46(4), 572–575 (2006)

Simulation Analysis of Pressure Regulation of Hydraulic Thrust System on a Shield Tunneling Machine*

Zhibin Liu, Haibo Xie**, and Huayong Yang

State Key Laboratory of Fluid Power Transmission and Control, Zhejiang University, Hangzhou, 310027, China
hbxie@zju.edu.cn

Abstract. Hydraulic thrust system is an important system on a shield tunneling machine. Pressure regulation of thrust cylinders is the most important function of thrust system during tunnel excavation. In this article, a hydraulic thrust system is illuminated, and a corresponding simulation model is carried out in order to study the system characteristics. Pressure regulation of a certain group's cylinders has a little influence of the other groups' cylinders. The influence will not affect the process much during tunnel excavation. Pump displacement may have a great effect on the pressure regulation. Oil supply flow rate should be adaptive to the system demand. A tough situation is simulated to explain how the pressure regulation works during tunnel excavation.

Keywords: Tunnel, hydraulic thrust system, pressure regulation, simulation.

1 Introduction

Shield tunnel machine is a large and complex machine for underground tunnel excavation. It is used in construction projects such as underground rail lines, urban pipelines, submarine tunnels and water conservancy, etc. Hydraulic technique is widely applied on shield tunneling machine. For instance, it is used in thrust system, cutter head drive system, screw conveyor, segment erector, etc.

Some research works about tunnel excavation are carried out. Sugimoto and Sramoon have built a theoretical model for tunnel excavation based on mechanics analysis [1]. Their simulation results have a good agreement with the observed data [2]. Maynar and Rodriguez used discrete numerical model to analyse the excavation process. Some studies about thrust force and torque are carried out [3]. Hehua Zhu et al. have discovered some relationship between working parameters during tunnel excavation [4]. Only a few studies about hydraulic system on the machine are done. Guoliang Hu et al. have done some works on a thrust system of a machine. Pressure and flow compound control is studies in article [5].

* This research is supported by the National Basic Research Program (973 Program) of China (No. 2007CB714000) and the National High Technology Research and Development Program (863 Program) of China (No. 2008AA042803).

** Corresponding author.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 493–503, 2009.

A different system from that of article [5] is studied in this research. Pressure regulation of thrust system is studied by using a simplified simulation model. And more cases about the regulation are discussed.

2 Thrust System

Thrust system is an important part of a shield tunneling machine. The system consists of power unit (electric motor and hydraulic pump), hydraulic valves and hydraulic cylinders as actuators. The number of thrust hydraulic cylinders is usually 16 or 32. If individual control of hydraulic cylinders is applied, more proportional control valves and pressure sensors are required. Moreover, machine operator has to control 16 or 32 pressure parameters during excavation. Individual control is rather complicated and expansive in cost. Nowadays, thrust cylinders are usually divided into 4 or 5 groups, and 4-group arrangement is more common. 4-group cylinders are divided into Group A, Group B, Group C and Group D, corresponding to right zone, lower zone, left zone and upper zone. The proportion of A: B: C: D is usually 4:5:4:3, as Fig.1 shows.

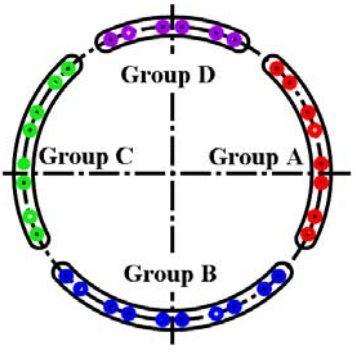


Fig. 1. Arrangement of 4-group hydraulic cylinders

More cylinders are required in Group B so as to counter the weight of cutter head and the higher earth pressure in the lower zone. Oil pressures in different cylinders are the same if they are in the same group. In this case, only 4 proportional control valves are required, and machine operator only has to control 4 pressure parameters.

During tunnel excavation, hydraulic cylinders are supplied with high pressure hydraulic oil, exerting on the segments, in order to generate a thrust force to push the machine forward. The machine operator may adjust hydraulic pressures of thrust cylinders to achieve steering control and machine posture adjustment. After a thrust process is over, segments will be set up to compose a new ring of tunnel. Meanwhile, thrust cylinders are pulled back in order to make room for those segments. When a segment is located, corresponding cylinders are pushed out. These cylinders exert on the segment, supporting the segment with a small force to finish the install.

3 Hydraulic Circuit and Modeling

The hydraulic circuit diagram for the thrust system is shown as Fig.2. The system mainly consists of a variable displacement pump 1, a pressure relief valve 2, a directional control valve 4, four pressure reducing valves 5 and sixteen hydraulic cylinders 8. Pump displacement is proportional to the input signal. Pressure relief valve works when the system is overloaded. Directional control valve is used to control the push out or pull back motion of cylinders. Pressure reducing valve is the most important component and is used to adjust the hydraulic pressure of cylinders.

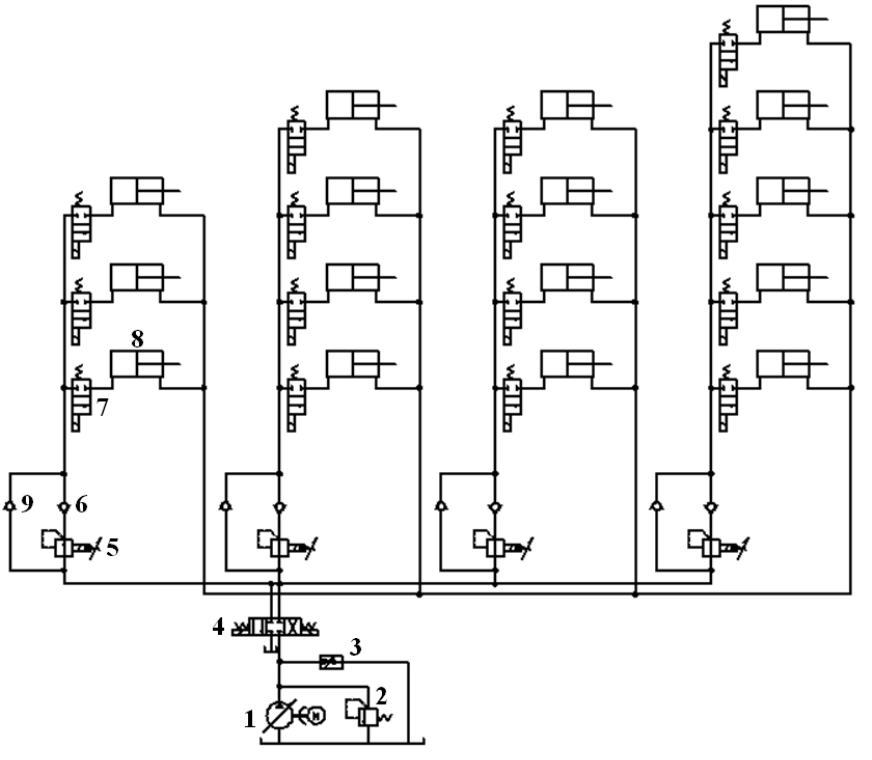


Fig. 2. Hydraulic circuit of thrust system

When cylinder is pushed out, high pressure oil comes from hydraulic pump 1, and then flows through directional control valve 4, pressure reducing valve 5, check valve 6, on/off valve 7, finally into the rear chamber of hydraulic cylinder 8. The back pressure oil flows out of the rod chamber of cylinder 8, then through directional control valve 4 and back into the oil tank at last. When cylinder is pulled back, oil flows through directional control valve 4, then into the rod chamber of hydraulic cylinder 8. The back pressure oil flow out of the rear chamber, then through on/off valve 7, check valve 9, directional control valve 4, finally into the oil tank.

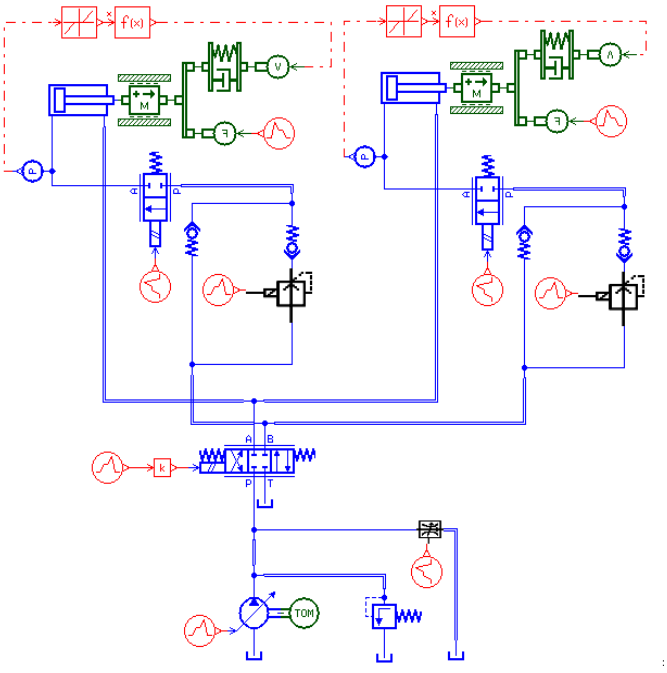


Fig. 3. Simulation model of the thrust system

We use commercial software AMESim as analysis tool. To simplify the model, we consider two zones of hydraulic cylinders for analysis. The simulation model is shown as Fig.3. All the hydraulic models are the normal models provided by the software except the pressure reducing valves and the flow control valve. These valves are built as "super-component", which is composed by other normal models. These valves give more details about not only the working parameters but also the internal structures.

We consider three kinds of forces in the cylinder load model: earth pressure force, friction and the force caused by the thrust motion. Earth pressure exerts on the cutter head and linings of the machine. The force exerted on the cutter head affects the thrust motion and the force on the lining can be ignored. Friction on linings of the machine caused by the earth pressure is considerable. It usually costs about half, or even more, of the thrust force. During excavation, however, these two forces almost maintain constantly. We set these two forces as constant in the model. Earth pressure is set by the input signal of the load model, and friction is set in the mass model of cylinder. Article [4] refers that there is an exponent relationship between advance speed and thrust force exerted on the excavated face. The higher of thrust force, the higher of thrust speed is. There is only slight different between exponent relationship and linear relationship. Fig.4. is cited from article [4]. The black dots represent the experimental result for a 54% open ratio cutter head, and the white dots represent the result for 36% open ratio. The black real line represents the exponent relationship, which is done in the article [4]. The red dashed line represents the linear relationship, which is supplemented in this paper. The dashed line also has a good agreement with the experimental result.

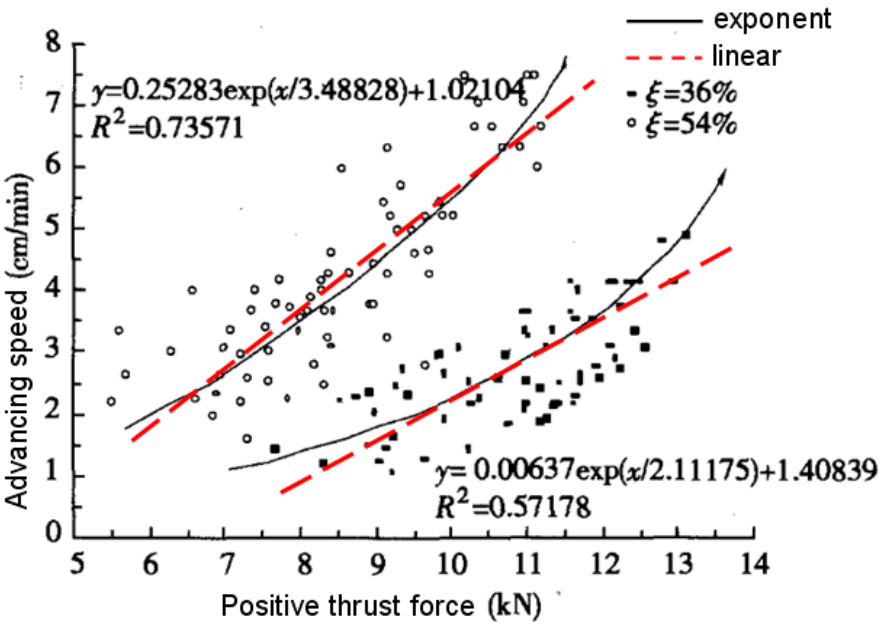


Fig. 4. Relationship between positive thrust force and advancing speed

We prefer linear to exponent so as to simplify the simulation. The thrust force should be greater than earth pressure exerted on the cutter head together with friction force on the lining, in order to push the machine advance. If thrust force is not great enough, the machine will not advance. In Fig.3, there is a dead band in the load model. The dead band represents the situation that thrust force is not great enough to push the machine forward. While the machine is advancing, increase of the thrust force will cause increase of advance speed. We apply this experiment result to the load model.

We mainly focus on the thrust pressure regulation during tunnel excavation. Pressure regulation determines the excavation working condition. Advancing orientation is due to the pressure distribution of thrust cylinders. Machine operator may control the posture of the machine by adjusting the pressures of cylinder groups. The machine may steer left or right, or advance along a small angle of slope if corresponding settings of pressure regulation are given. Normally, at the end of a tunnel excavation, the difference between actual tunnel axis and designed tunnel axis should be no more than 20 mm. The machine operator will adjust the machine advancing axis to agree with the designed axis. As a result, pressure regulation is a very important function during tunnel excavation.

4 Simulation

The main simulation parameters are set as follows:

Maximum displacement of hydraulic pump:	30 cc/rev
Nominal pump speed:	1500 rev

Relief pressure:	35.3 MPa
Piston diameter of hydraulic cylinder:	600 mm
Rod diameter of hydraulic cylinder:	480 mm
Force exerted on cutter head:	3200 KN (800KN on a cylinder)
Friction force:	6000 KN

The simulation parameters are obtained from a shield tunneling machine which is used in Nanjing subway construction. We take left and right thrust groups for simulation, and sixteen hydraulic cylinders are reduced to eight. A 63 cc/rev maximum displacement pump should be reduced to 31.5 cc/rev, and we take 30 cc/rev. Piston diameter of a real cylinder is 300mm, and rod diameter is 240mm. Either thrust group that contains 4 hydraulic cylinders is simplified to one hydraulic cylinder, which has the same work areas as the original cylinder group. As a result, the piston diameter and rod diameter of simulation cylinder are 600mm and 480mm, respectively. When oil pressure is about 8 MPa, which supplies about 9200 KN of thrust force, the machine begins to move. Article [6] refers that friction force costs about 53.5% ~ 73% of thrust force. And friction force is selected to be 6000 KN, which costs about 65% of thrust force. The rest 3200 KN is exerted on cutter head for tunnel excavation.

4.1 Pressure Regulation

Steering control is simulated in this case. Both hydraulic cylinders are set to about 14 MPa at first. The machine advances straight. We hope the machine to steer left. The pressure of right cylinder (P_{right}) should be higher than that of the left (P_{left}). P_{left} may maintain at 14 MPa. P_{right} is set to 20 MPa. Simulation result is shown as Fig. 5 and Fig. 6.

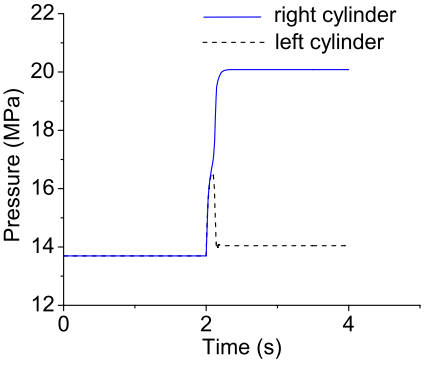


Fig. 5. Pressures of right and left cylinders during pressure regulation of left steering

P_{right} is adjusted to 20 MPa as the step signal is given. There is a small overshoot of the pressure in Fig. 5, for there is a mass-spring system in the pilot stage. Overshoot of the mass-spring system causes overshoot of the outlet pressure in the second stage. P_{left} has a pulse when P_{right} is adjusted and increases slightly after adjustment. The inlet pressure sudden change will cause the same pulse of outlet pressure at first.

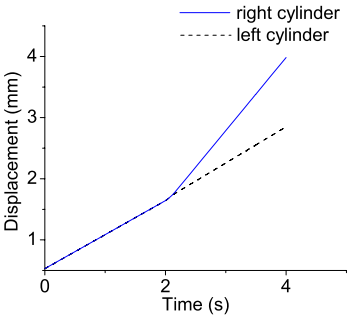


Fig. 6. Displacements of right and left cylinders during pressure regulation of left steering

Because the pilot stage frequency response is not high enough. When the pilot stage mechanism starts adjusting the opening of valve, outlet pressure will decrease to the set value. The small increase of P_{left} might be caused by the increase of pump outlet flow rate. The pump outlet flow rate will be discussed next. This small different in pressure will not affect the tunnel excavation. Higher advancing rate is achieved on the right side, as shown in Fig.6. A left steering motion is carried out.

4.2 Pump Displacement Adjustment

Pump displacement should be adjusted when pressure regulation is carried out. Three cases are illuminated to explain the problem. Pressure regulation is operated as the previous case.

Case 1: Displacement maintains at about 75% of maximum. Fig. 7 shows the result.

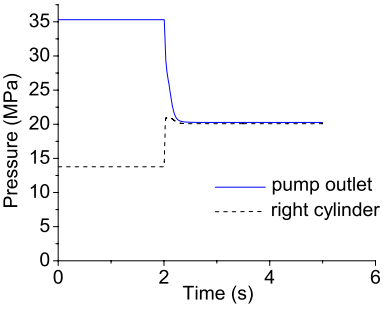


Fig. 7. Pressures of pump outlet and right cylinder when pump displacement maintains at 75% of maximum displacement

P_{right} is adjusted to 20 MPa. Pressure regulation is achieved. However, pump outlet pressure is at 35.3 MPa, which means pressure relief valve is open and the system is overloaded. Power loss caused by relief flow is considerable.

Case 2: Displacement maintains at about 50% of maximum. Fig. 8 shows the result. Pump outlet pressure is sensitive to the load pressure. There is no overload during pressure regulation. However, thrust pressure can not be adjusted to 20 MPa, for oil flow supply is not enough. Higher thrust pressure causes higher advance speed, which means the cylinder requires more oil flow rate. Pressure regulation is invalidated in this case.

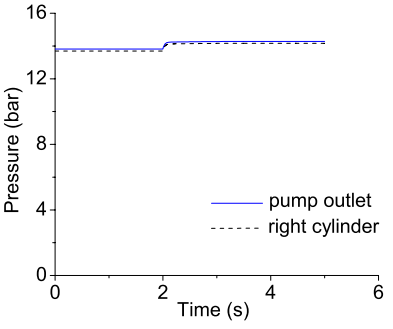


Fig. 8. Pressures of pump outlet and right cylinder when pump displacement maintains at 50% of maximum displacement

Case 3: Displacement is set at 50% at first, then 75% when pressure regulation is carried out. Fig. 9 shows the result. The system is not overloaded. Pump outlet pressure is sensitive to the load pressure. Thrust pressure can be adjusted to 20 MPa. This is because pump displacement is adaptive to the system. A pressure regulation should be together with a pump displacement regulation.

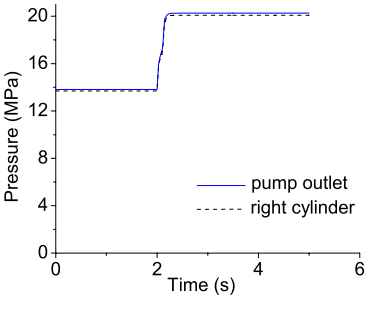


Fig. 9. Pressures of pump outlet and right cylinder when pump displacement is adjusted from 50% to 75% of maximum displacement when pressure regulation is applied

4.3 A Tough Situation

An optimal tunnel project is worked out by considers of citizen's demand, commercial demand, geological conditions, etc. Shield tunneling machine is expected to advance in good geological condition. Normally, geological survey is done before excavation

process. Though, it is impossible to know about everything that in front of the machine. Unexpected conditions may happen to the machine, such as encounters of hard rock, wooden poles, quicksand or underground river.

A tough situation is simulated in this case. The machine encounters some hard soil in the right side. The followings show how the pressure regulation works and how the machine behaves. The simulation result is shown as Fig. 10 and Fig. 11.

Note: Time domain is used as X-axis in the simulation. However, the time is not a practical data. Time setting is artificial for the purpose of the illustration. Time period may be much longer in practical tunnel process.

0s – 5s: The machine is advancing straight in soft soil with a normal speed. Thrust pressure is 14MPa.

5s - 15s: The machine is advancing from soft soil into hard soil on the right side. Right cylinder speed (V_{right}) decreases, because P_{right} is not enough to maintain the normal speed. This causes surplus pump outlet flow rate and increase of system pressure. And P_{left} and left cylinder speed (V_{left}) increase. System pressure increases until relief valve is open. Meanwhile, P_{left} and V_{left} drop to the normal.

At 20s: P_{right} is adjusted to 31MPa. A pulse occurs in P_{left} and causes a pulse in V_{left}, as illuminated before. Because an increase of flow rate occurs in right cylinder, flow rate of relief valve will decrease suddenly and there is a sudden drop of system pressure.

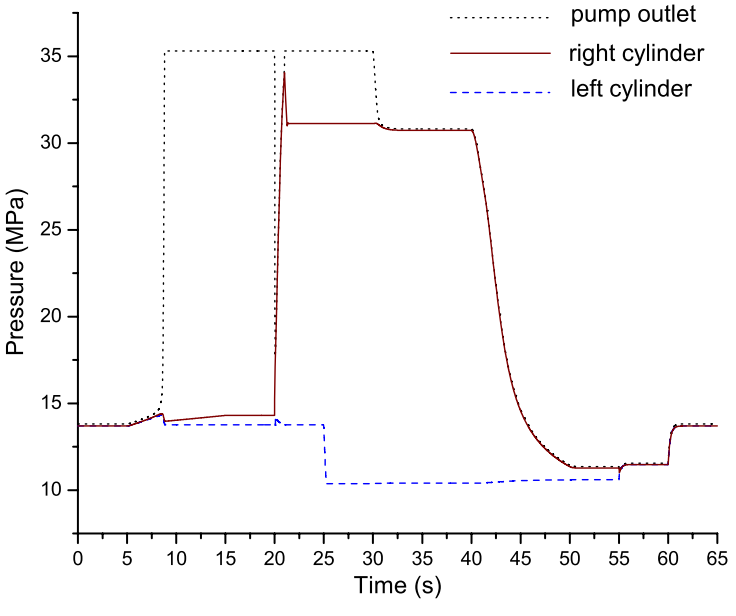


Fig. 10. Pressures of pump outlet, right cylinder and left cylinder as the machine advances into and then out of some hard soil

At 25s: V_{left} should decrease to ensure straight advance. For V_{left} is higher than V_{right}, unwanted right steering motion is carried out. Manual control to decrease P_{left} until the machine advances straight.

At 30s: Pump displacement is adjusted. After pressure regulation, the machine is now advancing at a lower speed. Pump displacement should decrease until overload is avoided, or over-adjustment will cause decrease of thrust pressure and advancing speed.

40s - 50s: The machine is advancing out of the hard soil. P_{right} decreases as the machine goes into soft soil little by little. P_{right} is rather high at the beginning, which causes speed increases. Oil supply is not enough for the system anymore as V_{right} increases. This causes the drop of system pressure, and V_{right} decreases at the moment.

At 55s: Thrust pressure is adjusted. Thrust pressure may be set to 14 MPa. Pressure regulation should take place before pump displacement adjustment. A slight increase of thrust speed happens due to the increase of thrust pressure. Oil supply is not enough, however, and pressure regulation is invalidated.

At 60s: Pump displacement is adjusted. As pump displacement increases, thrust pressure is adjusted to 14MPa and thrust speed increases. The machine is advancing at the normal speed from now on.

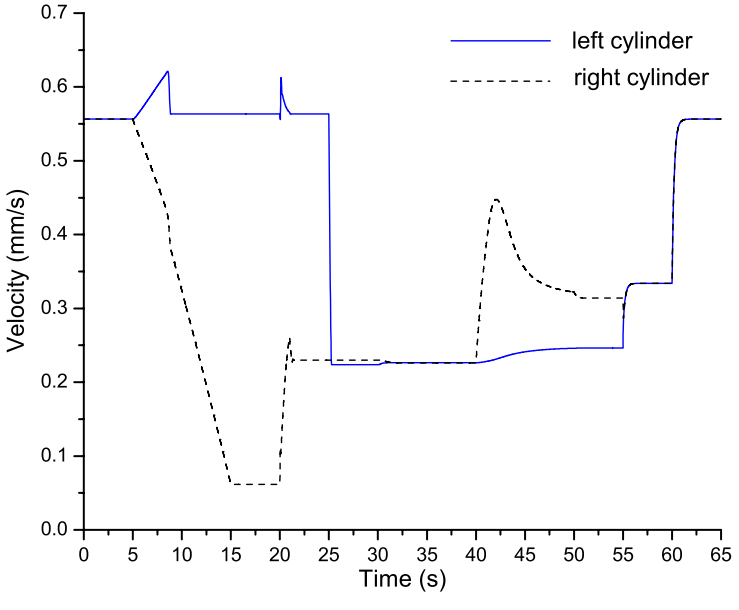


Fig. 11. Velocities of right cylinder and left cylinder as the machine advances into and then out of some hard soil

5 Conclusions

Pressure regulation of the thrust system on a shield tunneling machine is studied in this article. Pressure regulation includes two parts: pressure setting of thrust cylinders and pump displacement adjustment.

1) Pressure regulation of a cylinder group may influence the pressures of the other groups. The increases of a certain group's pressure and pump outlet flow rate may cause small pressure increase in other groups. This small change in pressure is acceptable in tunnel excavation process.
2) Pump displacement adjustment should be carried out when pressure regulation is applied. If pump outlet flow rate is more than the system demand, pressure regulation will also achieved with an unwanted pump pressure overload. If pump outlet flow rate is less than the system demand, pressure regulation is invalidated though pressure setting signal is given. The pressure regulation will be achieved as pump displacement increases enough.
3) A decrease of advancing speed may be a sign of encounter of hard soil. Compared with that, a decrease of thrust pressure may be a sign of encounter of soft soil. During excavation process, thrust pressure is set first, and then pump displacement adjustment should be carried out to finish the pressure regulation.

Acknowledgments

The authors would like to thank for the great help from the engineers in Xizi United Holding Corporation and HONGRUN Construction Group.

References

1. Sugimoto, M., Sramoon, A.: Theoretical Model of Shield Behavior During Excavation. I: Theory. J. Geotech. Geoenviron, 138–155 (February 2002)
2. Sramoon, A., Sugimoto, M., Kayukawa, K.: Theoretical Model of Shield Behavior During Excavation. II: Application, 155–165 (Feburary 2002)
3. Maynar, M.J.M., Rodríguez, L.E.M.: Discrete Numerical Model for Analysis of Earth Pressure Balance Tunnel Excavation. J. Geotech. Geoenviron., 1234–1242 (October 2005)
4. Zhu, H., Xu, Q., Zheng, Q., Liao, S.: Experimental Study on the Working Parameters of EPB Shield Tunneling in Soft Ground. China Civel Engineering Journal 40(9), 87–94 (2007) (in Chinese)
5. Hu, G., Gong, G., Yang, H.: Thrust Hydraulic System of Shield Tunnel Boring Machine with Pressure and Flow Compound Control. Chinese Journal of Mechanical Engineering 42(6), 124–127 (2006) (in Chinese)
6. Ling, J., Fan, L.: Shield Construction Technique and its Application in Guangzhou Underground Rail Lines. Guangzhong Technique 3, 25–32 (2000) (in Chinese)

Variable Rheological Joints Using an Artificial Muscle Soft Actuator and Magneto-Rheological Fluids Brake

Yuichiro Midorikawa and Taro Nakamura

Chuo University, Department of Precision Mechanics,
Faculty of Science and Engineering,
1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan
y_midorikawa@bio.mech.chuo-u.ac.jp,
nakamura@mech.chuo-u.ac.jp

Abstract. In recent times, the chances of robot-human contact have increased; hence, safety is necessitated with regard to such contact. Thus, manipulators using a pneumatic rubber artificial muscle, which is lightweight and flexible, are studied. However, this artificial muscle manipulator has faults such as slow response and limited instantaneous power due to operation by air pressure. Because of these faults, uncontrollable vibrations can occur, leading to instability in the arm when an object is held and lifted. In this study, an artificial muscle manipulator with 1 DOF and a variable rheological joint mechanism using MR fluid is developed. Vibration control of the arm using MR fluid is realized when an object is held and lifted, confirming the reduction in vibration due to the MR effect.

Keywords: MR brake, Artificial muscle, Soft manipulator, Position control, Vibration reduction.

1 Introduction

Partner robots for life support and power assist machinery for medical care have been developed recently. These robots are expected to be active in the human living environment in the future. Because humans and robots will share an environment, safety is necessitated in robot–human contact.

Therefore, we developed a manipulator using a straight-fiber-type pneumatic artificial muscle actuator. This type of artificial muscle has greater contraction ratio and power, and a longer lifetime than conventional McKibben-type muscles [2]–[6]. Furthermore, the muscles are very lightweight and flexible; hence, a manipulator made of this type artificial muscle has greater drivable range and torque. This ensures the safety of the manipulator in robot–human contact, which makes it suitable for performing collaborative activities with humans.

On the other hand, the muscle has high compliance. This causes the actuator to vibrate under a high load. Also, the former artificial muscle manipulator has some faults such as slow response and limited instantaneous power for operation by air pressure. Therefore, the control method for artificial muscles using feed-forward control that applies the mechanical equilibrium model and feedback control using proportional-integral (PI) control, and vibration control at the load was performed in the past.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 504–514, 2009.

Vibration control in the low-frequency area was confirmed, but vibration control in the high-frequency area was unrealizable due to the slow response of air pressure. In addition, variable joint stiffness control of the artificial muscle manipulator was performed [7]. However, it is difficult to control momentary joint stiffness, because the response of air pressure is slow.

On the other hand, magneto rheological (MR) fluids have been widely adopted as smart materials with properties that can be controlled by tuning an external magnetic field, which can transform from a fluid-like to a solid-like state within milliseconds. This fluid is applied to the joint of an artificial muscle manipulator to achieve a variable rheological joint. As a result, angle control and joint stiffness control in the high-frequency area, which cannot be controlled by air pressure, can be achieved.

In this study, an artificial muscle manipulator with 1-DOF and a variable rheological joint mechanism using MR fluid is developed, in which the vibration control of the arm using MR fluid is realized when an object is held and lifted.

2 Straight-Fiber-Type Artificial Muscle

2.1 Outline of the Artificial Muscle

Fig. 1 shows a schematic diagram of the straight-fiber-type artificial muscle. A tube is made from natural latex rubber with a carbon fiber seat fixed at either end by a terminal, which is inserted in the long-axis direction.

Since this suppresses axial expansion, the artificial muscle expands radially due to air pressure and exerts a contractile force axially. In addition, the ring installed in the muscle prevents an explosion due to excessive expansion, and we can modify its influence by adjusting the ratio between the length and radius of the muscle under different conditions.

The artificial rubber muscle using this structure is lightweight, flexible, and capable of high output. However, to move by air pressure, the pressure that can be applied is limited, and this actuator cannot achieve high stiffness.

2.2 Pressure Characteristics of the Artificial Muscle

Fig. 2 shows the pressure characteristics of this artificial muscle. As shown in this figure, we can control muscle contraction, according to the pressure characteristics and use the muscle as an actuator. However, the muscle's pressure characteristics are

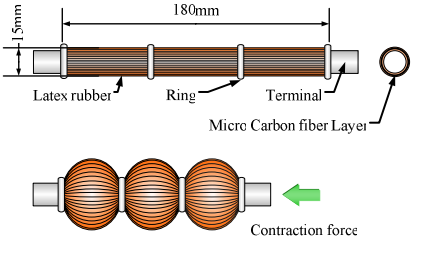


Fig. 1. Schematic diagram of artificial muscle

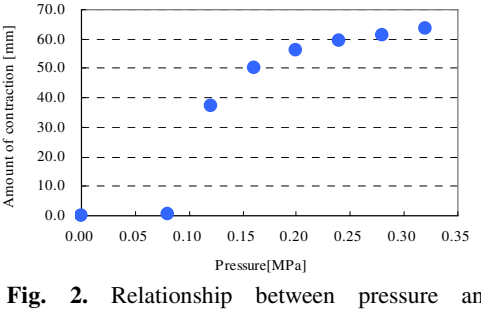


Fig. 2. Relationship between pressure and contraction

highly nonlinear due to the material used. These characteristics affect manipulator position control.

2.3 Mechanical Equilibrium Model

Since an artificial muscle has high nonlinearity, the input and output gains vary. Hence, it's highly likely that movement becomes unstable in certain situations. Therefore, a mechanical equilibrium model for the artificial muscle has been proposed in the past.

Table 1 shows the parameters of the muscle and the manipulator. In this table, the subscript number discriminates between artificial muscles 1 and 2; the common equation for both muscles uses the subscript *i*. From the mechanical equilibrium model [7], an expression of the relationships between contraction *x*, load *F*, and pressure *P* is obtained.

F is expressed as

$$F(\phi_{0i}, P_i) = \frac{PG_{3i}(\phi_{0i}) - G_{1i}(\phi_{0i})}{G_{2i}(\phi_{0i})} \quad (1)$$

The model was derived from mechanical equilibrium relations, when the artificial muscle is contracted, a cross section of the axial direction is assumed to have the circular arc shape seen in Fig. 3. ϕ_0 is the central angle of the shape, M is the coefficient for the fibers of the artificial muscle, and 0 < M < 2 is the range of M.

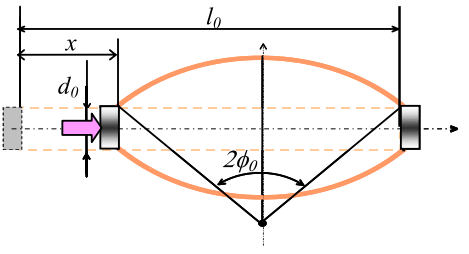


Fig. 3. Shape model of artificial muscle viewed along the Z-axis

Table 1. Parameters of the Artificial Muscle

θ_d	[rad]	Desirable angle
τ	[Nm]	Load torque
P_1	[Pa]	Pressure
P_2		
F_1	[N]	Pull force
F_2		
x_{d1}	[m]	Desirable contraction
x_{d2}		
ψ_1	[rad]	Angle of slack wire
ψ_2		
b_1	[mm]	Width of glass fiber
b_2		
α_1		Approximation constant number
α_2		

n		Number of the glass fiber
K_j	[Nm/rad]	Joint Stiffness
K_1		Fixed stiffness number
K_2		
l_{01}	[m]	Length between cap and ring
l_{02}		
d_{01}	[m]	Diameter of artificial muscle
d_{02}		
t_{01}	[m]	Thickness of artificial muscle
t_{02}		
M		Fiber constant number

$$\phi_0 = \frac{2\alpha l_0^{1.5} x}{(l_0 - x)^2 + \alpha^2 x l_0} \quad (\alpha = 1.4) \quad (2)$$

$$G_1(\phi_0) = \frac{4Kt}{d_0}\left[\frac{l_0}{d_0}\right]^2\left[\frac{\sin\phi_0 - \phi_0\cos\phi_0}{\phi_0^2}\right] \quad (3)$$

$$G_2(\phi_0) = \frac{M\tan\phi_0}{d_0 nb}, \quad (4)$$

$$G_3(\phi_0) = 2\left[\frac{l_0}{d_0}\right]^2\left[\frac{\phi_0 - \sin\phi_0\cos\phi_0}{\phi_0^2}\right] + 4\frac{l_0}{d_0}\frac{\sin\phi_0}{\phi_0} - \frac{\pi d_0 M}{4nb}\tan\phi_0 \quad (5)$$

Fig. 4 compares the theoretical and experimental results for the relationship between the contraction and the force as a parameter of pressure. It is observed that there is

good agreement between the theoretical and experimental results, and that the mechanical equilibrium model has sufficient accuracy for use in position control.

Next, the outline of an artificial muscle manipulator is shown in Fig. 5. Here, F is the contraction force, F_0 is the balance force, x is amount of contraction, θ is the joint angle, r is the pulley radius, and τ'_1 is the driving torque.

When the balance force is considered, the driving force F' is expressed as

$$F' = F - F_0 \tag{6}$$

Further, from Fig. 5, the contraction x and driving torque τ'_1 are expressed as

$$x = r\theta \quad , \tag{7}$$

$$\tau = F'r \cdot \tag{8}$$

From (7) and (8), the pressure P_1 and P_2 of the artificial muscle depends on the joint stiffness K_{jd} and the desirable angle θ_d. P_1 and P_2 are expressed as

$$\begin{aligned} P_1(\theta_d, \tau) = & \left[G_{11}(\phi_{01})G_{22}(\phi_{02}) - G_{12}(\phi_{02})G_{12}(\phi_{01}) \right. \\ & \left. + \frac{K_j}{K_{a2}} G_{21}(\phi_{01})G_{32}(\phi_{02}) + \frac{\tau}{r} G_{21}(\phi_{01})G_{22}(\phi_{02})\right] \\ & \Big/ \left[G_{22}(\phi_{02})G_{31}(\phi_{01}) + \frac{K_{a1}}{K_{a2}} G_{21}(\phi_{01})G_{32}(\phi_{02}) \right] \end{aligned} \tag{9}$$

$$P_2(\theta_d, \tau) = \frac{K_j}{K_{a2}} - \frac{K_{a1}}{K_{a2}} P_1 \text{,} \tag{10}$$

where

$$\phi_{0i}(x_{di}') = \frac{2\alpha_i l_{0i}^{1.5} x_{di}'^{0.5}}{(l_{0i} - x_{di}')^2 + \alpha^2 x_{di}' l_{0i}} \text{,} \tag{11}$$

$$G_{1i}(\phi_{0i}) = \frac{4K_i t_i}{d_{0i}} \left[\frac{l_{0i}}{d_{0i}}\right]^2 \left[\frac{\sin\phi_{0i} - \phi_{0i}\cos\phi_{0i}}{\phi_{0i}^2}\right] \text{,} \tag{12}$$

$$G_{2i}(\phi_{0i}) = \frac{M\tan\phi_{0i}}{d_{0i} nbi} \text{,} \tag{13}$$

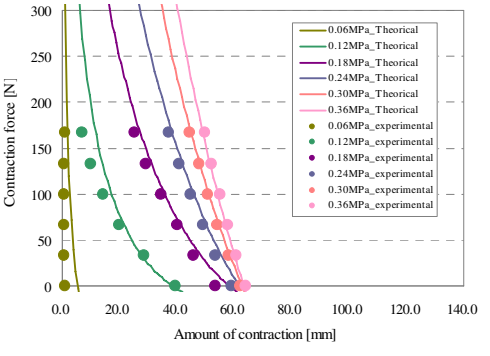


Fig. 4. Comparison with theoretical curve and experimental results

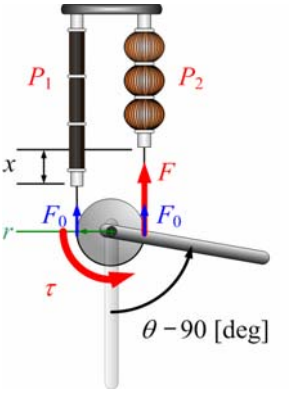


Fig. 5. Schematic diagram of manipulator joint

$$G_{3i}(\phi_{0i}) = 2\left[\frac{l_{0i}}{d_{0i}}\right]\left[\frac{\phi_0 - \sin\phi_{0i}\cos\phi_{0i}}{\phi_{0i}^{\ 2}}\right] + 4\frac{l_{0i}}{d_{0i}}\frac{\sin\phi_{0i}}{\phi_{0i}} - \frac{M\pi d_{0i}}{nb_i}\tan\phi_{0i} \tag{14}$$

From (9) and (10), stabilization of joint angle control can be achieved.

3 MR Brake

In this study, MR fluids are applied to the joint of an artificial muscle manipulator to realize a variable rheological joint. Hence, an MR brake that has small size, high output, and high response was studied.

3.1 MR Fluids

MR fluids have been widely adopted as smart materials with properties that can be controlled by tuning an external magnetic field, which can transform them from a fluid-like to a solid-like state within milliseconds. Hence, MR fluid is useful in many applications such as a clutch and brake, an actuator and vibrating-type damper. Its use in inclined and angular sensors has also been proposed.

3.2 MR Brake

In this study, MRB-2107-3 of the LORD Co. is used as an MR brake device. Fig. 6 shows the schematic view of MRB-2107-3, and Table 2 shows its specifications. This device distributes MR fluid to the circumference of an internal disk part and changes the viscous friction of the disk part surface by changing the magnetic field. As a result, the torque can be continuously controlled for revolving motion. The response speed is several millimeters per second. Therefore, the artificial muscle's response delay is avoided by MR brake. Moreover, because the MR brake has the damper function, the achievement of the system that has flexibility becomes possible. The features of the MR brake are shown below.

a) It is small and produces a high output due to its high torque density.
b) It is strong with a simple structure.
c) The response speed is several millimeters per second.
d) It is continually variable.
e) It has function of damper.
f) It is stable.

From these characteristics, wide-ranging rigidity control and vibration control are possible.

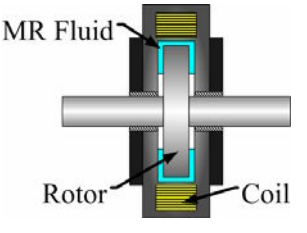


Fig. 6. Configuration of the MR brake

3.3 Characteristics of the MR Brake

In this study, the vibration of the arm is controlled with the MR brake when an object is held and lifted. Therefore, the voltage-torque characteristics of the MR brake were

investigated. Fig. 7 shows the relationship between the voltage and the torque of the MR brake. This graph shows that the torque of the MR brake rises as the impressed voltage increases.

Table 2. Specification of the MR Brake

Diameter [mm]	92
Length [mm]	36
Weight [kg]	1.4
Maximum torque [Nm]	5.65
Minimum torque [Nm]	<0.34
Max current value [A]	1
Temperature range [℃]	−28.9 to 71.1

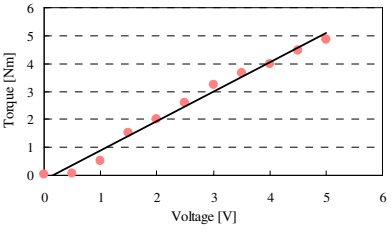


Fig. 7. Relationship between the voltage and the torque of the MR brake

4 Artificial Muscle Manipulator

This study examines vibration control of the arm by the MR fluid when an object is held and lifted. Therefore, a 1-DOF artificial muscle manipulator equipped with the MR brake in the joint is designed.

Fig. 8 shows a schematic diagram of the manipulator produced in this study, and Table 3 shows its specifications. The artificial muscle manipulator is a mechanism that pulls with two artificial muscles working together, and transmits the contraction force of the muscle to the rotation axis through the pulley. Also, the MR brake is fixed to the first link side, making it possible to apply the brake to the rotation axis. This manipulator includes an encoder to detect joint angle and a strain gauge to detect the load of the link point.

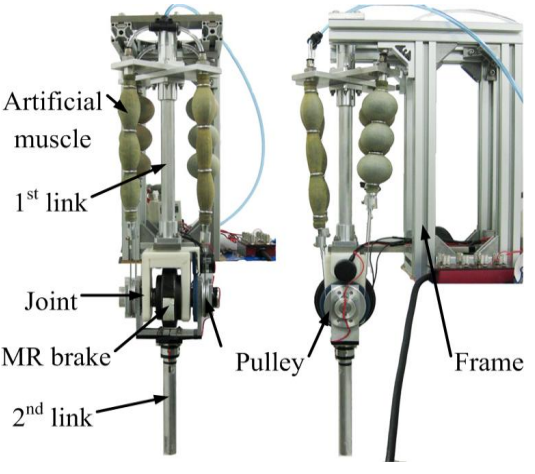


Fig. 8. Schematic diagram of the manipulator

Table 3. Specification of the manipulator

Mass of 2nd link [kg]	0.168
Length of 2nd link [mm]	250
Center of gravity 2nd link [mm]	240
Movable range [rad]	$-\pi/2 \quad \theta \quad \pi/6$
Inertia of 2nd link [kgm^2]	4.80×10^{-6}
Parallel number of artificial muscle	2
Series number of artificial muscle	1

5 Design of Control System

5.1 Control of Artificial Muscle

Fig. 9 shows a block diagram of an artificial muscle manipulator. Currently, the control technique of the artificial muscle uses feed-forward control [8] that applies the mechanical equilibrium model and feedback [8] control using PI control.

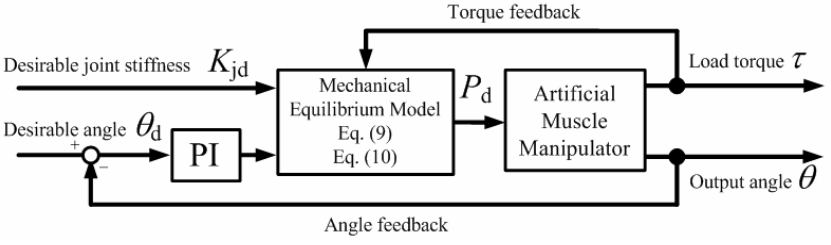


Fig. 9. Block diagram of feed-forward control

1) Feed-Forward Control

An artificial muscle manipulator is controlled by air pressure. However, the motion characteristic of the muscle is nonlinear. Also, the manipulator generates rapid motion to the target angle. Therefore, the position control may become unstable. Hence, the manipulator uses feed-forward control that applies the mechanical equilibrium model, and input and output are linearized.

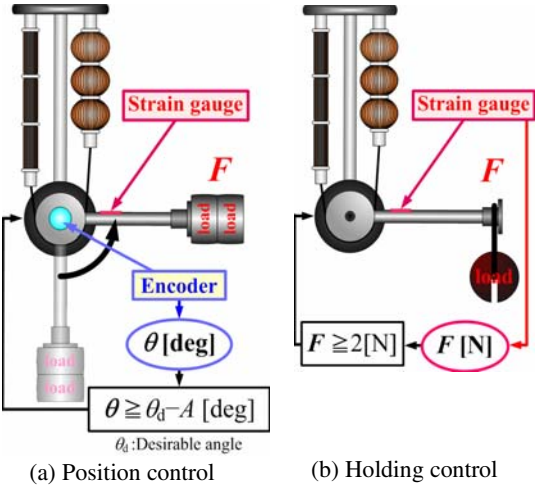
2) Feedback Control

With the dynamic equilibrium model, linearization of the input and output angles in the steady state is realized. However, an error occurs in output angle for the accuracy of the equilibrium model and the load that hangs from the joint.

Therefore, feedback control using PI control is applied to the manipulator. As a result, exact position control becomes possible, because angle compensation to a target angle is performed.

5.2 Control of the MR Brake

This study examines vibration control of the arm by the MR fluid when an object is held and lifted. Therefore, two control methods for the MR brake are produced. Fig. 6 shows the control system of the MR brake when an object is held and lifted.



(a) Position control (b) Holding control

Fig. 10. Control method for experimental systems

1) Position control
In position control (Fig.10 (a)) the MR brake is applied to the rotation axis, when the joint angle detected by the encoder reaches before *A*[deg] of a desirable angle when lifting an object. Moreover, the brake torque applies the value calculated from the load detected with the strain gauge and the length of the 2nd link.

2) Holding control
In holding control (Fig.10 (b)), the MR brake is applied to the rotation axis, when the link point load is detected by the strain gauge when holding an object. The break torque calculation technique is the same as the position control.

6 Vibration Control Experiments

6.1 Position Control Experiment

In the experiment, a load of 1.7 kg was installed at the link point of a manipulator, and the experiment examined the lift by step input of which the desirable angle is 90[deg]. This experiment examined stability in the following control systems.

Control system a) Angle feedback using PI control.
Control system b) Angle feedback using PI control and torque feedback to feed-forward control (Eq. (9), (10)).
Control system c) Angle feedback using PI control, torque feedback with MR brake. (*A* was set to 40 [deg].)

Here, in all of control system, each PI gain was adjusted so that the rise time is set to 0.4 [s], and the stability of the step responses were compared.

6.2 Holding Control Experiment

In the experiment, the arm was maintained in the position of 90[deg], and vibration control was realized when a load of 2.55 kg was put on the link point. In the case of this experiment, the above-mentioned control methods were applied.

7 Results of Experiments

7.1 Results of Position Control Experiment

Fig. 11 shows the experimental results of the step response for position control. When control system is (a), the arm vibrates due to the spring characteristics of the artificial muscle and high disturbance torque. In addition, when control system is (b), torque feedback control realized stable position control. However, a momentary vibration cannot be controlled by the slow response of air pressure. On the other hand, when control system is (c), we could confirm vibration reduction of a large vibration at the time of arm stop. From this figure, MR brake torque is impressed according to the disturbance load from the position of 50 degrees.

However, the arm causes a rapid acceleration change, because high brake torque occurs suddenly at the desired angle. This results in unsteady lift operation, because the lifted object may be thrown when a rapid stop occurs. Therefore, in the future the brake torque will be raised gently to create a smooth stop.

7.2 Results of Holding Control Experiment

Fig. 12 shows the experimental results for holding control. When control system is (a), the arm falls and vibrates due to the load of the weight. In addition, when control system is (b), torque feedback control realized stable holding control. However, a momentary vibration cannot be controlled by the slow response of air pressure. On the other hand,

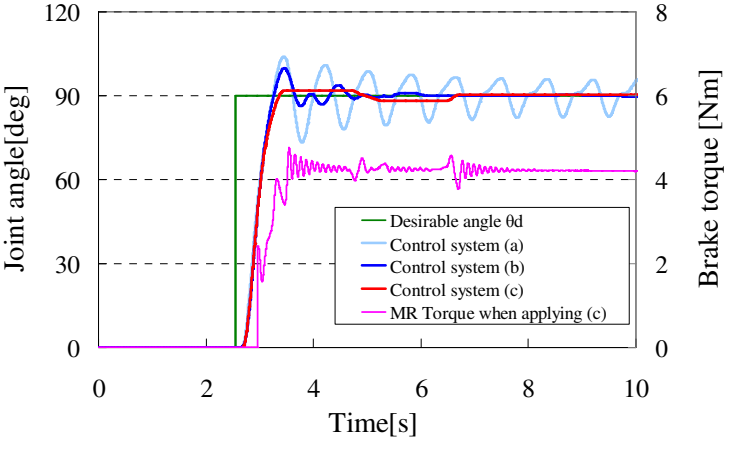


Fig. 11. A experimental results of the step response for position control

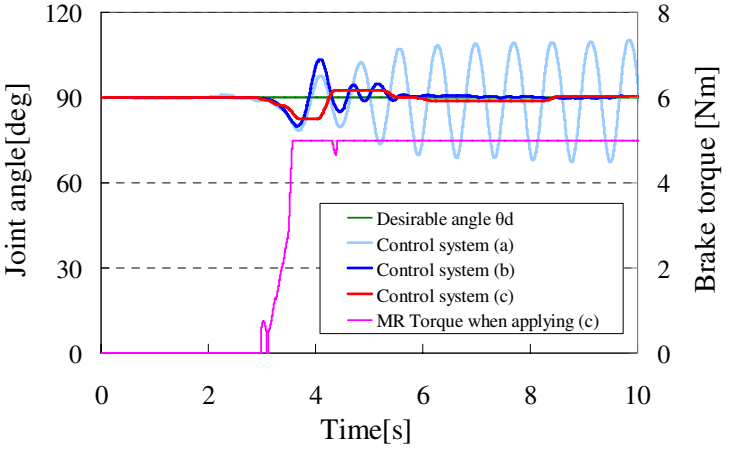


Fig. 12. A experimental results of the step response for Holding control

when control system (c), we could confirm the control of this fall and vibration. From this figure, MR brake torque is impressed according to the disturbance load.

However, a steady state error occurred, because the joint torque resulting from the load exceeds the maximum generating torque of the MR brake, and the arm fell due to the artificial muscle's high compliance characteristic. Moreover, the figure shows compensatory movement caused by PI control to counteract the angle error due to the fall. However, the arm was saturated reaching up to the desired angle.

As the cause, increased muscle burden due to the brake and load be considered. It is thought that we control an artificial muscle and MR brake independently. Therefore, the burden to the artificial muscle increases by mutually generated torque, because MR brake and an artificial muscle cannot compensate mutually. As a result, high pressure is transmitted to an artificial muscle, and the arm's angle is changed forcibly. Hence, a solution would be a control system design that enables a change of the maintenance by the artificial muscle and MR brake.

8 Conclusion

A 1-DOF artificial muscle manipulator with a variable rheological joint mechanism using MR fluid is developed. To verify its effectiveness, vibration control was investigated when an object is held and lifted. From the experimental results, we confirmed that vibration restraint was achieved by a high output and the response of the MR brake in the area where it cannot be achieved by air pressure.

In the future, the rapid acceleration change caused by the brake will be controlled continuously to realize more stable operation when an object is lifted. In addition, the vibration that occurs due to the high-speed response of the artificial muscle will be controlled. And we aim to achieve quicker operation by an artificial muscle manipulator, and the running operation and the throw operation by the artificial arm or an artificial leg.

References

[1] Park, B.J., Park, C.W., Yang, S.W., Kim, H.M., Choi, H.J.: Core-Shell Typed Polymer Coated-Carbonyl Iron Suspension and Their Magnetorheology. In: ERMR 2008, p. 102 (2008)

[2] Nickel, V.L., Perry, M.D.J., Garrett, A.L.: Development of useful function in the severely paralysed hand. Journal of Bone and Joint Surgery 45A(5), 933–952 (1963)

[3] Gavrilovic, M.M., Maric, M.R.: Positional servo-mechanism activated by artificial muscles. Medical and Biological Engineering 7, 77–82 (1969)

[4] Klute, G.K., Czernieki, J.M., Hannaford, B.: McKibben Artificial Muscles: Pneumatic Actuators with Biomechanical Intelligence. In: Proceedings of the IEEE/ASME International Conference on Advanced Intelligent Mechatronics 1999, pp. 221–226 (1999)

[5] Chou, C.P., Hannaford, B.: Static and Dynamic Characteristics of McKibben Pneumatic Artificial Muscles. In: Proceedings of IEEE International Conference On Robotics and Automation 1994, pp. 281–286 (1994)

[6] Nakamura, T.: Experimental Comparisons between McKibben type Artificial Muscles and Straight Fibers Type Artificial Muscles. In: SPIE International Conference on Smart Structures, Devices and Systems III (2006)
[7] Nakamura, T., Maeda, H., Nagai, H., Saito, H.: Position and compliance control of an artificial muscle manipulator using a mechanical equilibrium model. In: Proc. Annual Conference of the IEEE Industrial Electronics Society, IECON 2008 (2008), doi:10.1109/IECON.2008.4758512
[8] Nakamura, T., Shinohara, H.: Position and Force Control Based on Mathematical Models of Pneumatic Artificial Muscles Reinforced by Straight Glass Fibers. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA 2007), pp. 4361–4366 (2007)

A Dexterous and Self-adaptive Humanoid Robot Hand: Gesture-Changeable Under-Actuated Hand

Wenzeng Zhang, Demeng Che, Qiang Chen, and Dong Du

Key Laboratory for Advanced Materials Processing Technology, Ministry of Education, Dept. of Mechanical Engineering, Tsinghua University, Beijing 100084, China
wenzeng@tsinghua.edu.cn

Abstract. A novel concept called gesture-changeable under-actuated (GCUA) function is proposed to improve the dexterities of traditional under-actuated hands and reduce the control difficulties of dexterous hands. Based on the GCUA function, a new humanoid robot hand, GCUA Hand is designed and manufactured. The GCUA Hand can grasp different objects self-adaptively and change its initial gesture dexterously before contacting objects. The hand has 5 fingers and 15 DOFs, each finger is based on screw-nut transmission, flexible drawstring constraint and belt-pulley under-actuated mechanism to realize GCUA function. The analyses on grasping static forces and grasping stabilities are put forward. The analyses and Experimental results show that the GCUA function is very nice and valid. The hands with the GCUA function can meet the requirements of grasping and operating with lower control and cost, which is the middle road between traditional under-actuated hands and dexterous hands.

Keywords: Humanoid robot, multi-fingered hand, passive adaptive grasp, gesture-changeable under-actuated function, stability, dexterity.

1 Introduction

Humanoid robotic hand plays an important role as end-effecter of humanoid robot. Over the past 30 years, the study on dexterous hand has gained plentiful achievements. A dexterous hand has 3~5 fingers with 2~4 DOFs in each finger, whose joints are mostly driven by actuators actively. Examples of dexterous robot hands are Utah/MIT Hand [1], Robonaut Hand [2], UB Hand [3], BH series hands by Beihang University and HIT series hands [4] by Harbin Institute of Technology. Dexterous robot hands can grasp different kinds of objects agilely and stably. However, dexterous hands cannot grasp objects self-adaptively, which makes them highly depend on their sensors and control system. The complexity of dexterous hands makes them high cost and low reliability.

Under-actuated robotic hands can overcome some weaknesses of dexterous hands whose joints are mostly driven by actuators forwardly. Thus under-actuated robot hands are more and more important in recent 10 years. In the research of under-actuated robot hands, Laval University (Birglen et al. [5]) in Canada, MIT (Dollar et al. [6]) in USA, HIT (Liu et al. [7]), Beihang University (Guo et al. [8]), and Tsinghua University (Zhang et al. [9]), have done many jobs. Under-actuated robotic hands can grasp ob-

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 515–525, 2009.

jects with different shapes and sizes self-adaptively, which makes hands reduce the requirements of sensor and control system.

The traditional under-actuated finger has the gesture which cannot be changed (mostly keep straight) until the finger touches objects, which makes the finger less human-like and less stable to grasp different objects. For instance, with the constraint of spring, each finger of TH-3 hand must keep straight (the angle between middle- and terminal-segment is straight angle) only if it is blocked by objects, which reduces the stability and dexterity of grasping.

2 Principle of GCUA Function

2.1 Concept of GCUA Function

A novel concept called gesture-changeable under-actuated (GCUA) function is proposed to improve the dexterities of traditional under-actuated hands and reduce the control difficulties of dexterous hands. It is called "gesture-changeable": change the finger initial gesture before grasping. The gesture-changeable under-actuated (GCUA) mechanism can change the finger's initial gesture flexibly according to the different sizes and shapes of the grasped objects, and then grasp objects passively adaptively. The gesture-changeable under-actuated function can make the finger pinch relatively small objects easily, simultaneously make the grasping process more humanlike (People always bend the middle joints of fingers to grasp relatively small objects).

In fact, adding the GCUA function makes under-actuated finger more like dexterous finger rather than traditional under-actuated finger. The GCUA mechanism adds the dexterity of dexterous hands to the under-actuated mechanism by changing the finger's initial gesture, at the same time, eliminates the weakness that dexterous hands highly depend on their sensors and control systems by the self-adaptive grasping function. The GCUA hand will be the middle road between under-actuated hand and dexterous hand.

2.2 Grasping Process of Finger with GCUA Function

Fig. 1. shows the under-actuated grasping process of a finger with GCUA function. The motor rotates on the base, driving the gear, and then the first joint-shaft rotates. Before I-segment contacts the grasped object, a spring located around the second joint-shaft and a spring around the third joint-shaft keep II-segment and III-segment being static relative to I-segment as long as possible, resulting in rotation of the three segments as a rigid body around the first joint-shaft.

When I-segment is blocked by the grasped object and unable to continue rotating, the motor keep rotating to drive II-segment and III-segment rotates against the elastic force of the return spring as a rigid body, until also touching the object. When II-segment is blocked by the grasped object, the III-segment rotates continuously.

Fig. 2. shows the second movement process of the GCUA function: the finger changes the initial gesture firstly and then grasps the object self-adaptively. The motor located in II-segment rotates, driving the gear. Through the screw-nut transmission and flexible drawstring, the motor drives II-segment to rotate against the elastic force of the return spring around the second joint-shaft. II-segment has already rotated an angle

before the motor stops rotating. The flexible drawstring can prevent II-segment from rotating backward, however, allow II-segment to rotate forward with the driving of other actuators. Therefore, the pre-bending motion has no effect on the under-actuated grasping process.

The process of II-segment rotating backward is like this: the motor rotates backward, driving the gear. Then with the screw-nut transmission, the motor drives the flexible drawstring not restrict II-segment any more. The deformation of the return spring gets smaller gradually, with II-segment rotating backward to be the initial status.

These two processes mentioned above, can be used alone or combined under the different requirements of grasping. The GCUA function can not only be self-adaptive to the shapes and sizes of the grasped objects with the under-actuated grasping process, but also grasp small size and light objects or gliding objects with the pre-bending motion stably.

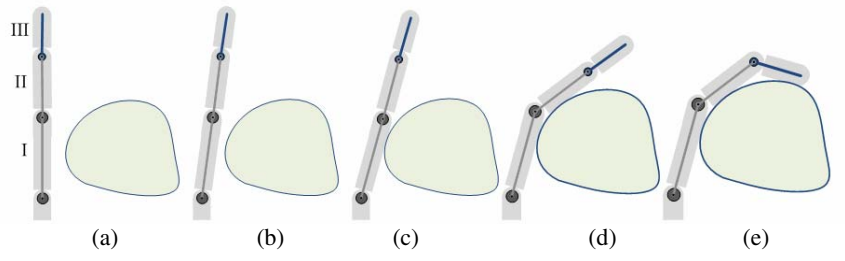


Fig. 1. Under-actuated grasping process of the GCUA finger

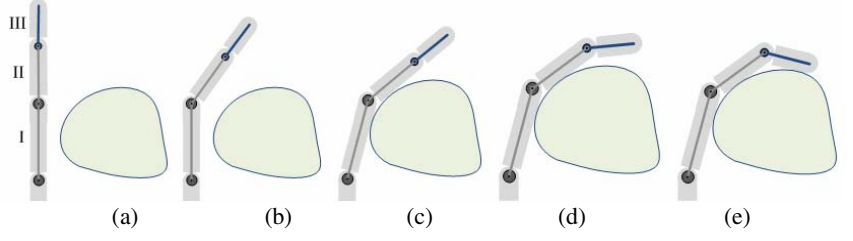


Fig. 2. Pre-bending motion and Under-actuated grasping

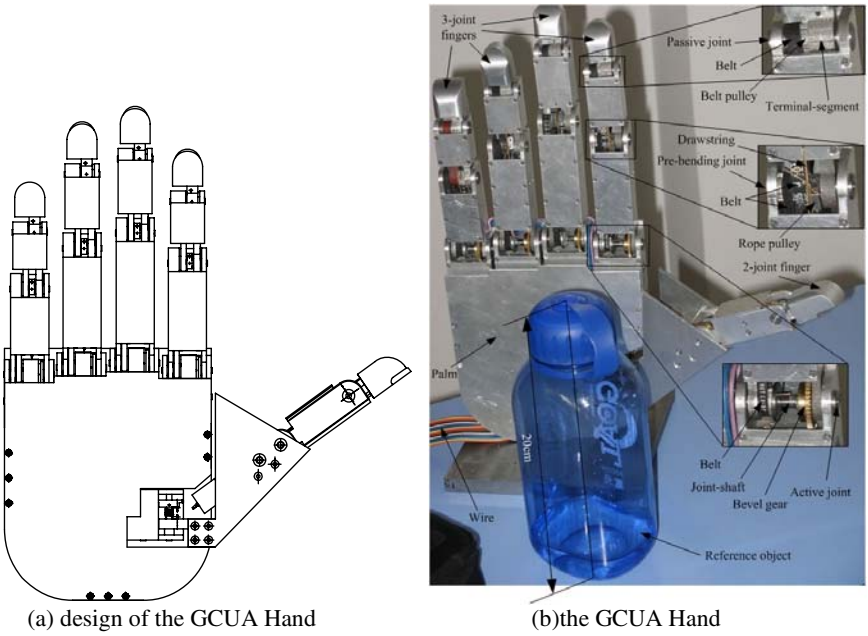
3 Design of the GCUA Hand

We will describe the design of GCUA Hand which is based on the gesture-changeable under-actuated function. The hand has 5 fingers and 15 DOFs(6 active DOFs, 5 pre-bending DOFs, 4 passive DOFs).

GCUA Hand is the double size with an adult's hand, with a length 350mm, palm length of 180mm, palm width of 152mm, palm thickness of 30mm, all the fingers can rotate angle range of 0~90° for all joints. The details of the hand and the placement of components are shown in Fig. 3.

3.1 Design of the 2-Joint Finger

The component of the 2-joint finger mechanism is shown in Fig. 4. The first motor is fixed with the base, whose output shaft is fixed with the first bevel gear, the first bevel gear is meshing with the second one which is fixed with the base joint-shaft, the base joint-shaft is sleeved within the base. The first straight gear is also fixed with the base joint-shaft, which meshes with the second one fixed with the first joint-shaft. The first-shaft is sleeved within the base, the second joint-shaft is sleeved within the middle-segment, and fixed with the terminal-segment.



(a) design of the GCUA Hand (b)the GCUA Hand

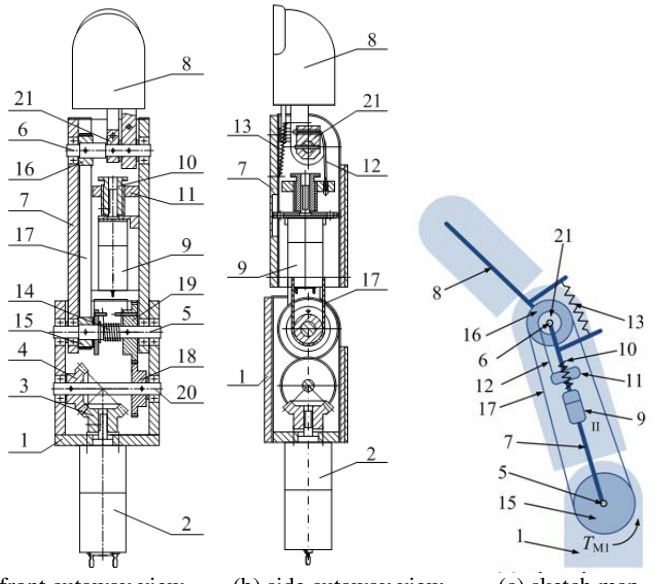
Fig. 3. Details of GCUA Hand and placement of components

At the same time, the return spring around the second joint-shaft connects the middle and terminal segment by its two ends, the active spring around the first joint shaft connects the active pulley and the first joint, which can eliminate the torque produced when the active pulley reverses.

The second motor is fixed in the middle-segment, whose output shaft is fixed with the screw, and the screw is connected with the nut. The flexible drawstring connects the nut and the rope pulley by its two ends, the rope pulley is fixed with the second joint-shaft.

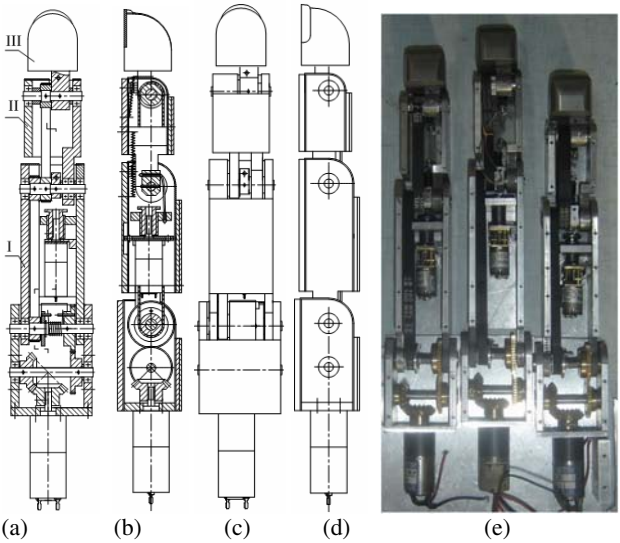
3.2 Design of the 3-Joint Finger

The 3-joint GCUA finger has two motors to drive 3 DOFs, which can change the initial gesture (the angle between I-segment and II-segment) with the pre-bending motion and grasp different objects with the under-actuated grasping process self-adaptively.



(a) front cutaway view (b) side cutaway view (c) sketch map

Fig. 4. The GCUA finger with 2 joints. 1-base; 2-first motor; 3-first bevel gear; 4-second bevel gear; 5-first joint-shaft; 6-second joint-shaft; 7-middle-segment; 8-terminal-segment; 9-second motor; 10-screw; 11-nut; 12-flexible drawstring; 13-return spring; 14-active spring; 15-active pulley; 16-passive pulley; 17-belt; 18-first straight gear; 19-second straight gear; 20-base joint-shaft; 21-rope pulley.



(a) (b) (c) (d) (e)

Fig. 5. The GCUA finger with 3 joints. (a) front cutaway view; (b) side cutaway view; (c) front view; (d) side view; (e) photo of three 3-joint fingers.

Fig. 5. shows the design of the 3-joint GCUA finger. The 3-joint GCUA finger uses the same principle like 2-joint GCUA finger, which has the first motor located in the base to drive the finger to grasp different objects self-adaptively (Under-actuated grasping process), the second motor located in I-segment to drive the finger to change the initial gesture (Pre-bending motion). The four 3-joint fingers of GCUA Hand have the same structure but different sizes, each of which can grasp objects independently.

4 Analyses of the GCUA finger

4.1 Force Analysis of Pre-bending Motion

Fig. 6a. shows the force analysis of 3-joint GCUA finger when it changes its initial gesture. The letter "I" stands for the I-segment, "II" stands for the II-segment.

T_{M2}-the torque caused by the second motor to the screw with regard to II-segment, second motor torque for short, Nmm;

F_{M2}-the force exerted by the screw to the flexible drawstring, which is caused by T_{M2}, N;

T_{S1}-the torque caused by the return spring between I-segment and II-segment with regard to the II-segment joint-shaft, spring torque for short, Nmm;

F_P-the grasping force exerted by the II-segment to the object, N;

r_3,r_P-the radii of passive pulley and rope pulley, mm;

h_P- the arm of force F_P with regard to the II-segment joint-shaft, mm;

θ_P-the rotational angle of the II-segment, rad;

ψ-the lead angle of the screw, rad.

According to the static force analysis, when the II-segment touches the grasped object, using the principle of virtual work, one obtains

$$F_{M2}=CT_{M2} \tag{1}$$

$$F_{M2}\cdot r_P+T_{S1}=F_P\cdot h_P \tag{2}$$

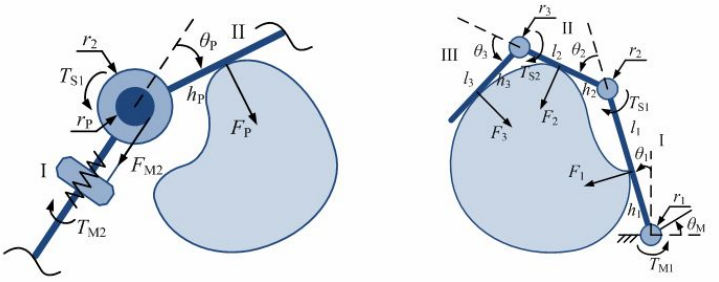
where $C=2/(d\cdot\tan\psi)$, $T_{S1}=-K_1\theta_P$, K_1 is the stiffness of the return spring. Combining eq. (1), (2), one obtains

$$F_P=J_P^{\ T}T_P \tag{3}$$

where $J_P=[2r_P/(dh_P\tan\psi)\ 1/h_P]^T$, $T_P=[T_{M2}\ T_{S1}]$. The main object of the pre-bending motion is changing the angle between I-segment and II-segment, at the same time, it can also provide grasping force.

4.2 Force Analysis of Under-Actuated Grasping Process

Although pre-bending motion can provide grasping force, the main grasping force is provided by under-actuated grasping process. With the unilateral constraint of the flexible drawstring, pre-bending motion has little effect on under-actuated grasping process, therefore, we can consider under-actuated grasping process only. Fig. 6b. shows the static force analysis of 3-joint GCUA finger when under-actuated grasping process carries on. I stands for I-segment whose length is l_1, mm, II stands for II-segment, whose length is l_2, mm, III stands for III-segment whose length is l_3, mm.



(a) changing the initial gesture (b) under-actuated grasping process

Fig. 6. Static force analysis of the GCUA finger

F_1-the grasping force caused by the I-segment to the object, whose magnitude is equal to the reaction force caused by the object against the I-segment, N;

F_2-the grasping force caused by the II-segment to the object, whose magnitude is equal to the reaction force caused by the object against the II-segment, N;

F_3-the grasping force caused by the III-segment to the object, whose magnitude is equal to the reaction force caused by the object against the III-segment, N;

T_{M1}-the torque caused by the first motor to the active pulley with regard to the I-segment joint-shaft, motor torque for short, Nmm;

T_{S1}-the torque caused by the return spring between I and II segments with regard to the II-segment joint-shaft, first spring torque for short, Nmm;

T_{S2}-the torque caused by the return spring between II and III segments with regard to the III-segment joint-shaft, second spring torque for short, Nmm;

r_1, r_2, r_3- the radii of the active, first passive and second passive pulley, mm;

θ_M-the rotational angle of the active pulley, rad;

θ_1 - the rotational angle of the I-segment, rad;

θ_2 - the rotational angle of the II-segment, rad;

θ_3 - the rotational angle of the III-segment, rad;

h_1- the arm of force F_1 with regard to the I-segment joint-shaft, mm;

h_2- the arm of force F_2 with regard to the II-segment joint-shaft, mm;

h_3- the arm of force F_3 with regard to the III-segment joint-shaft, mm.

According to the principle of virtual work, one gets

$$T^{\mathrm{T}}\Omega = F^{\mathrm{T}}V \tag{4}$$

where T is input torque vector by the first motor and the return springs, F is the grasping force vector by the three segments of the finger, V is the velocity vector on the contact points, i.e.

$$T = \begin{bmatrix} T_{M1} \\ T_{S1} = -K_1\theta_2 \\ T_{S2} = -K_2\theta_3 \end{bmatrix} \tag{5}$$

$$F=[F_1\ F_2\ F_3]^{\mathrm{T}} \tag{6}$$

where K_1, K_2 are the stiffness of the return spring located around the second joint-shaft and the third joint-shaft respectively.

$$V=\begin{bmatrix} h_1 & 0 & 0 \\ l_1\cos\theta_2+h_2 & h_2 & 0 \\ l_1\cos(\theta_2+\theta_3)+l_2\cos\theta_3+h_3 & l_2\cos\theta_3+h_3 & h_3 \end{bmatrix}\begin{bmatrix} \dot{\theta}_1 \\ \dot{\theta}_2 \\ \dot{\theta}_3 \end{bmatrix} \tag{7}$$

$$\Omega=\begin{bmatrix} \dot{\theta}_M & \dot{\theta}_2 & \dot{\theta}_3 \end{bmatrix}^T \tag{8}$$

where $\dot{\theta}_M$ is the rotational velocity of the first motor, $\dot{\theta}_1$ is the rotational velocity of the I-segment, $\dot{\theta}_2$ is the rotational velocity of the II-segment, $\dot{\theta}_3$ is the rotational velocity of the III-segment. Let

$$A_V=\begin{bmatrix} h_1 & 0 & 0 \\ l_1\cos\theta_2+h_2 & h_2 & 0 \\ l_1\cos(\theta_2+\theta_3)+l_2\cos\theta_3+h_3 & l_2\cos\theta_3+h_3 & h_3 \end{bmatrix} \tag{9}$$

Considering the rotary process of the finger, one gets

$$\theta_M=\theta_1+\alpha_1+\alpha_2,\ \alpha_1 r_1=r_2\theta_2,\ \alpha_2 r_1=r_3\theta_3 \tag{10}$$

where α_1 is the rotational angle of the II-segment after the I-segment is blocked by the grasped object, α_2 is the rotational angle of the III-segment after the II-segment is blocked by the grasped object, from eq. (10), one gets

$$\theta_M=\theta_1+\frac{r_2}{r_1}\theta_2+\frac{r_3}{r_1}\theta_3 \tag{11}$$

Differentiating the eq. (11), one can obtain

$$\dot{\theta}_M=\dot{\theta}_1+\frac{r_2}{r_1}\dot{\theta}_2+\frac{r_3}{r_1}\dot{\theta}_3 \tag{12}$$

Combining eq. (8) and eq. (12), one obtains

$$\Omega=A_\Omega\Theta \tag{13}$$

where $\Theta=\begin{bmatrix} \dot{\theta}_1 & \dot{\theta}_2 & \dot{\theta}_3 \end{bmatrix}^T$, $A_\Omega=\begin{bmatrix} 1 & r_2/r_1 & r_3/r_1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. Combining eq. (4), (5), (6), (7), (9), and eq. (13), one gets

$$T^T A_\Omega=F^T A_V \tag{14}$$

Calculate the determinants of matrix A_Ω and A_V, we know $|A_\Omega|=1>0$, $|A_V|=h_1h_2h_3$, when h_i isn't zero, with matrix manipulation, one obtains

$$F=A_V^{-T}A_\Omega^T T \tag{15}$$

Let $A_F=A_V^{-T}A_\Omega^T$, one obtains

$$F=A_F T \tag{16}$$

Let $k_i=(l_i\cos\theta_{(i+1)}+h_{(i+1)})/(h_i h_{(i+1)})$ (i=1,2), $R_i=r_{(i+1)}/r_1$ (i=1,2) and $s=l_1\cos(\theta_2+\theta_3)/h_2$, one gets

$$A_F = \begin{bmatrix} \frac{1-h_2k_2R_2}{h_1} - R_1k_1 + h_2k_1k_2R_2 - sR_2 & \frac{R_1}{h_2} - R_2k_2 & \frac{R_2}{h_3} \\ -k_1 & \frac{1}{h_2} & 0 \\ h_2k_1k_2 - s - \frac{h_2k_2}{h_1} & -k_2 & \frac{1}{h_3} \end{bmatrix}^{\mathrm{T}} \tag{17}$$

The return spring torques T_{S1} and T_{S2} is quite small compared to the motor torque T_{M1}, and might as well be omitted. From eq. (16), one knows F_3 is always large than zero. Let l_i, h_i, R_i, T_{M1} as constants (l_i=2, h_i, R_1=2, R_1=1, T_{M1}=100, SI). The grasping stability area of the finger can get, shown in Fig. 7.

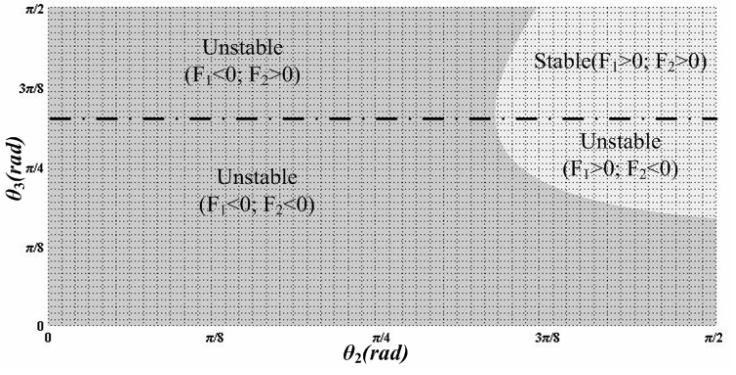


Fig. 7. Grasping stability area when R_1=2, R_2=1

When state point (θ_2, θ_3) is in the stable area, all the three grasping force are positive, thus, the finger can grasp the object stably. On the contrary, if point (θ_2, θ_3) is in the unstable area, at least one force is negative, so the finger cannot grasp the object stably. Fig. 8. shows that the stability with different magnitudes of R_i.

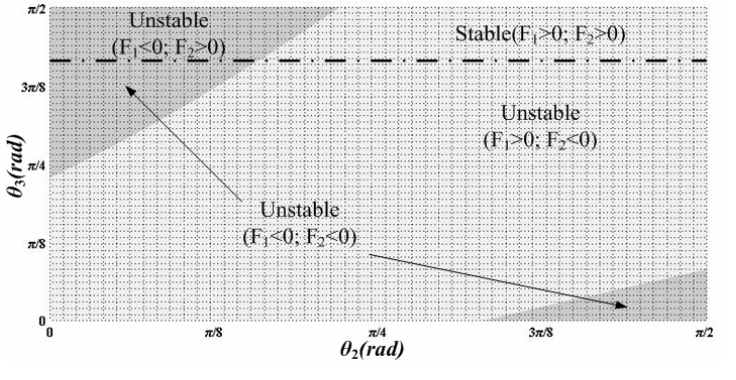


Fig. 8. Grasping stability area when R_1=1.5, R_2=1

5 Grasping and Operation Experiments

Fig. 9 shows some pictures of the hand grasping experiments. The hand can grasp objects of different sizes and shapes with under-actuated grasping process like traditional under-actuated hand. Moreover, the hand can also grasp many small or gliding objects, like tennis ball, apple MP3, and so on. What's more, the hand can do some operations or poses like touching keyboard, playing electronic organ, and so on, which in shown in Fig. 10.

With the GCUA function, the hands can not only grasp objects but also operate instruments. Combining pre-bending motion with under-actuated grasping process makes the finger more humanoid, more dexterous, easier to control, lower manufacture and maintenance cost.

Fig. 9. GCUA Hand grasping different objects

Fig. 10. Operations and poses of the GCUA Hand

6 Conclusions

This paper proposed a design idea of gesture-changeable under-actuated (GCUA) hand which can achieve the pre-bending motion and the under-actuated grasping process. The analyses and Experimental results show that the GCUA function is very nice and valid. GCUA Hand can grasp different objects self-adaptively like traditional under-actuated hands. Furthermore, it can also operate some instruments or make humanoid poses as a dexterous hand. Although it is less dexterous than dexterous hand, it can meet many needs of grasping or operating with a lower manufacture and

maintenance cost. The hands with the GCUA function will be the middle road between traditional under-actuated hands and dexterous hands.

Acknowledgements

This paper was funded by the National Natural Science Foundation of China (No. 50905093), Hi-Tech R&D Program (2007AA04Z258).

References

1. Jacobsen, S.C., Iversen, E.K., Knutti, D.F., et al.: Design of the UTAH/M.I.T. dextrous hand. In: Proc. IEEE Inter. Conf. on Robotics and Automation, April 1986, pp. 1520–1532. IEEE Computer Society Press, San Francisco (1986)
2. Lovchik, C.S., Diftler, M.A.: The robonaut hand: a dexterous robot hand for space. In: Proc. IEEE inter. Conf. on Robotics & Automation, May 1999, pp. 907–912. IEEE Computer Society Press, Detroit (1999)
3. Lotti, F., Tiezzi, P., Vassura, G.: Development of UB hand 3: early results. In: Proc. IEEE Inter. Conf. on Robotics & Automation, Barcelona, Spain, April 2005, pp. 4488–4493 (2005)
4. Liu, H., Meusel, P., Seitz, N., et al.: The modular multisensory DLR-HIT-Hand. Mechanism and Machine Theory 42, 612–625 (2007)
5. Birglen, L., Gosselin, C.M.: Kinetostatic analysis of underactuated fingers. IEEE Transactions on Robotics and Automation 20(2), 211–221 (2004)
6. Dollar, A.M., Howe, R.D.: Towards grasping in unstructured environments: grasper compliance and configuration optimization. Advanced Robotics 19(5), 523–543 (2005)
7. Liu, H., Gao, X., Shi, S.: Under-actuated self-adaptive artificial hand: Chinese Patent: CN 1292719C (November 2007)
8. Guo, W., Liu, B., Zhang, Y., et al.: Humanoid robot hand mechanism: Chinese Patent: CN 1283429C (November 2006)
9. Zhang, W., Che, D., Liu, H., et al.: Super under-actuated multi-fingered mechanical hand with modular self-adaptive gear-rack mechanism. Industrial Robot: An International Journal 36(3), 255–262 (2009)

Design of the Upper Limb Rehabilitation Support Device Using a Pneumatic Cylinder

Koichi Kirihara[1], Norihiko Saga[2], and Naoki Saito[1]

[1] Akita Prefectural University, Department of Machine Intelligence and System Engineering
84-4 Tsuchiya, Yurihonjo, Akita, Japan
[2] Kwansei Gakuin University, Department of Human System Interaction
2-1 Gakuen, Sanda, Hyogo, Japan
{m10a004,naoki_saito}@akita-pu.ac.jp,
saga@kwansei.ac.jp

Abstract. This paper describes a device to support rehabilitation of a patient's upper limb motion. For safety, light weight, and flexibility, it uses a pneumatic cylinder, for which the optimum arrangement is presented. The rehabilitation support device has two rehabilitation modes corresponding to different rehabilitation contents. A compliance control system and a position control system are applied for those modes. We evaluate the effectiveness of the rehabilitation support mode through some experimentation.

Keywords: Rehabilitation support, Pneumatic actuator.

1 Introduction

In recent years in Japan, the number of handicapped people has increased because of sickness, injury, aging, etc. Furthermore, according to an annual report on the aging society in Japan, a hyper-aging society will exist in the near future [1], [2]. Because of the physical handicaps of elderly people and handicapped people, activities of daily life (ADL), such as eating, will become difficult. They cannot live without the exertion of great effort. Moreover, the burdens will increase not only for handicapped people, but also for caregivers such as families [3]. Therefore, handicapped people will undergo rehabilitation to establish or increase their ability to use upper limbs.

In rehabilitation training, handicapped people (patients hereinafter) and the occupational therapists (OTs hereinafter) train together. An OT demonstrates and facilitates motions that give a constant or varied load to a patient's upper limbs and moves their upper limbs, repeating joint's flexion and extension at a constant speed. The OT repeats the operation while considering the level of the patient's trouble. Therefore, we think that the machine can often substitute for the OT's motion in rehabilitation. The device is expected to facilitate rehabilitation for patients and OTs. Some rehabilitation devices and robots have been developed [4]–[11]. However, most are large and complex mechanisms. Moreover, few examinations have investigated rehabilitation devices and robots using the pneumatic actuator. In a clinical setting,

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 526–534, 2009.

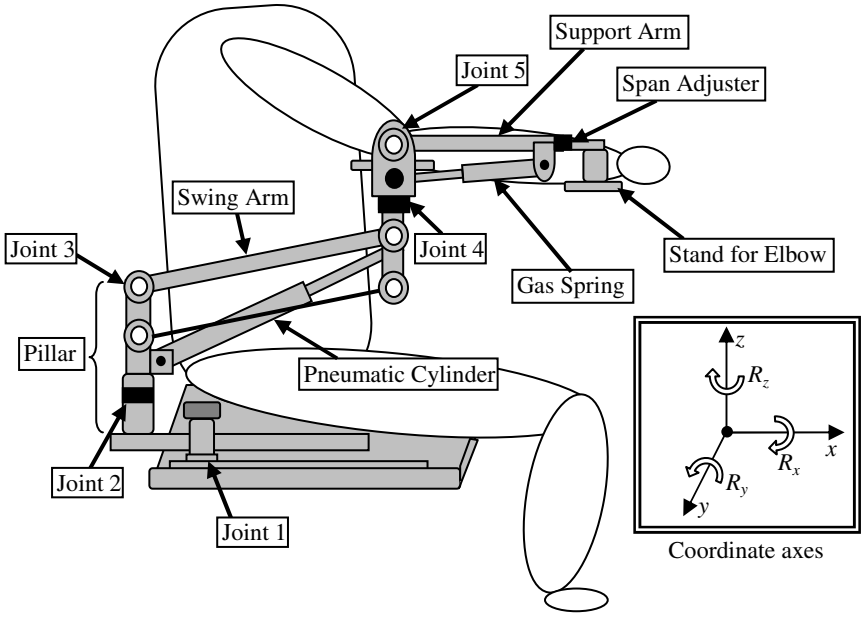


Fig. 1. A structure of the rehabilitation support device

such a device is expected to be a simple mechanism with a simple control system that is easy to use. For this study, we developed a compact, lightweight, upper-limb rehabilitation support device using a pneumatic cylinder with a link mechanism. Medical and welfare apparatus and systems such as the rehabilitation support device that we developed must be safe, flexible, and lightweight because the device must have contact with patients during its operation. An electric motor and a hydraulic actuator are used for almost all robots. The system would become complex and bulky if we were to use these actuators as medical and welfare apparatus. Therefore, we used a pneumatic cylinder to operate the device because shock can be absorbed using compressibility of air: it is lightweight and slim, and it has a high power–weight ratio.

In this study, a position control system and a compliance control system are applied on the device to replace an OT's motion of rehabilitation training. Some experiments were performed to evaluate the device and its control system.

2 Rehabilitation Support Device

Fig.1 depicts the structure of the upper limb-rehabilitation support device. The device has five degrees of freedom (DOF) by virtue of its link mechanism [12] and consists of joint 1 – joint 5. Joint 1 reciprocates on the *x*-axis by a linear guide to support the upper limb for the reach motion. Joint 3, with an attached pneumatic cylinder, rotates around the *y*-axis to support flexion and extension of the shoulder joint actively. Joint 5, with an attached gas spring (Y0061, Tokico; Hitachi Ltd.), rotates around the *y*-axis

to support the elbow joint's external rotation and internal rotation. Joint 2, with an attached rotation joint, rotates around the z-axis to support the shoulder joint's horizontal flexion and extension. Joint 4, with an attached rotation joint, rotates around the z-axis to support the elbow joint's flexion and extension.

2.1 Actuator Layout

In designing the device, we considered the actuator layout. In this report, we suggest two layouts that the pneumatic cylinder uses for the actuator of the device. We then draw a comparison of the two methods based on their respective ranges of motion and torque variation.

Fig. 2(a) portrays a schematic diagram showing the arrangement of the pneumatic cylinder, arranged diagonally in the device. In addition, Fig. 2(b) shows a schematic diagram of the pneumatic cylinder arranged vertically in the device.

Here, the overall length of the pneumatic cylinder (diagonally arranged type) is denoted as l_d; the overall length of the pneumatic cylinder (vertically arranged type) is denoted as l_v. The angle between the swing arm and pneumatic cylinder is denoted as α, the angle between the vertical line and pneumatic cylinder is represented as γ. Each relation is as shown below.

$$l_d = \sqrt{a^2 + L^2 - 2aL\cos\theta} \tag{1}$$

$$l_v = \sqrt{\{L(1-\cos\beta)\}^2 + (A+B)^2} \tag{2}$$

$$\alpha = \cos^{-1}\left(L^2 + {l_d}^2 - a^2 / 2Ll_d\right) \tag{3}$$

$$\gamma = \cos^{-1}\left[l^2 + (A+B)^2 - \{L(1-\cos\beta)\}^2 / 2l(A+B)\right] \tag{4}$$

Therein, θ is the angle between a pillar and swing arm; L denotes the length of the swing arm. In addition, a stands for the distance from center of rotation to the pneumatic cylinder's connector. Torque τ_d and τ_v are expressed as follows.

$$\tau_d = FL\sin\alpha \tag{5}$$

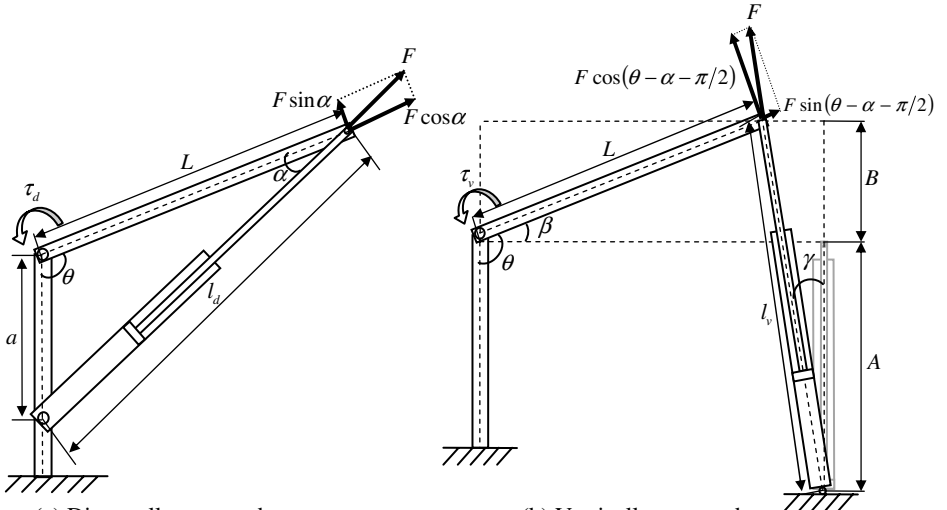
$$\tau_v = FL\cos(\theta - \alpha - \pi/2) \tag{6}$$

The force F generated by the pneumatic cylinder is obtained from the experiment. Moreover, the strokes of the pneumatic cylinder x_d and x_v are expressed as shown below.

$$x_d = l_d - l_i \tag{7}$$

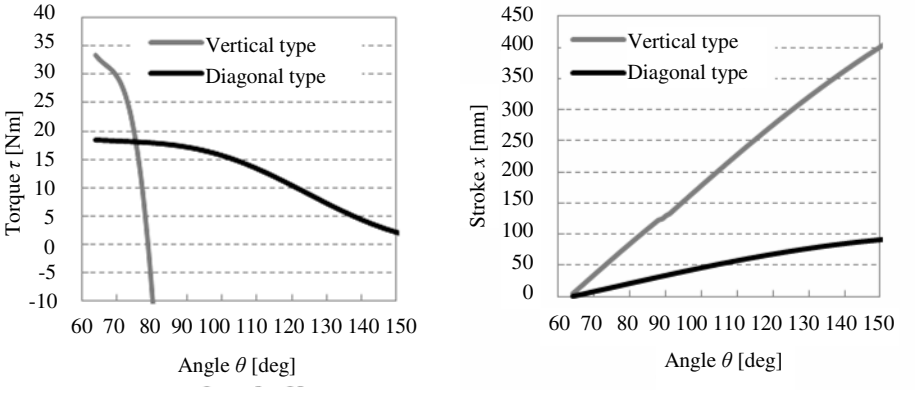
$$x_v = l_v - l_i \tag{8}$$

In those equations, l_i signifies the initial length of the pneumatic cylinder.



(a) Diagonally arranged type (b) Vertically arranged type

Fig. 2. Schematic diagram of two suggested arrangements



(a) Calculated results for rotation angle and torque

(b) Calculated results for the rotation angle and pneumatic cylinder stroke

Fig. 3. Actuator layout characteristics

Fig. 3(a) shows calculated results for the rotation angle and torque. Fig. 3(b) portrays the calculated results for the rotation angle and stroke of the pneumatic cylinder. As shown in Fig. 3(a), for a small rotation angle, the torque is very high for the vertically arranged type. However, the torque decreases rapidly as the rotation angle increases. Meanwhile, the torque is stable throughout the whole range of motion in the case of a diagonally arranged type. Furthermore, as presented in Fig. 3(b), the stroke in the case of the vertically arranged type is longer than that of the diagonally

arranged type. This result confirms that the device becomes large and bulky in the case of vertically arranged type. We infer that the diagonally arranged type has an advantage over the vertically arranged type. Therefore, the diagonally arranged type was adopted.

2.2 Rehabilitation Support Functions

The rehabilitation support device has two rehabilitation modes corresponding to the difference of rehabilitation contents.

The first mode is muscular power recovery and movable region expansion mode (Mode A). In this mode, the device gives a load to the patient's upper limb, as portrayed in Fig. 4(a). Then, the patient repeats flexion and extension while resisting the load from the device. In this mode, compliance control was applied to operate as an OT. The patient can conduct rehabilitation to the level of the patient's muscular power.

The second mode is a practical function of the upper limb recovery mode (Mode B). In this mode, the device supports training, which operates the upper limb on the desk, as portrayed in Fig. 4(b). The patient trains to use a "peg board" etc. while the device supports the arm. A position control (rotation angle of joint 3) is applied to support an upper limb's vertical motion. As shown for mode B, a patient with trouble operating the upper limb to resist gravity can easily train on a desk.

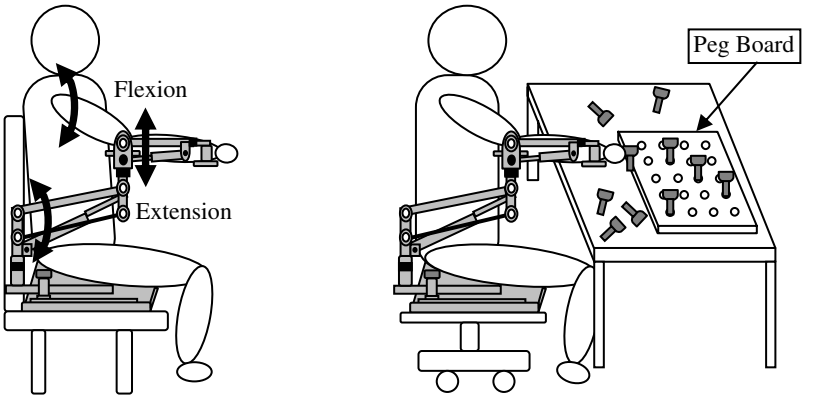


Fig. 4. Rehabilitation support modes

2.3 Device Control System

Fig. 5 depicts the device control system. The electro pneumatic regulator (ETR200-1; Koganei Corp.) regulates the pneumatic cylinder's (T-DA20×100; Koganei Corp.) inner pressure. A rod in the pneumatic cylinder expands and contracts when the pneumatic cylinder's inner pressure changes. The swing arm rotates around the *y*-axis. The rotation angle of joint 3 is measured using the rotary position sensor. The load

cell (LMA-A-100N; Kyowa Electronic Instruments Co. Ltd.), installed in a stand for the elbow, measures the force that the patient is adding.

A compliance control system is applied to change the stiffness of joint 3 [13], [14]. The compliance control equation is written as

$$\tau = K(\theta_d - \theta). \tag{9}$$

However, θ_d is the desired angle, θ is the measured angle, τ is the torque of the joint 3, and K is the constant of stiffness. In addition, $d\theta$ is defined as the difference between the desired angle and the measured angle ($\theta_d - \theta$).

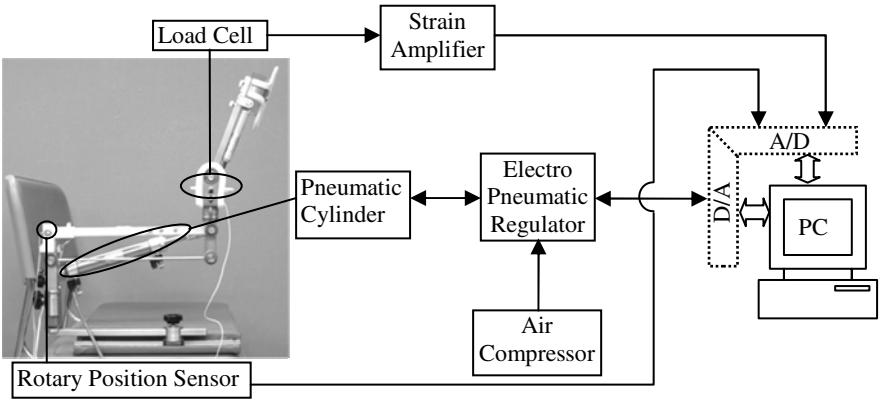


Fig. 5. Control system of the device

3 Experiments and Evaluation

In this section, we describe a compliance control for Mode A and a position control for Mode B. Furthermore, we evaluate the effectiveness of the rehabilitation support mode through experimentation.

3.1 Experiment Confirming Compliance Control for Mode A

The rehabilitation support device is fixed with a jig so that the rotation angle θ might be 90 deg. We measured the $d\theta$ and generated torque τ.

Fig. 6 presents experimental results of compliance control. The solid line represents the theoretical value of the generated torque from eq. (9). The gray solid line shows torque according to the weight of the arm of a typical adult male (65.7 kg body weight; arm weight 3.2 kg). Comparison of experimental results and theoretical values shows a strong correlation. Fig. 6 shows the generated torque as 18 Nm; the torque by the arm weight is 8.9 Nm, as depicted by the gray solid line. Sufficient margins exist from the torque by the weight of the arm to the limit of the generation torque. Therefore, the patient can add force from the state to put the arm on the device.

We confirmed that the stiffness of joint 3 rose by raising the constant of the stiffness through this experiment. When actually using the device for rehabilitation, we

assume that the constant of stiffness is set low for a patient with weak muscles, and that the constant of stiffness is set high for patients with strong muscles, presumably those in advanced stages of recovery.

3.2 Experiment Position Control for Mode B

This experiment is performed with and without a load (wrist part, 1 kg; elbow part, 1.8 kg), which assumes the weight of a human's arm. The loads of the wrist and elbow part were estimated using the ratio of the weight of each part to the weight of a human. Moreover, the target value was given from 110 deg to 90 deg in the ramp input, which was assumed to represent extension of the arm (shoulder joint).

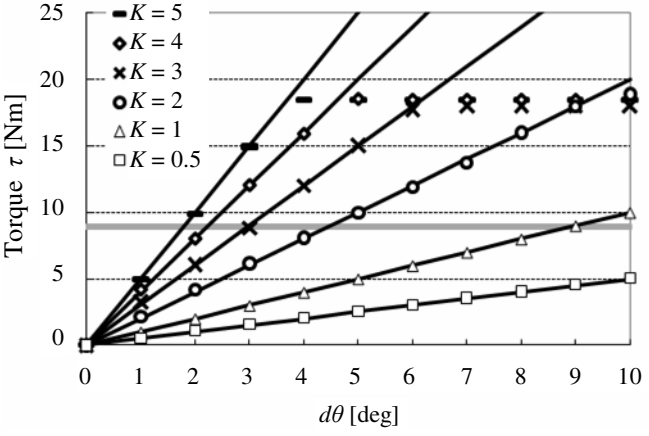


Fig. 6. Experimental results of compliance control

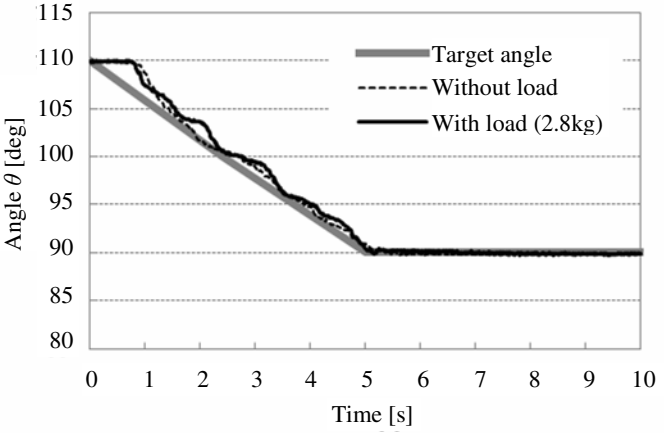


Fig. 7. Experimental results of position control

Fig. 7 portrays experimental results of position control. The rotation angle smoothly followed the target value without overshooting. It converged to the target angle (90 deg). Therefore, when the device is used for assistance of rehabilitation training on a desk, the patient's arm can be moved to the position that the patient desires. The device is useful safely, without giving discomfort to the patient.

4 Conclusions

In this study, we developed an upper-limb-rehabilitation support device using a pneumatic cylinder. A summary of the obtained results is presented as follows.

By arranging the pneumatic cylinder optimally, the device is compact, but provides a wide range of movement.

The device has two rehabilitation support modes corresponding to different rehabilitation contents. In Mode A, to support a patient's muscular power recovery and movable region expansion, a compliance control system was applied. In Mode B, to support recovery of a patient's practical function of the upper limb, a position control system was applied.

A compliance control performance for Mode A was verified experimentally, showing high correlation with measured values and theoretical values of torque of joint 3.

The position control performance for Mode B was verified experimentally. The results confirm that the rotation angle of joint 3 smoothly followed the target angle.

These results confirmed that the device we developed can support a patient's training.

References

1. Government of Japan, Cabinet office: Annual report on the Aging Society (2006)
2. Saga, N., Saito, N., Chonan, S.: Development of a Support Arm System Using Artificial Muscle Actuator and Gas spring. In: 2nd Frontires in Biomedical Devices Conference, Irvine (2007)
3. Saga, N., Saikawa, T., Okano, H.: Flexor Mechanism of Robot Arm Using Pneumatic Muscle Actuators. In: IEEE International Conference on Mechatronics & Automation, Canada, pp. 1261–1266 (2005)
4. Kiguchi, K., Tanaka, T., Watanabe, K., Fukuda, T.: Design and Control of an Exoskeleton System for Human Upper-Limb Motion Assist. In: IEEE/ASME International Conference on Advanced Intelligent Mechatronics, Japan, pp. 926–931 (2003)
5. Bien, Z., Kim, D.-J., Chung, M.-J., Kwon, D.-S., Chang, P.H.: Development of a Wheelchair-based Rehabilitation Robotic System (KARESII) with Various Human-Robot Interaction Interfaces for the Disabled. In: Advanced Intelligent Mechatronics, Japan, pp. 902–907 (2003)
6. Noritsugu, T., Ando, F., Yamanaka, T.: Rehabilitation Robot Using Rubber Artificial Muscle (1st Report Realization of Exercise Motion with Impedance control). Journal of RSJ 13(1), 141–148 (1995)

7. Yamamoto, K., Hyodo, K., Ishii, M., Matsuno, T.: Development of Power Assisting Suit for Assisting Nurse Labor. JSME International Journal Series B 45(3), 703–711 (2002)
8. Kiguchi, K., Esaki, R., Tsuruta, T., Watanabe, K., Fukuda, T.: An Exoskeleton System for Elbow Joint Motion Rehabilitation. In: IEEE/ASME International Conference on Advanced Intelligent Mechatronics, Japan, pp. 1228–1233 (2003)
9. Hakogi, H., Ohaba, M., Kuramochi, N., Yano, H.: Torque Control of a Rehabilitation Teaching Robot Using Magnetro-Rheological Fluid Clutches. JSME International Journal Series B 48(3), 501–507 (2005)
10. Takaiwa, M., Noritsugu, T.: Development of Wrist Rehabilitation Equipment Using Pneumatic Parallel manipulator. In: IEEE International Conference on Robotics and Automation, Spain, pp. 2302–2307 (2005)
11. Koyanagi, K., Furusho, J., Ryu, U., Inoue, A.: Rehabilitation Sytem with 3-D Machine for Upper Limb. In: IEEE International Conference on Advanced Intelligent mechatronics, Japan, pp. 1222–1227 (2003)
12. Kushida, D., Nakamura, M.: Development of Meal Assistance Orthosis and Its Controller for Challenged Persons. IEEJ Trans. EIS 124(6), 1296–1304 (2004)
13. Tsumugiwa, T., Fuchikami, Y., Kamiyoshi, A., Yokogawa, R., Yoshida, K.: Stability Analysis for Impedance Control of Robot in Human-Robot Cooperative Task System. Journal of Advanced Mechanical Design, Systems, and Manufacturing 1(2), 133–121 (2007)
14. Richardson, R., Brown, M., Bhakta, B., Levesley, M.: Impedance control for a pneumatic robot-based around pole-placement, joint space controllers. ELSEVIER Control Engineering Practice 13, 291–303 (2004)

Spike: A Six Legged Cube Style Robot

Christopher Coyte, Mark Beckerleg, and John Collins

School of Engineering
AUT University
Auckland 1010, New Zealand
{chrcoy92,mark.beckerleg,john.collins}@aut.ac.nz

Abstract. This paper describes a six legged cube based robot named 'Spike', which uses three axes of symmetry with a collinear pair of legs mounted on each axis. Spike is designed to implement a novel form of locomotion which uses a tilting and falling motion as a result of leg movements to form its gait. Due to the triangular symmetry inherent in the robot's footprint, each step the robot takes is limited to one of three directions. At rest, the robot has three points of contact with the ground and all sides share equal length to form a tripedal stance. The gait is generated by directing a single leg toward the bisector of the opposing side, causing the robot to tumble, and fall into a new leg configuration. It was found that the robot was able to move along an angle of trajectory, where the accuracy of following the trajectory over time was dependant on the number of steps made.

Keywords: Six Legged Robotics, Tumbling motion, Motion control, Tripod Stance.

1 Introduction

A robot is a term that describes a large variety of devices that utilize many levels of autonomy, intelligence and motion. The main determining features of a robot are; they can sense and interact with their environment, have a degree of intelligence, can make deliberate movements and appear to have intent [1]. Robots are used in a wide variety of areas including; industrial, service, domestic, research applications and many other sectors.

Ground based robotic motion can be broadly split into legged and non-legged motion, each with advantages and disadvantages inherent in their design. A legged approach provides greater ability to negotiate obstacles because they are not limited by wheel height as with traditional wheeled mobile robotics. Legged robots can be less destructive to the environment [2], and have the ability to handle uneven natural terrain [3]. Wheeled approaches are better suited where speed is required on a smoother even surface [4], and have a simpler design complexity [5].

The applications for legged robotics are vast and much research has been done on a multitude of variations of legged robots. Commonly known legged robot applications include bipedal humanoid style robots such as HOAP-3 by Fujitsu and ASIMO by Honda and multi-legged designs such as The Ambler [6] and HAMLET [7]. There

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 535–544, 2009.

have also been robotic designs that are biologically inspired such as a Snake's motion [8] and four legged animals [3] like the BigDog by Boston Dynamics. There is also growing research development on mimicking biological evolution as a means of control systems for robotics [9].

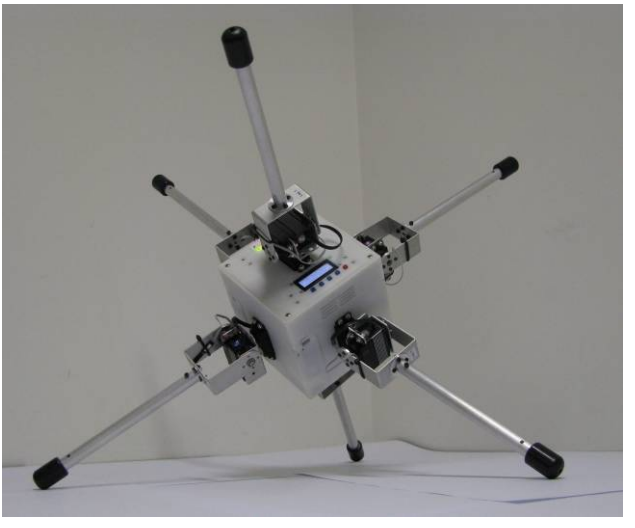


Fig. 1. The Spike Robot

2 Background

In this research, a legged robot named 'Spike' was built to evaluate a unique design and gait. The legged robot uses a total of six legs that are each fitted to a single face of its cube chassis. At rest, the robot forms a tripedal stance, with three legs facing the ground and three legs facing upward. Light sensors attached to each foot, allows the robot to determine which three of the six legs are in contact with the ground. The robot's motion is generated by moving a single leg towards the bisector of the opposing side. This action shifts the robot's centre of mass toward the fulcrum created by the two remaining standing legs. At the point where the centre of mass extends past the fulcrum point, the robot becomes unbalanced and tumbles forward into a new tripedal leg configuration. The physical robot prototype is shown in Fig 1.

Spike is programmed to follow a user defined trajectory by taking successive tumbling steps in the direction toward the desired heading. This form of motion is unique to the authors' knowledge.

The robot is controlled by a wireless remote controller, that is used to transmit an angle relative to the robot's starting position in its environment. The robot will travel along a given angle by moving the legs on its faces to force a weight imbalance by shifting its centre of mass, causing the tipping motion. The robot continues to shift towards the target autonomously until the user stops the robot. Relevant data is transmitted continuously from the robot for analysis on a PC.

It is desired that the robot will move along a line in any direction from its starting position, however the robot can not directly do this, as at any one time it can only move in one of three possible directions due to its triangular footprint. To overcome this, the robot's trajectory after each step is recalculated, to allow the robot to compensate for the previous directional errors. Thus, the accuracy of the overall motion improves with the number of steps taken.

It is planned that the robot will be available for further research on evaluating evolutionary computation based control systems for its motion, with an objective of finding new gaits that are able to move differently and perhaps more efficiently than the manually programmed gait used.

3 Design and Electronics

3.1 Leg Joint Construction

The geometric form of the chassis is in the shape of a cube, where each leg is at 90° to each cube face of the chassis. At rest, this creates three points of contact with the ground surface. Spike travels by forcing one of the three legs touching the ground to move towards the direction of travel, causing the cube to become unbalanced and fall forward. To enable the legs to move in any direction, two servo motors were needed to form a joint as shown in Fig 2. The first 'base' motor (α) rotates the second 'joint' motor (β) that actually tilts the leg, such that the leg is driven to the bisector of the opposing side, and hence the robot falls in line with the driving leg's direction.

The twelve digital servo motors used were titanium geared HSR-5995TG and HSR-5990TG manufactured by Hi-tec which provide 30kg/cm torque at stall and 39kg/cm standing torque, with a maximum current consumption of 5200mA at 7.4V per motor at stall. Both motors are able to provide 180° travel and are controlled via standard servo motor pulse width modulation. A detailed description of servo motors and their PWM timing requirements is described in [10]. The range of motion created by the two degrees of freedom allowed the foot to be positioned at any point around a half sphere created by each joint.

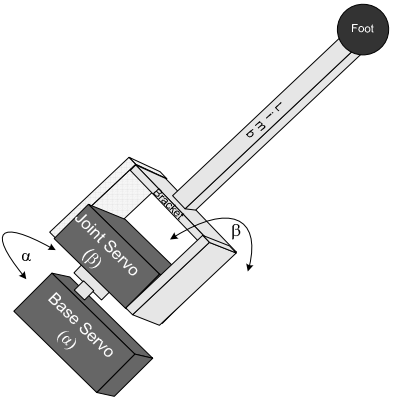


Fig. 2. Leg Mechanism

For Spike's gait, the base motor's rotation is limited such that the foot's direction of travel is always directly toward the bisector of the opposing side. This is achieved by the leg's base motor rotating through ±45° from its start position, and tilting the joint motor by a full ±90° from its starting position. This is illustrated in Fig 2. By adjusting the base motor by 45°, it allows the full drive capacity of the joint motor to act perpendicular to the ground plane. Each leg has two degrees of freedom, thus Spike has a total of twelve degrees of freedom available.

The limb and bracket are off-the-shelf components made from aluminium which is a benefit for weight and strength characteristics in this application. The foot is turned from acetyl on a lathe, and rounded at the ground contact points for improved sliding ability against the frictional surface.

3.2 Chassis Design

The cube was fabricated using a 3D plastics rapid prototyping printer. This allows the cheaply and precisely manufactured cube shape to include complex internal structures for mounting servo motors, batteries and circuit boards. With the design developed in plastics, the chassis had sufficient strength to be suitable for the impact experienced with each leg movement and the robot's tumbling motion on the ground. As the chassis was custom designed, all components and mounting points were perfectly suited for the robot design. As plastic was used, the cube was light enough to prevent unnecessary stress on the servo motors, which would affect battery life and mechanical wear.

The cube measures 110mm internally and 118mm externally and is comprised of two parts; a lid, and the remaining body. The internal circuit boards are stacked together and mounted to the lid panel, allowing all the electronics to be removed from the cube simultaneously. This provides easy access to components and simplifies maintenance. Four cavities inside the cube were provided for battery placement. Each cavity included a catch mechanism, similar to that found on a television remote control, allowing the battery to be easily removed. Ventilation grills providing airflow were included for heat dissipation. The plastic used to create the cube was transparent to radio frequencies, allowing wireless devices to be included inside the cube, creating a finished and professional look.

3.3 Electronics

Custom circuit boards were designed for system control, interfacing to external I/O devices and wireless communication. The robot contained nine discrete circuit boards, of which the two primary boards are: the control board, and the I/O interface board. The control board is used to implement the gait and navigation. It also reads the light sensors, battery and temperature probe states that are measured by the I/O interface board. The control board is detachable and has two versions; an Atmel microcontroller and an Altera FPGA allowing the control system to be interchanged to make use of microcontroller, FPGA and other processing technologies. The I/O interface board provides PWM control to servo motors and power control for the entire robot. The inter-processor communication between the control and I/O board is achieved using the Serial Peripheral Interface (SPI) communication bus. A block diagram of the interconnections is shown in Fig 3.

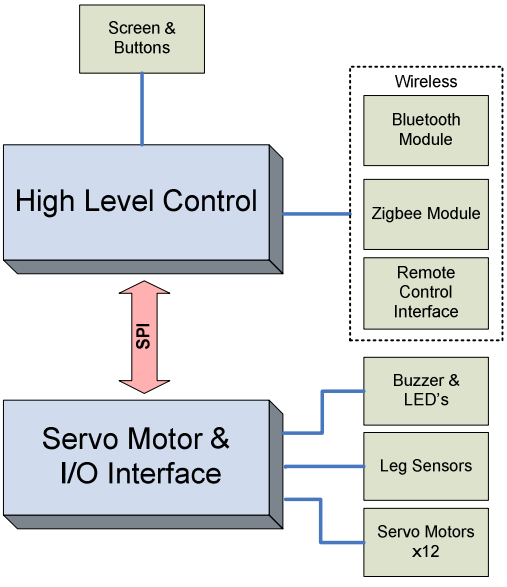


Fig. 3. Communication and Electrical Overview

Control software for Spike is run on two Atmel Mega128 microcontrollers, running at 12MHz and 8 MHz for the I/O interface and control boards respectively. The control board processor receives its commands from a wireless controller and a button interface. The button interface selects the operational mode, whereas the wireless controller provides a user specified bearing for the robot to follow. The control processor performs the algorithms to determine how the robot gait will achieve the desired bearing and the corresponding sequence of servo motor positions that will allow the robot to follow the bearing. These servo motor positions are then sent to the I/O interface board. A Bluetooth link is also incorporated for PC data analysis of calculations and status readings.

The I/O interface board receives details of desired motor angle position, which is then converted into the required PWM for the servo motors. The I/O interface board is also responsible for reading analogue measurements and for control of the power systems for the robot.

Light sensors are mounted inside each leg's foot. The relative differences in light levels from each sensor allowed the robot to determine which three of the six feet are in contact with the ground. The robot legs are grouped into three axes with an opposing pair of legs on each. The relative difference in light levels is measured between two legs on the same axis and the sensor with the lower light reading is assumed to be in contact with the ground. This reading is necessary to assist with the trajectory planning of the robot in its environment.

The robot is autonomous, thus power is supplied by four lithium polymer high capacity 7.4V 1250mAh batteries that have a maximum discharge rate of 25 amps. To avoid power fluctuations on the electronics, one battery is used to power the electronics whilst the remaining 3 batteries are evenly divided between the servo motors.

Servo motor current consumption fluxuates during motion, largely due to high demands placed on the servo motors, as forces on the legs vary. Generally at rest with 3 legs on the ground, current consumption is approximately 1.6A per motor giving approximately 20 minutes of running time for the robot. The robot also has an external DC source connection for non-autonomous applications.

3.4 Mechanical Forces

The robot is inherently stable at rest as its centre of mass is balanced in equilibrium by its tripod base formed using the 3 axes of symmetry. The forces acting on the robot legs are shown in Fig 4. The robot weighs approximately 2.2kg and its height varies during the tumbling motion. The maximum breadth between two feet is 650mm.

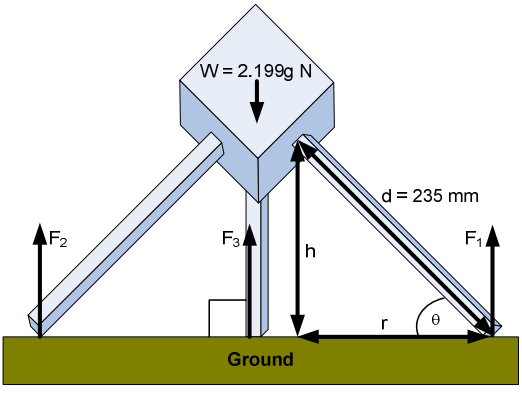


Fig. 4. Forces on Legs

To find the maximum opposing force acting on the servo motor to determine the ability of the servo to shift the robot, we initially calculate the torque, shown in (1)-(7):

$$\tau = Fd \tag{1}$$

In the test situation the friction coefficient of the ground is taken as approximately k≈0.7. Thus we have:

$$\tau = \frac{W}{3} d\left(\cos\theta + k\sin\theta\right) \tag{2}$$

To calculate θ when τ is maximum, we differentiate:

$$\frac{d\tau}{d\theta} = 0 \tag{3}$$

$$-\sin\theta + k\cos\theta = 0 \tag{4}$$

$$\tan\theta = \frac{\sin\theta}{\cos\theta} = k \tag{5}$$

Therefore we have θ = 35° when opposing torque is a maximum with k=0.7. The maximum opposing torque on the servo motor is:

$$\tau = \frac{2.199}{3} \times 0.235 \times (\cos 35 + 0.7 \times \sin 35) \tag{6}$$

$$\tau = 0.21N \tag{7}$$

The servo motors used in this application are capable of delivering 0.3Nm torque at stall, and 0.39 Nm standing torque, thus the motors are able to shift the robot in a worst case angle scenario.

3.5 Mechanics of Motion

In Fig 5, the gait strategy for Spike is shown. Before commencing movement, it is assumed that the centre of mass is balanced evenly in the centre of the cube. To illustrate Spike's motion, consider the following example. If the target heading is due north of the robot, the legs need to arrange themselves in a way to allow the robot to fall in the direction nearest due north. In Fig 5, this would be performed by driving leg 1 directly toward the ground, in line with the bisector of the line drawn between legs 2 and 3. As the driving leg progresses through its motion, legs 2 and 3 act as a fulcrum turning point as the side of leg 1 rises up and toward the tipping point directly over the two other standing legs. When the robot's mass shifts past these two standing legs, the robot falls onto leg 4, and is hence in a new leg configuration and orientation. The leg positions in a step sequence are shown pictorially in Fig 6.

3.6 Dynamic Recalibration of the Trajectory Heading

Spike will attempt to move along a heading provided by the user via the remote control. The heading is a bearing relative to the robot's starting position which is taken to be 0°. Spike has a triangular footprint and thus can move in one of only three possible directions each 120° apart. In order to move along a bearing lying between these angles, a combination of steps must be taken. In order to determine which leg to move, a mathematical analysis is repeatedly calculated to produce a direction in which the robot should travel. It is likely that the required direction cannot be exactly followed, thus the closest direction in which the robot can move is determined and the robot then steps in that direction. The desired direction must be recalculated after every step, to compensate for the accumulation of directional errors.

This method found that the accuracy of following the starting heading was dependant on the number of steps made by the robot. A sample plot is shown in Fig 7 of the robot shifting towards a 55° heading. The bulleted points on the plot are the actual positions of the robot throughout its journey. The linear line is the desired trajectory

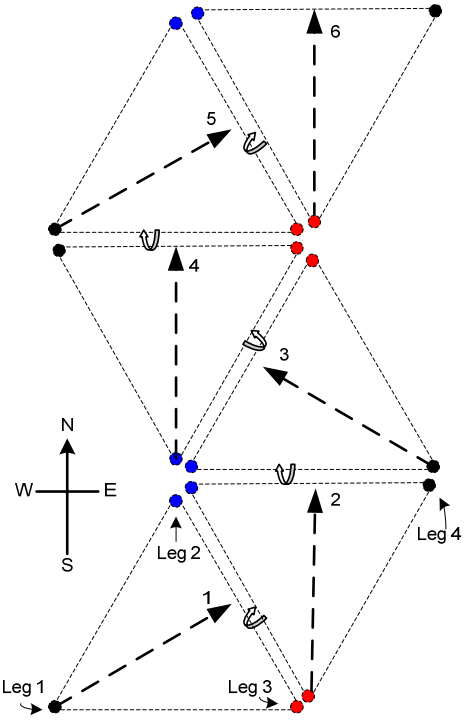


Fig 5. Gait Strategy

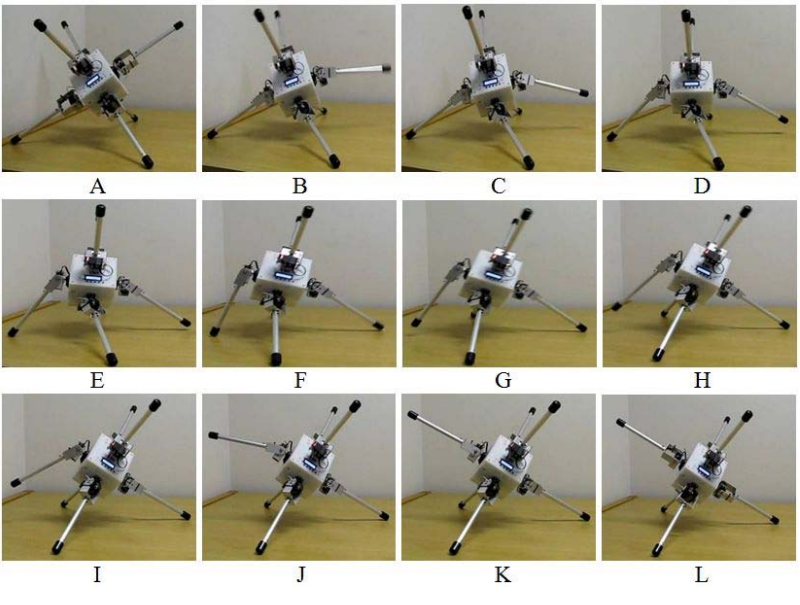


Fig. 6. Gait Operation

for the robot to follow. It can be seen that as the robot begins to drift away from the target trajectory, the robot recalculates its position to make correction steps, such that the robot returns to the required heading.

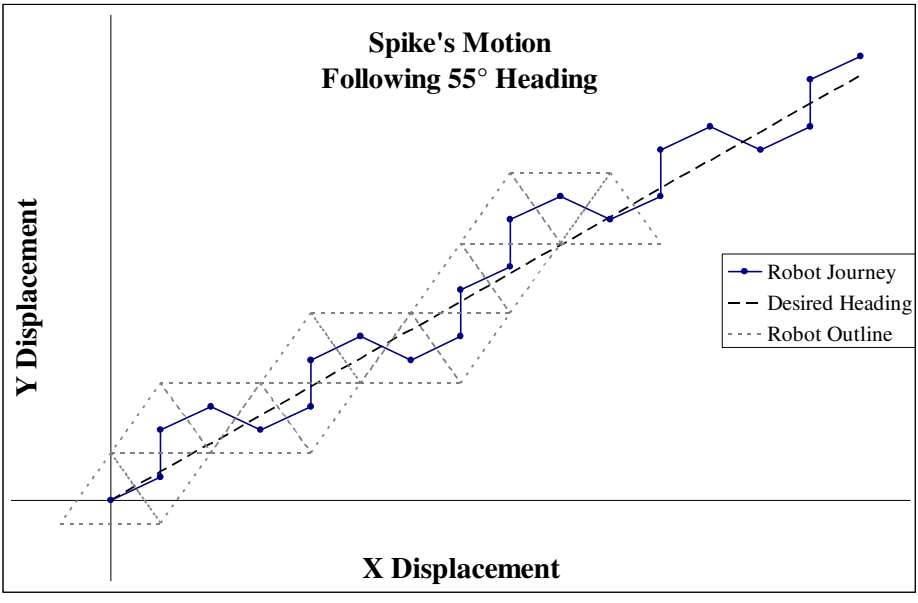


Fig. 7. Direction Following

4 Conclusions

In this paper, it has been shown that a six legged robot using a tripedal gait and tumbling motion can successfully produce a motion along a user input heading. Emphasis has been placed on describing the leg mechanics as a main part of the robot's gait design. The symmetry of the triangular ground footprint is discussed with diagrams that explain the axes and tipping points of the robot. The experiments carried out exhibited the motion of the robot in its environment, and a frame by frame account of the gait was shown.

References

1. Braunl, T.: Embedded Robotics: Mobile Robot Design and Applications with Embedded Systems (2006)
2. Todd, D.J.: Walking machines: an introduction to legged robots/D.J. Todd. Kogan Page (1985)
3. Railbert, M., Blankespoor, K., Nelson, G., Playter, R.: BigDog, the Rough-Terrain Quadruped Robot. In: Proceedings of the 17th World Congress, The International Federation of Automatic Control, pp. 10822–10825 (2008)

4. Pablo, G., Iez de, S., Elena, G., Joaquin, E.: Quadrupedal Locomotion: An Introduction to the Control of Four-legged Robots. Springer, New York (2006)
5. Kuo, A.D.: Choosing Your Steps Carefully. IEEE Robotics & Automation Magazine 14, 18–29 (2007)
6. Bares, J.E., Whittaker, W.L.: Walking robot with a circulating gait. In: IEEE International Workshop on Intelligent Robots and Systems 1990. Towards a New Frontier of Applications, Proceedings. IROS 1990, vol. 802, pp. 809–816 (1990)
7. Fielding, M.R., Dunlop, R., Damaren, C.J.: Hamlet: force/position controlled hexapod walker - design and systems. In: Proceedings of the, IEEE International Conference on Control Applications (CCA 2001), pp. 984–989 (2001)
8. Junyao, G., Xueshan, G., Wei, Z., Jianguo, Z., Boyu, W.: Design and research of a new structure rescue snake robot with all body drive system. In: IEEE International Conference on Mechatronics and Automation, 2008. ICMA 2008, pp. 119–124 (2008)
9. Tan, K.C., Chew, C.M., Tan, K.K., Wang, L.F., Chen, Y.J.: Autonomous robot navigation via intrinsic evolution. In: Proceedings of the 2002 Congress on Evolutionary Computation, vol. 2, pp. 1272–1277 (2002)
10. Pinckney, N.: Pulse-width modulation for microcontroller servo control. IEEE Potentials 25, 27–29 (2006)

The Similarity Design of Heavy Forging Robot Grippers

Qunming Li, Yonghong Wu, and Hua Deng

School of Mechanical and Electrical Engineering,
Central South University, Changsha, 410083, China
liqm@mail.csu.edu.cn

Abstract. In order to design and study the gripping mechanism of heavy forging robots and to save money and delivery-time, it is necessary to optimize the structural dimensions of the gripper, and design a downsizing model gripper based on similarity theory, which can be had experiments to test the physical features and performances of the prototype of heavy grippers. This paper has studied the methods of similarity design and dimension optimization of the gripper mechanisms of heavy forging robots. Firstly, by using similarity theory, the similar criterion and similar indicators of the gripping system are presented, and the relationship between the geometry similarity ratio and other parameters' similarity ratio is defined. Secondly, based on the design requirement of the gripping design, the non-dimensional objective function is established. By using multi-objective optimization genetic algorithm, the optimization of the push-bar gripper is carried out. Finally, based on the simulation of a model and a prototype and the relationship of parameters of the gripping systems, the similarity of the experimental models is analyzed, demonstrating the feasibility of the similarity design for heavy forging robot grippers.

Keywords: Heavy forging robots, similarity design, multi-objective optimization, genetic algorithm.

1 Introduction

Heavy duty forging robots play an important role in forging automation. Now the required amount of large-scale heavy duty forging robots is increasing greatly year by year, and the gripping capacity of these robots is increasing to more than 100 tons. It is well known that it is almost impossible to develop a real heavy forging robot to have modal test and performance analysis because of the shortage of cost, field, material, and delivery time. The method of simulating tests or the model measurements with the aid of computer, which closely combines theory with experiment, is perhaps the breakthrough in the future [1]. So a downsizing model of the heavy forging robot is generally accepted based on similarity theory to study and test its performances and features before the real forging robot is developed and experimented.

The main difference of heavy forging robots from usual industrial robots is that the gripped gravity and torque are very large. It can save huge costs and lots of delivery

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 545–553, 2009.

time by similarity design, optimization and model experiments. The gripper is one of the key parts of the forging robots. It is the interface between the forged work piece and the whole forging manipulator. Most of the heavy gripping systems that are installed in industrial automations and robots are designed to be mechanical two symmetrical-finger grippers, which are considered as the simplest efficient grasping configuration. The optimization of gripping mechanisms and the method of similarity design are the main work for model design and experiments.

A lot of methods from robotic researches have been proposed to optimize multifingered grasps [2-4]. These methods mainly focus on the grasping force distribution and optimization, and don't consider the fingers' geometrical dimensions. All of them are supposed that the contact model between fingers and the grasped work-piece is a friction point contact, and the contact forces of the finger-tips can be controlled. The design of the two-finger grippers must consider multi-objective parameters simultaneously. According to the optimum objective functions, optimum gripper dimensions have been found out by genetic algorithm [5], [6]. These researches are not concerned to similarity design. The method of similarity design between the physical prototype and its model system is studied in [7], but it is not a mechanical similarity. This paper is focused on the method of similarity design and dimension optimum of the forging grippers.

2 Conditions of Similarity Design

Geometrical similarity is a fundamental requirement between the model and the prototype of forging grippers. Geometrical similarity includes similar configuration of the gripping systems and similar structural dimensions. For a stable gripping manipulation during the forging process, the gripping system must be ensured that the high temperature forged work-piece will not slip relative to the tongs, whenever the actions are to rotate, lift up and down, sway the gripper, or forge and impact the work-piece. So the other two similarity requirements are kinematics similarity and dynamics similarity. The similarity rules are mainly based on the dynamics similarity of a mechanical system.

To grip the object stably, the gripping system of the forging robot must exert enough contact forces to the object through the tongs, and the contact status has great influences on the gripping stability [8]. And contact problems in robot dynamics represent a very complex field of research. From the standpoint of mechanics, the dynamic model of the gripping system of the forging robot has to consider the issues of multi-body dynamics, elasticity, friction and contact. So the dynamic model of the gripper can be represented as follows:

$$F(f,\theta,\omega,\dot{\omega},l,v,a,m,\rho,E,\varepsilon,\sigma,\mu,g,t)=0 \tag{1}$$

where: f— forces
θ— rotation angle of the gripper
ω— rotation angle velocity of the gripper
$\dot{\omega}$— rotation angle acceleration of the gripper
l— joints' length of the gripper

v— motion velocity of the gripper
a— motion acceleration of the gripper
m— masses of the system
ρ — density of the material
E— elastic modulus
ε — elastic strain
σ — elastic stress
μ — friction coefficient
g — gravity acceleration
t — time

There are 3 fundamental physical variables: forces f, length l, time t among the 15 quantities. According to the second theorem of similarity, 12 dimensionless variables (criterions) are defined as follows:

$$\begin{cases} \Pi_1=\theta,\Pi_2=\mu,\Pi_3=\dfrac{E}{F/l^2}, \\ \Pi_4=\dfrac{g}{v^2/l},\Pi_5=\dfrac{t}{l/v},\Pi_6=\dfrac{m}{Fl/v^2}, \\ \Pi_7=\dfrac{\rho}{F/l^2v^2} \\ \Pi_8=\dfrac{\omega}{v/l},\Pi_9=\dfrac{\dot{\omega}}{v^2/l^2},\Pi_{10}=\dfrac{a}{v^2/l}, \\ \Pi_{11}=\varepsilon,\Pi_{12}=\dfrac{\sigma}{F/l^2} \end{cases} \tag{2}$$

In the dynamic model of the gripper, θ, l, v, m, E, μ, g, t, ρ, ω, $\dot{\omega}$ are given or measured, which are known variables, and a, ε, σ are unknown variables which need to be solved. The π-criterions are called qualified criterions when they are consisted of known and fundamental variables, or are called quantified criterions when they are consisted of unknown and fundamental variables. If subscript 1 represents prototype of the gripper, and subscript 2 represents model, then according to the first theorem of similarity, nine similarity conditions can be obtained as follows:

$$\begin{aligned} &\theta_1=\theta_2,\mu_1=\mu_2,\frac{E_1}{f_1/l_1^2}=\frac{E_2}{f_2/l_2^2}, \\ &\frac{g_1}{v_1^2/l_1}=\frac{g_2}{v_2^2/l_2},\frac{\rho_1}{f_1/(l_1^2v_1^2)}=\frac{\rho_2}{f_2/(l_2^2v_2^2)}, \\ &\frac{t_1}{l_1/v_1}=\frac{t_2}{l_2/v_2},\frac{m_1}{f_1l_1/v_1^2}=\frac{m_2}{f_2l_2/v_2^2}, \\ &\frac{\omega_1}{v_1/l_1}=\frac{\omega_2}{v_2/l_2},\frac{\omega_1}{v_1^2/l_1^2}=\frac{\omega_2}{v_2^2/l_2^2}, \end{aligned} \tag{3}$$

And four output solution equations are obtained as follows:

$$\begin{aligned}
&\varepsilon_1 = \varepsilon_2, \\
&a_1 = \left(\frac{a_2}{v_2^2 / l_2}\right)\left(\frac{v_1^2}{l_1}\right), \\
&\sigma_1 = \left(\frac{\sigma_2}{f_2 / l_2^2}\right)\left(\frac{f_1}{l_1^2}\right), \\
&f_1 = \left(\frac{l_1}{l_2}\right)^2 f_2,
\end{aligned} \tag{4}$$

3 Selections of Similarity Scales

Define the similarity variables of the design parameters as follows:

$$\lambda_\theta = \frac{\theta_1}{\theta_2},\ \lambda_F = \frac{f_1}{f_2},\ \lambda_\mu = \frac{\mu_1}{\mu_2}, \lambda_\varepsilon = \frac{\varepsilon_1}{\varepsilon_2},$$

$$\lambda_l = \frac{l_1}{l_2},\ \lambda_\omega = \frac{\omega_1}{\omega_2},\ \lambda_{\dot\omega} = \frac{\dot\omega_1}{\dot\omega_2},\ \lambda_\rho = \frac{\rho_1}{\rho_2},$$

$$\lambda_v = \frac{v_1}{v_2},\ \lambda_a = \frac{a_1}{a_2},\ \lambda_\sigma = \frac{\sigma_1}{\sigma_2},\ \lambda_E = \frac{E_1}{E_2},$$

$$\lambda_g = \frac{g_1}{g_2},\ \lambda_t = \frac{t_1}{t_2},\ \lambda_m = \frac{m_1}{m_2}$$

Substitute the above relationships into (3) and (4), and then the similarity indicators can be obtained:

$$\begin{aligned}
&\lambda_\theta = 1,\ \lambda_\mu = 1,\ \lambda_\varepsilon = 1, \\
&\lambda_v^2 = \lambda_g \lambda_l,\ \lambda_l = \lambda_v \lambda_t,\ \lambda_l \lambda_F = \lambda_m \lambda_v^2, \\
&\lambda_\rho \lambda_v^2 \lambda_l = \lambda_F,\ \lambda_v = \lambda_\omega \lambda_l,\ \lambda_v^2 = \lambda_{\dot\omega} \lambda_l^2, \\
&\lambda_v^2 = \lambda_a \lambda_l,\ \lambda_F = \lambda_\sigma \lambda_l^2,\ \lambda_F = \lambda_E \lambda_l^2
\end{aligned} \tag{5}$$

Table 1. Similarity Scales of the design variables of the gripper

Variables	Similarity scale	Variables	Similarity scale
l	λ_l	m	λ_l^2
θ	1	a	1
ω	$1/\sqrt{\lambda_l}$	E	1
$\dot\omega$	$1/\lambda_l$	μ	1
f	λ_l^2	σ	1
t	$\sqrt{\lambda_l}$	ε	1
v	$\sqrt{\lambda_l}$	g	1

For the similarity design of the gripper of a forging robot, the model and the prototype are established in the same gravity field, and is designed in the same configuration with the same materials for similar parts, so we have: $\lambda_g = 1$, $\lambda_a = 1$, $\lambda_\mu = 1$, and $\lambda_E = 1$. According to the similar indicators of equation (5), the similarity scales of various variables relative to the dimensional scale λ_l can be obtained as shown in table 1. When the similarity scale λ_l is given, all other design parameters can be derived from table 1.

4 Optimization of the Dimensions of a Forging Gripper

The mechanisms of a robot gripper are classified as two types: pull with a sliding block and push with a long lever. For a similarity design, it is important to optimize the gripping mechanism to obtain the best dimensions and performances from model to prototype. Fig.1 is an example of a typical configuration of forging grippers which is driven by a push lever connected to a hydraulic cylinder. When the lever is pushed, the two tongs will close to grip object, if pulled, it will open.

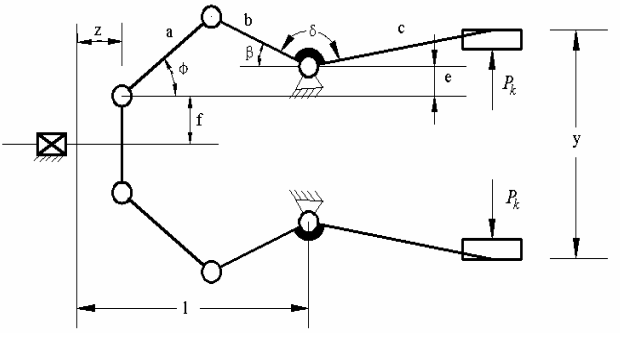


Fig. 1. The configuration of a typical heavy forging gripper

4.1 Design Parameters and Constraint Equations

For this gripper, the design variables are depended on a,b,c,e,f,l,δ. Where a, b, c, e, f, l are dimensions of the gripper and δ is the angle between elements b and c of the gripper. So the decision vector is $\mathbf{X}$=(a,$b,c,e,f,l,\delta)^T$. Suppose that F is the actuating force of the cylinder, ϕ and β are angles of links with horizontal reference lines. P_k is the griping force exerted by fingers of robot gripper on the work piece. The geometrical dependences of the gripper mechanism are shown in Fig.2.

According to force balance equations and the geometrical relationship, the following equations are obtained.

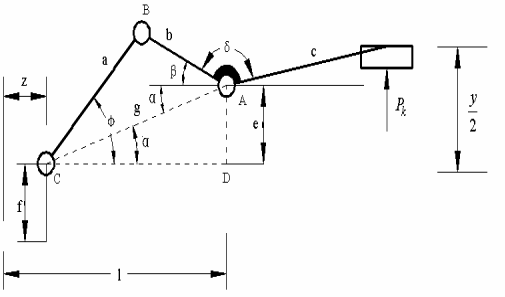


Fig. 2. Geometrical dependencies of the gripper

$$\frac{F}{2} = R\cos(\phi), \quad R = \frac{F}{2\cos(\phi)} \tag{6}$$

$$Rb\sin(\beta+\phi) = P_k c\cos(\pi-\beta-\delta) \tag{7}$$

$$P_k = \frac{Rb\sin(\beta+\phi)}{c\cos(\pi-\beta-\delta)} = \frac{bF\sin(\beta+\phi)}{2c\cos(\phi)\cos(\pi-\beta-\delta)} \tag{8}$$

$$g^2 = e^2 + (l-z)^2, \quad g = \sqrt{e^2+(l-z)^2} \tag{9}$$

$$tg(\alpha) = \frac{e}{l-z}, \quad \alpha = artg(\frac{e}{l-z}) \tag{10}$$

4.2 Optimum Results Using NSGA-II

The definition and selection of objective functions must satisfy the forging process requirements of the forging robots. Based on equations (6)-(10), considering the geometrical compact and gripping efficiency, the objective functions can be evaluated as the same as reference [5]. The Non-dominated Sorting Genetic Algorithm (NSGA-II) is used to solve the multi-objective optimum problem.

A model design example of NSGA-II is optimized using the following data:

1. Geometrical constraints (length unit: mm):
$170 \le a \le 500$, $120 \le b \le 550$,
$200 \le c \le 600$, $100 \le e \le 450$,
$50 \le f \le 500$, $200 \le l \le 700$, $90^0 \le \delta \le 180^0$
2. The parameters of NSGA-II:
* Population size=200, number of generations=150
* Crossover rate=0.6, mutation rate=0.001

The multiple objectives are combined into scalar objective via weight vector. Table 2 gives the model optimum results from the Pareto fronts and according to the similarity design theory introduced above, from table 1, when the model parameters are given ($\lambda_l = 1/2$), a prototype parameters can also be derived. In table 2, number 1 represents model, and 2 represents prototype.

Table 2. An optimum design parameters by NSGA-II

Model	a(mm)	b(mm)	c(mm)	e(mm)	f(mm)	l(mm)	i
1	224.8	190.7	391.1	110.1	70.36	339.6	2.472
2	449.2	381.4	782.2	220.2	140.7	679	2.472

5 Simulation Analysis of Similarity Design

In order to analyze the similarity between the model and prototype based on the theory of similarity design, we have designed two different solid model of the forging gripper according to the fundamental optimum geometrical dimensions and the length similarity scale λ_l, and simulate the similarities of their dynamic and kinematics characteristics by using multi-body dynamics software ADAMS. Define subscript 1 represent model 1, and subscript 2 represent model 2, which 1 is an optimum design model based NSGA-II, and 2 is a prototype of 1 designed by similarity theory as described in this paper. Their dimensional scale is given as $\lambda_l = l_1 / l_2 = 1/2$. The dimension sizes is given in table 2. Then the other parameter can be obtained from table 1 as follows:

$$\frac{m_1}{m_2} = (\frac{l_1}{l_2})^2 \quad \frac{v_1}{v_2} = (\frac{l_1}{l_2})^{\frac{1}{2}},$$

$$\frac{\omega_1}{\omega_2} = (\frac{l_2}{l_1})^{\frac{1}{2}} \quad \frac{F_1}{F_2} = (\frac{l_1}{l_2})^2,$$

$$\frac{t_1}{t_2} = (\frac{l_1}{l_2})^{\frac{1}{2}}, \quad \sigma_1 = \sigma_2$$

The 3-dimension solid configuration of model 1 and 2 is shown in fig.3.

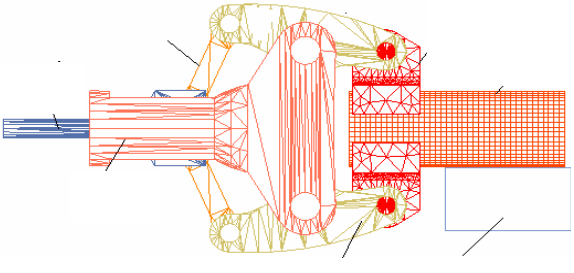


Fig. 3. Simulation model of the forging grippers

During the dynamic simulations, now let gripper model 1 grip an object and lift up in velocity v_1, and let gripper model 2 lift up in a similarity scale velocity $v_2 = v_1(1/\lambda_l)^{\frac{1}{2}}$. And the time size is $t_1 = t_2(\lambda_l)^{\frac{1}{2}}$. If the simulation results of model 1 and 2 read from ADAMS dynamic calculation at the similarity time satisfy the similarity indicators as described in table 1, it demonstrates that the similarity criterions and

scales is reasonable and effective, which supports its application in model optimum design and prototype experiments.

Define that the gripping force of the up tong of model 1 is F_1, and that of model 2 is F_2, and now output their strains at the similar position of the two tongs, the results calculated by ADAMS during the lifting operation is shown in fig.4 and fig.5.

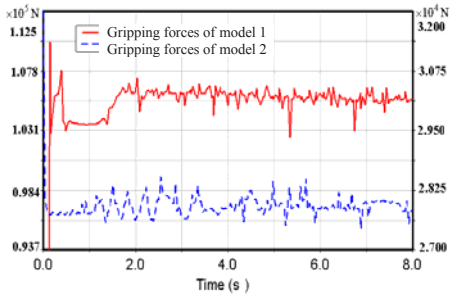


Fig. 4. Gripping forces of model 1 and 2

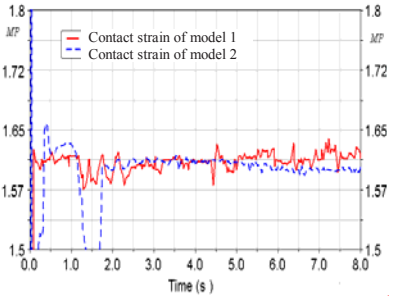


Fig. 5. Contact strains at a central point of model 1 and 2

It can be seen from the simulation results that the change trends of the gripping forces and contact strain forces are consistent, and the ratio of model 1 and 2 at the similar time satisfies the similarity scale. It also shows that the similarity design method for heavy forging grippers presented in this paper is effective, and can be applied for downsizing model's physical-similarity design and experiment of various grippers, which can save money, field and time than prototype design and experiments.

6 Conclusions

This paper has studied the methods of similarity design and dimension optimization of the gripper mechanisms of heavy forging robots. Based on similarity theory, the similar criterion and similar indicators of the gripping system are presented and the relationship between the geometry similarity ratio and other parameters similarity ratio is defined. Then, according to the design requirements of the gripping system, the non-dimensional objective function is established. By using multi-objective optimization genetic algorithm, optimization of the push-bar gripper is carried out. Simulation results demonstrate the feasibility of the similarity design for heavy duty forging robot grippers.

Acknowledgments. This work was supported by National Basic Research Program of China (Grant No.2006CB705400). Authors acknowledge Dr. Yi Chen for providing NSGA-II code on Web used in this study.

References

1. Shi, Z., Ding, C., Chen, J.: Dimensional Analysis and Physical Similarity of Lossy Electromagnetic Systems. Chinese Physical Letters 10(6), 347–350 (1994)
2. Wantanabe, T., Yoshikawa, T.: Grasping optimization using a required external force set. IEEE Trans. Robotics and Automation 44(1), 52–66 (2007)
3. Buss, M., Hashimoto, H., Moore, J.B.: Dextrous hand grasping force optimization. IEEE Trans. Robotics and Automation 12(3), 406–417 (1996)
4. Abu Zitar, R.A., Nuseirat, A.M.: Optimum gripper using ant colony intelligence. Industrial Robot: An International Journal 32(1), 17–23 (2005)
5. Saravanan, R., Ramabalan, S., Godwin Raja Ebenezer, N., et al.: Evolutionary multi criteria design optimization of robot grippers. Applied Soft Computing 9, 159–172 (2009)
6. Filomeno Coelho, R., Bersini, H., Bouillard, Ph.: Parametrical mechanical design with constraints and preferences: application to a purge value. Computer Methods in Applied Mechanics and Engineering 192, 4355–4378 (2003)
7. Chen, J.-H.: Theoretical Analysis of Multi-Objective Genetic Algorithms— Convergence Time. In: Population Sizing and Disequilibrium, IEEE NNS Walter Karplus Research Grant, (2003)
8. Li, Q., Gao, D., Deng, H.: Influence of Contact Forces on Stable Gripping for Large-Scale heavy Manipulator. In: Xiong, C.-H., Liu, H., Huang, Y., Xiong, Y.L. (eds.) ICIRA 2008. LNCS (LNAI), vol. 5315, pp. 839–847. Springer, Heidelberg (2008)

The Research of Mechanism Synthesis Based on Mechanical System Chaos Anti-control Methods

Youxin Luo

College of Mechanical Engineering, Hunan University of Arts and Science Changde, 415000, P.R. China
LLYX123@126.com

Abstract. Many questions in natural science and engineering are usually transformed into solving nonlinear equation. Newton iterative method is an important technique to one dimensional and multidimensional variables and iterative process exhibits sensitive dependence on initial guess points. For first time, the nonlinear feedback chaos anti-control method of mechanism synthesis was put forward. The motion of mechanical centrifugal governor systems was transformed from periodic motion to chaotic motions by the addition of nonlinear feedback and all solutions of the nonlinear questions of mechanism synthesis were found by utilizing chaotic sequences of chaotic motions of mechanical system to obtain locate initial points. As an example the problem of function generation for planar four-linkage guide mechanism was considered. This makes that multi-projects selecting could be possible. This method is adaptive to planar multi-linkage and spatial mechanism. This provides a new simple realization method for mechanics design.

Keywords: Mechanical systems, chaos anti-control, Matlab software, Planar-linkage, nonlinear equations.

1 Introduction

The question of motion synthesis in mechanism has the characteristic of multi-solutions. How to find the dimensional scheme which satisfies the given kinematic mechanism and get the optimization is the meaningful research.

The mathematic model of mechanism synthesis is a vast nonlinear polynomial system. How to find the solutions of the system is a difficult problem in the mechanism fields. Chaotic motion in the system comes from the characteristic of nonlinear system. But nonlinear is not the sufficient condition but the necessary condition. How to use the chaotic iterative method to solve the nonlinear equations in mechanism, literature [1-7] had done much work, which provided a new way to solve the mechanism problems. Applying the chaotic characteristic of mechanical system into the mechanism synthesis is an important research problem. The objective is to make use of the chaotic characteristic of mechanical system to fleetly solve the problem of mechanism synthesis. Chaos control can suppress or eliminate the chaotic dynamical behavior.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 554–561, 2009.

Chaotic anti-control mainly is that through the external input or the adjustment in internal parameters the original non-chaotic system becomes chaos or chaos of the original system becomes stronger [8]. In this paper, a novel method to find all the solutions of mechanism synthesis question was introduced, in which the initial point of Newton iterative method were the chaotic sequences generated by chaotic anti-control of mechanical system. As an example, by setting the guidance synthesis of the planar four-linkage mechanism**,** this paper found multi-project which satisfies the given motion condition and it provides the basis for project optimization.

2 Anti-control of Chaos for Mechanical Centrifugal Flywheel Governor Systems

Kinematic equations of dimensionless 3D non-autonomous system for centrifugal velometer are shown as follows [8].

$$\begin{cases} \dot{x} = y \\ \dot{y} = (e + n^2 z^2)\sin x \cos x - (e + \dfrac{9.8}{l})\sin x - by \\ \dot{z} = \dfrac{\alpha \cos x - F}{I} \end{cases} \tag{1}$$

In the autonomous system when system parameter $n = 3, l = 1.5, b = 0.4$, $F = 0.3, I = 1.2, e = 0.3, \alpha = 0.46$, the system is in the periodic motion. System diagram is as Fig.1. Using nonlinear control in formula (1), the anti-control equations are as follows.

$$\begin{cases} \dot{x} = y \\ \dot{y} = (e + n^2 z^2)\sin x \cos x - (e + \dfrac{9.8}{l})\sin x - by \\ \dot{z} = \dfrac{\alpha \cos x - F}{I} + \varepsilon z|z| \end{cases} \tag{2}$$

When $\varepsilon = -0.17$, system is transformed from periodic motion into chaotic motion. The diagram is as Fig. 2.

If periodic load is given by velometer, the kinematic equations of dimensionless 3D non-autonomous system for centrifugal velometer are as follows [8].

$$\begin{cases} \dot{x} = y \\ \dot{y} = (e + n^2 z^2)\sin x \cos x - (e + \dfrac{9.8}{l})\sin x - by \\ \dot{z} = \dfrac{\alpha \cos x - F}{I} - a\sin(\omega t) \end{cases} \tag{3}$$

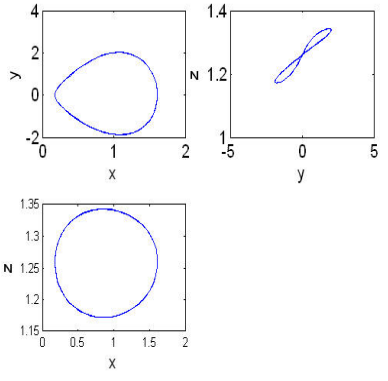


Fig. 1. Autonomous system diagram ($\alpha = 0.46$)

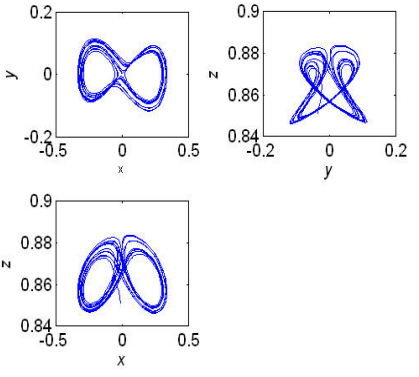


Fig. 2. Autonomous system diagram after control ($\varepsilon = -0.17$)

In formula (3) of non-autonomous system, when the system parameters are given by $n = 3, l = 1.5, b = 0.4, F = 0.3, I = 1.2$, $e = 0.3, \alpha = 0.49, a = 0.8, \omega = 1$, the system is in periodic motion, the diagram is shown as Fig.3. Nonlinear control is loaded in formula (3), the anti-control equations are as follows.

$$\begin{cases} \dot{x} = y + \varepsilon x|x| \\ \dot{y} = (e + n^2 z^2)\sin x \cos x - (e + \dfrac{9.8}{l})\sin x - by \\ \dot{z} = \dfrac{\alpha \cos x - F}{I} - a\sin(\omega t) \end{cases} \tag{4}$$

When $\varepsilon = 0.03$, the system is transformed from periodic motion into chaotic motion, the diagram is as Fig. 4.

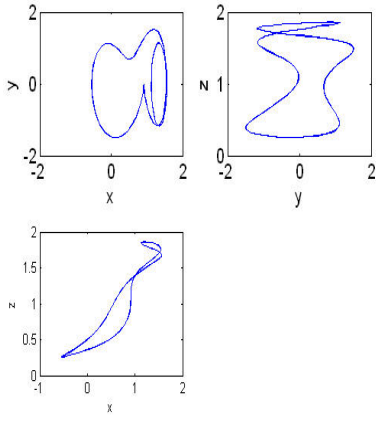


Fig. 3. Autonomous system diagram ($\alpha = 0.49$)

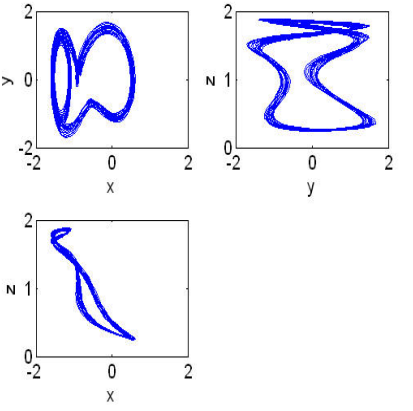


Fig. 4. Autonomous system diagram after control ($\varepsilon = 0.03$)

3 Newton Iterative Method of Chaotic Anti-control Based on Mechanical System

Based on mechanical system chaotic anti-control the steps of using Newton iterative method to find all solutions of mechanism are as follows:

Step 1: Establish the iterative formula N(x) of Newton optimal iterative, mechanism synthesis can be transformed into multi-equations.

$$F(x) = 0 \tag{5}$$

Where,

$$F(x) = [f_1(x), f_2(x), \cdots, f_n(x)]^{\mathbf{T}}$$
$$x = [x_1, x_2, \cdots, x_n]^{\mathbf{T}}$$

Where n is the mechanism synthesis point number.

The iterative function of multi-Newton optimization method is as follows.

$$x_{n+1} = N(x) = x_n - H(x_n)^{-1} \nabla F(x_n) \tag{6}$$

Step 2: Generate the iterative initial points

Use the chaotic anti-control equation (2) in autonomous system of mechanical centrifugal flywheel governor to generate the initial sequences $x00(i,k)$ $(i = 1,2,\cdots,n, k = 1,2,\cdots,N)$, where n is the number of variable , j is the length of data sequences) which is the same to the number of variables, then transform the chaotic sequences $x00(i,k)$ into variable interval, we get $x0(i,k)$. Take $x0(:,k)$

as the initial point of Newton iterative of equations (6), we can obtain all the solutions or most of solutions of synthesis question of accurate point.

Step 3: Substitute every real solution in equations (5) to check the functional values, choose the functional values which is less than positive ε according to the design accuracy. Determine the optimal dimensional scheme according to the nonnegative and not long length, existed crank, embranchment sequence and transmission angle.

4 Guidance Synthesis of Planar Four-Linkage Mechanism

The planar four-linkage mechanism is shown in Fig. 5. The guidance synthesis of rigid body requests that linkage $A_1B_1P_1$ is through certain locations, every linkage has θ_{1j} ($j = 1,2,\cdots,n$) angles compared to the first location. We divide the mechanism into two dyadic linkages $A_0A_1P_1$ and $B_0B_1P_1$. Since two dyadic linkages have the same given condition and kinematic synthesis equations, the results of kinematic synthesis are the same. We analyze one dyadic linkage as Fig.6. We take the coordinates A_{0x} and A_{0y} of stationary joint central A_0 and the coordinates A_{1x} and A_{1y} of the first location as the design variables denoted by x_1 , x_2 , x_3 and x_4 . On the constraint condition and transformation matrix of stable linkage length, the synthesis equations are as follows [9].

$$f_j(x) = P_{j1}x_1x_3 + P_{j2}x_1x_4 + P_{j3}x_2x_3 + P_{j4}x_2x_4 + P_{j5}x_1 + P_{j6}x_2 + P_{j7}x_2 + P_{j8}x_3 + P_{j9} \quad (7)$$

Where, $j = 2,3,\cdots,n, n$ is number of the synthesis points.

$$\begin{aligned}
P_{j1} &= 1-\cos(\theta_{1j}) \qquad P_{j2} = \sin(\theta_{1j}) \\
P_{j3} &= -\sin(\theta_{1j}) \qquad P_{j4} = 1-\cos(\theta_{1j}) \\
P_{j5} &= -(Pjx - P_{1x})\cos(\theta_{1j}) + P_{1y}\sin(\theta_{1j}) \\
P_{j6} &= -(P_{jy} - P_{1x}\sin(\theta_{1j})) - P_{1y}\cos(\theta_{1j}) \\
P_{j7} &= -\cos(\theta_{1j})P_{j5} - \sin(\theta_{1j})P_{j6} \\
P_{j8} &= \sin(\theta_{1j})P_{j5} - \cos(\theta_{1j})P_{j6} \\
P_{j9} &= -(P_{j5}^2 + P_{J6}^2)/2
\end{aligned}$$

In equations (7), there are four design variables and n-1 design equations. So the number of accuracy synthesis is five. When $n \le 5$, use Homotopy or chaotic to synthesize, and, when $n > 5$, only use Homotopy optimization method or chaotic optimization method to make approximately synthesis.

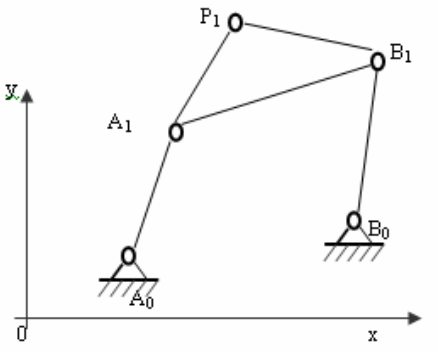


Fig. 5. Planar four-linkage mechanism

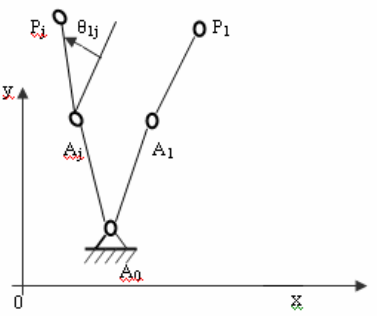


Fig. 6. Two locations of dyadic linkages

5 Examples

If there is a rigid body guidance mechanism of planar four-linkage, θ_{1j} and coordinate is shown as Table 1. Find all solutions.

On the condition of n=5, use accurate point synthesis method to find the solutions. Now use the method proposed in this paper and compile Matlab/Simulink program to find all real solutions (the chaotic sequences are produced by Eq.(2). Put every real solution into formula (5) to check the value of objective functions, then get the the dyadic linkage (Table 2) that satisfies the demand.

Table 1. θ_{1j} and corresponding coordinate

j	1	2	3	4	5
P_j	(0,2)	(-1,1.5)	(-1.5,0.5)	(-0.5,0.5)	(0,1)
θ_{1j}	0	3	17	32	26

Table 2. Synthesis results of dyadic linkage guidance synthesis

No	x_1	x_2	x_3	x_4
1	-1.08806	-1.36136	-.15230	-0.28868
2	5.12738	1.86785	5.77998	1.12524
3	3.41455	-2.00586	1.89649	3.82482
4	3.8457	-2.8608	1.66389	5.03998

Integrating two of them, we can get six planar mechanisms (Table 3) which satisfies the demand. Simulation result demonstrate that the accuracy satisfies the design demand and the synthesis results of mechanism are 1,2,3 and 4 which are the same to literature [9].

Table 3. Six planar four-linkage mechanism of synthesis

N0.	A_{0x}	A_{0y}	A_{1x}	A_{1y}	B_{0x}	B_{0y}	B_{1x}	B_{1y}	Notes
1	-1.08806	-1.36136	-1.15230	-0.28868	3.41455	-2.00586	1.89649	3.82482	Satisfy demand
2	-1.08806	-1.36136	-1.15230	-0.28868	3.84571	-2.86082	1.66389	5.03998	Satisfy demand
3	3.41455	-2.00586	1.89649	3.82482	5.12738	1.86785	5.77998	1.12524	Satisfy demand
4	5.12738	1.86785	5.77998	1.12524	3.84571	-2.86082	1.66389	5.03998	Satisfy demand
5	-1.08806	-1.36136	-1.15230	-0.28868	5.12738	1.86785	5.77998	1.12524	Not
6	3.41455	-2.00586	1.89649	3.82482	3.84571	-2.86082	1.66389	5.03998	Satisfy demand

6 Conclusions

The novel mechanism synthesis method is firstly proposed which based on chaos anti-control method of mechanical system. The rigid body guidance synthesis of planar four-linkage mechanism is researched with the method. The result shows the novel method does not construct the initial mechanism and can get groups of mechanism project. It makes a good basis for multi-project optimization-selection. It has the meaning of investigation of planar and spatial mechanism synthesis. The method can also adapt to a mechanism approximate synthesis.

Acknowledgments. This research is supported by the grant of the 11th Five-Year Plan for the construct program of the key discipline (Mechanical Design and Theory) in Hunan province (XJT2006180), National Science Foundation of China (50845038).

References

1. Luo, Y.X., Guo, H.X., Xiao, W.Y.: Research on Newton Optimization Iteration Chaos Methods of Mechanism Synthesis. Journal of Mechanical Transmission 29(6), 11–13 (2005)
2. Luo, Y.X.: Optimization-chaos method for synthesis of approximate guidance on planar four-bar linkage. Machine Design 20(11), 34–36 (2003)
3. Zhang, L.T., Dong, X.H.: Chao Solution of Non - linear Equations and Its Application to Mechanism Synthesis. Machine Tool & Hydraulics 31(1), 165–169 (2003)
4. Luo, Y.X., Li, D.Z.: Finding all solutions to forward displacement analysis problem of 6-SPS parallel robot mechanism with chaos-iteration method. Engineering Design 10(2), 34–36 (2003)
5. Xie, J., Chen, Y.: Application of Chaos Theory to Synthesis of Plane Rigid Guidance. Mechanical Science and Technology 19(4), 524–526 (2000)
6. Xie, J., Chen, Y.: Chaos in the Application of Newton Raphson Iteration to Computational Kinematics. Journal of Mechanical Transmission 24(1), 4–6 (2000)

7. Xie, J., Chen, Y.: A Chaos-based Approach to Obtain the Global Real Solutions of Burmester Points. China Mechanical Engineering 13(7), 608–710 (2002)
8. Xun, X.F., Luo, G.W.: Nonlinear Feedback Anti-control of Chaos for Mechanical Centrifugal Flywheel Governor Systems. China Mechanical Engineering 16(15), 1373 (2005)
9. Liu, A.X., Yang, T.L.: Design of Mechanical System Kinematics. China Petrochemical Press, Beijing (1999)

A Method of Stiffness Analysis of Parallel Mechanisms

Boqiang Xu, Tiemin Li, and Jun Wu

Institute of Manufacturing Engineering, Department of Precision Instruments and Mechanology, Tsinghua University, Beijing, 100084, China
xbqwss@sina.com, litm@tsinghua.edu.cn

Abstract. This paper presents a method, called Parameter-Separation Method, for stiffness analysis of parallel mechanisms. By treating the effect of different parameters (i.e. external force and torque, translational and rotational displacement) separately, the method produces 4 criteria of stiffness analysis. A further method based on the first method is proposed when we conduct stiffness analysis on some parallel mechanisms whose inputs are of the same unit. It applies a special way to formulate dimensionally homogeneous Jacobian matrix with consideration of the input need, and comes up with 2 criteria to judge the stiffness of parallel mechanisms. These two methods are explained in detail, and their effects and relationship are demonstrated by stiffness analysis of a 4RRR redundantly actuated parallel mechanism.

Keywords: Stiffness analysis, parallel mechanism, 4RRR parallel mechanism, Parameter-Separation Method.

1 Introduction

Parallel mechanisms (PM) are claimed to be able to offer a higher stiffness-to-mass ratio and better accuracy compared with those traditional serial mechanisms, which make them feasible to be used in high speed machining [1]. Many researches, concerning different aspects of a variety of PMs, have been done by scholars, among which stiffness analysis is of vital importance [2] [3].

A traditional method of calculating stiffness is to analyze the Jacobian matrices of PMs, taking into account the compliance of the actuated joints [4]. This theory can apply to manipulators having degrees of freedom in either translational or rotational direction only, but not combination of both. Otherwise, if the mechanism has complex degrees of freedom, the Jacobian matrix is dimensionally inconsistent. In this situation, methods for formulating a dimensionally homogeneous Jacobian matrix will commonly be applied, which have been studied by many researchers [5][6]. However, these methods have some drawbacks, such as failing to maintain a mathematically symmetric structure of the mechanism, obscuring physical interpretation of some parameters, etc. This paper introduces a new method for stiffness analysis which does not have such problems.

The rest of the paper is organized as follows. In section 2, the Parameter-Separation Method (PSM) and its derivative method, Combined-Parameter-Separation Method (CPSM), are proposed, explained and compared. In section 3, we apply the PSM and CPSM to stiffness analysis of redundantly actuated 4RRR PM. In section 4, the advantages of these two methods are summarized at last.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 562–570, 2009.

Method

2.1 Parameter-Separation Method

Some methods of stiffness analysis describe the anisotropy behavior of stiffness of PM by inviting the stiffness matrix **C**, so the relationship between external wrench and displacement of moving platform can be described in the following formula:

$$\mathbf{W} = \mathbf{J}^{\mathbf{T}}\mathbf{K}\mathbf{J}\Delta\mathbf{p} = \mathbf{C}\Delta\mathbf{p} \tag{1}$$

Where **J** denotes the inverse Jacobian matrix, **K** denotes the stiffness of all actuators, **Δp** denotes the displacement of moving platform and **W** denotes the external wrench. The minimum eigenvalue of **C**, which represents the minimum stiffness in certain direction, is the focus of researches.

The external wrench exerted on the moving platform can be divided into external force (EF) and external torque (ET), the displacement of moving platform can be divided into translational displacement (TD) and rotational displacement (RD). Generally, these 4 parameters do not share the same unit. So these parameters should be divided up, and the effect of EF on TD, ET on TD, EF on RD and ET on RD should be examined separately. Then formula (1) can be written as

$$\begin{bmatrix}\mathbf{Dt}\\ \mathbf{Dr}\end{bmatrix} = \Delta\mathbf{p} = \mathbf{C}^{-1}\mathbf{W} = \begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{B}^{\mathbf{T}} & \mathbf{D}\end{bmatrix}\begin{bmatrix}\mathbf{F}\\ \mathbf{T}\end{bmatrix} \tag{2}$$

Where $\mathbf{Dt} = [dx \quad dy \quad dz]^T$, $\mathbf{Dr} = [d\alpha \quad d\beta \quad d\gamma]^T$, $\mathbf{F} = \begin{bmatrix}F_x & F_y & F_z\end{bmatrix}^T$, $\mathbf{T} = \begin{bmatrix}T_x & T_y & T_z\end{bmatrix}^T$ represent the vector of TD, RD, EF, and ET respectively. Formula (2) can be transformed to

$$\begin{bmatrix}dx\\ dy\\ dz\end{bmatrix} = \mathbf{A}\begin{bmatrix}F_x\\ F_y\\ F_z\end{bmatrix} + \mathbf{B}\begin{bmatrix}T_x\\ T_y\\ T_z\end{bmatrix}, \quad \begin{bmatrix}d\alpha\\ d\beta\\ d\gamma\end{bmatrix} = \mathbf{B}^{\mathbf{T}}\begin{bmatrix}F_x\\ F_y\\ F_z\end{bmatrix} + \mathbf{D}\begin{bmatrix}T_x\\ T_y\\ T_z\end{bmatrix} \tag{3}$$

A, B, $\mathbf{B}^{\mathbf{T}}$, D reflect the effect of EF on TD, ET on TD, EF on RD and ET on RD respectively. Then we have 4 types of stiffness: the translational stiffness for force (*TSF*), the translational stiffness for torque (*TST*), the rotational stiffness for force (*RSF*) and the rotational stiffness for torque (*RST*), which can be defined by the inverse of the 2-norms of **A, B, $\mathbf{B}^{\mathbf{T}}$, D** respectively as in formula (4).

$$TSF = \frac{1}{\|\mathbf{A}\|_2}, TST = \frac{1}{\|\mathbf{B}\|_2}, RSF = \frac{1}{\|\mathbf{B}^{\mathbf{T}}\|_2}, RST = \frac{1}{\|\mathbf{D}\|_2} \tag{4}$$

2.2 Combined-Parameter-Separation Method

In most PMs, the inputs share the same unit, for example, 3RRR parallel mechanism has 3 revolute inputs, and Stewart platform has 6 translational inputs. In this situation, TD and RD can be combined together through formulating a dimensionally

homogeneous Jacobian matrix in a way to be introduced later which treats only the effect of EF and ET separately. This method produces two combined criteria, termed as *CSF* and *CST*, which represent the combined stiffness for EF and combined stiffness for ET.

To formulate the dimensionally homogeneous Jacobian matrix, it is necessary to define a weight *kr* to balance the importance of TD and RD. In order to define the *kr*, we should first examine the importance of the TD (*It*) and the importance of RD (*Ir*). The 'importance' of certain type of displacement can be judged from the magnitude of input needed to render a unit displacement of such type. So *It* is defined as the maximum magnitude of input vector $\mathbf{\Delta q}$ which will result in a unit vector of TD, *Ir* is defined as the maximum magnitude of input vector $\mathbf{\Delta q}$ which will result in a unit vector of RD in certain direction. In PMs, the relationship between input displacement $\mathbf{\Delta q}$ and outputs displacement $\mathbf{\Delta p}$ can be expressed as

$$\mathbf{\Delta q} = \mathbf{J}\mathbf{\Delta p} = \begin{bmatrix}\mathbf{G} & \mathbf{H}\end{bmatrix}\begin{bmatrix}\mathbf{Dt}\\ \mathbf{Dr}\end{bmatrix} \tag{5}$$

From formula (5) and definition of *It* and *Ir*, we see *It* and *Ir* can be expressed as the 2-norms of $\mathbf{G}$ and $\mathbf{H}$ respectively. Further, we define:

$$kr = It / Ir \tag{6}$$

Where $It = \|\mathbf{G}\|_2$, $Ir = \|\mathbf{H}\|_2$, *kr* represents the ratio of magnitude of input vector to render a unit RD and that to render a unit TD. To be noted, *kr* is not uniform throughout the workspace because the ratio of importance of the TD to that of RD is not constant at different configurations.

After calculating the *kr*, we have

$$\mathbf{\Delta q} = \mathbf{J}\mathbf{\Delta p} = \mathbf{J}\begin{bmatrix}\mathbf{I}_1 & 0\\ 0 & kr\times\mathbf{I}_2\end{bmatrix}\begin{bmatrix}\mathbf{Dt}\\ \frac{1}{kr}\mathbf{Dr}\end{bmatrix} = \mathbf{JV}\begin{bmatrix}\mathbf{Dt}\\ \frac{1}{kr}\mathbf{Dr}\end{bmatrix} = \mathbf{J}'\begin{bmatrix}\mathbf{Dt}\\ \frac{1}{kr}\mathbf{Dr}\end{bmatrix} = \mathbf{J}'\mathbf{\Delta q}' \tag{7}$$

Where $\mathbf{I}_1, \mathbf{I}_2$ are unit diagonal matrices. At last, formula (1) can be expressed as

$$\begin{bmatrix}\mathbf{Dt}\\ \frac{1}{kr}\mathbf{Dr}\end{bmatrix} = \mathbf{V}^{-1}\mathbf{C}^{-1}\begin{bmatrix}\mathbf{F}\\ \mathbf{T}\end{bmatrix} = \begin{bmatrix}\mathbf{M} & \mathbf{N}\end{bmatrix}\begin{bmatrix}\mathbf{F}\\ \mathbf{T}\end{bmatrix} \tag{8}$$

where $\mathbf{M}$ and $\mathbf{N}$ reflect the effect of EF, ET on the combined stiffness, and we get two criteria of combined stiffness of EF and ET, namely *CSF*, *CST* in formula (9):

$$CSF = \frac{1}{\|\mathbf{M}\|_2}, CST = \frac{1}{\|\mathbf{N}\|_2} \tag{9}$$

Using *CSF* and *CST*, the stiffness analysis of parallel mechanism can be more meaningfully conducted.

2.3 Relationship of PSM and CPSM

From formula (2), (7) and (8), we can get:

$$\mathbf{M}=\begin{bmatrix}\mathbf{A}\\ \frac{1}{kr}\mathbf{B}^{\mathrm{T}}\end{bmatrix},\ \mathbf{N}=\begin{bmatrix}\mathbf{B}\\ \frac{1}{kr}\mathbf{D}\end{bmatrix} \tag{10}$$

Formula (4),(9) and (10) show that the relationship of *CSF*, *TSF*, *RSF* is similar to that of the springs in serial, shown in Fig. 1, where *CSF* represents the whole stiffness and *TSF*, $\frac{1}{kr}\times RSF$ represent the stiffness of solo springs in serial. And the relationship can be expressed in formula (11)

$$\frac{1}{CSF}\approx\frac{1}{TSF}+\frac{1}{RSF/kr} \tag{11}$$

A similar analysis can be applied on the relationship of *CST*, *TST* and *RST*, and this relationship can be expressed in formula (12)

$$\frac{1}{CST}\approx\frac{1}{TST/kr}+\frac{1}{RST} \tag{12}$$

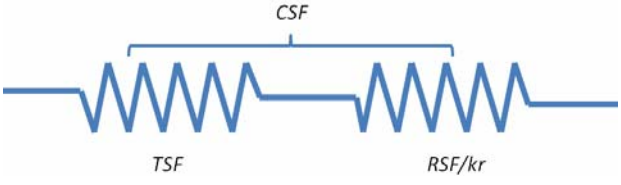


Fig. 1. Relationship in *CSF*, *TSF* and *RSF*

3 Stiffness Analysis of 4RRR Parallel Mechanism

3.1 4RRR Redundant Parallel Mechanism

4RRR redundant parallel mechanism is derived from 3RRR parallel mechanism by adding a redundant kinematic link in a symmetrical way to increase stiffness and avoid singularity. It has four kinematic chains and each chain has three revolute joints.

As shown in Fig. 2, four kinematic chains $A_iB_iC_i$ (i=1,2,3,4) connect the moving platform $C_1C_2C_3C_4$ to the base $A_1A_2A_3A_4$. The side length of $C_1C_2C_3C_4$ and $A_1A_2A_3A_4$ are denoted as *h* and *k* respectively. Lengths of A_iB_i and B_iC_i are denoted as l_1 and l_2. We locate the origin of the global frame at A_1 and the origin of the moving frame at the center of the moving platform, O_N. The orientation of the moving platform is

expressed with α, which is the angle between the axis X and the axis X_N. The position of moving platform is denoted by the coordinate of O_N, namely *x* and *y*. Four revolute inputs are denoted as θ_i (i=1,2,3,4).

In the following analysis, the parameters are set to $l_1 = l_2 = 800\,\text{mm}, h = 200\,\text{mm}$ and $k = 1000\,\text{mm}$, and we make moving platform move from p_1(0,500) to p_2 (1000, 500). At each point on the trace, the stiffness will be calculated when the moving platform rotates from 0 to -140 degree (minus indicates a clockwise rotation).

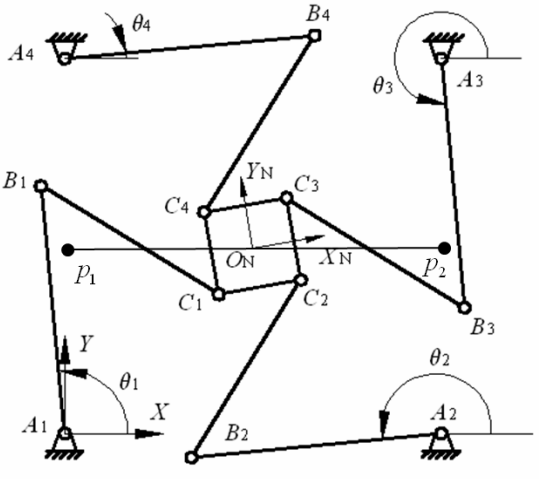


Fig. 2. 4RRR redundant mechanism

3.2 Stiffness Analysis Based on Parameter-Separation Method

The following part will apply PSM on 4RRR parallel mechanism. Fig. 3 and Fig. 4 demonstrate the distribution of 4 types of stiffness in 3D and contour graphs.

From these figures, following conclusions of the stiffness of 4RRR PM can be arrived in:

1. All 4 types of stiffness are relatively high when the moving platform is in the middle of the base platform (TD=500 mm) and the stiffness decreases when the moving platform moves apart from this center.

2. When the moving platform is in the center of the base platform, the *TST*, *RSF* are infinite, indicating that EF will result in little RD, and ET will lead to little TD, resulted from the symmetrical configuration of the mechanism. However, when the moving platform moves apart from the center, both the *TST* and *RSF* decrease, indicating the EF will result in the RD and the ET will result in TD.

3. *TSF*, *RST* first decrease and then increase with the rotation of the platform. At about -56 degree, *TSF*, *RST* drop to 0, the singularity occurs.

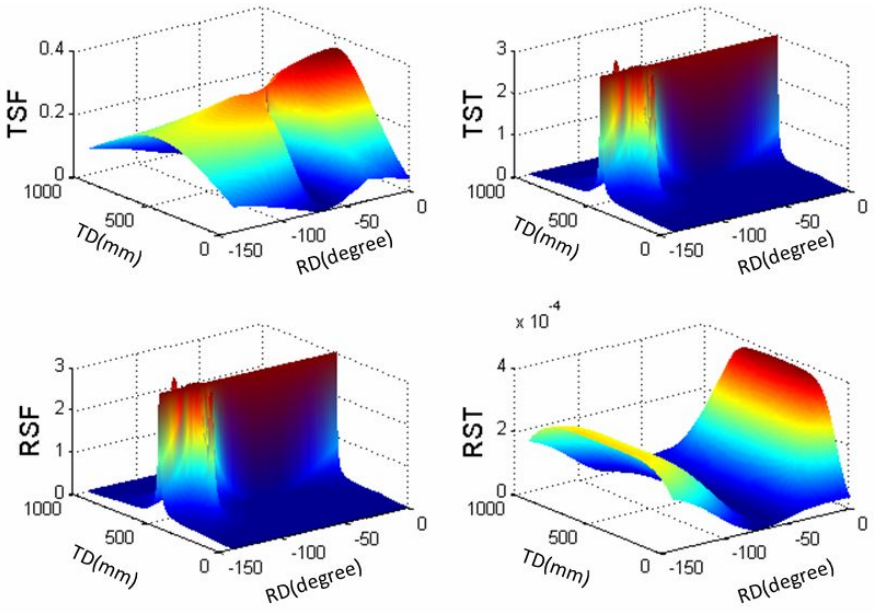


Fig. 3. 3D graph: four types of stiffness

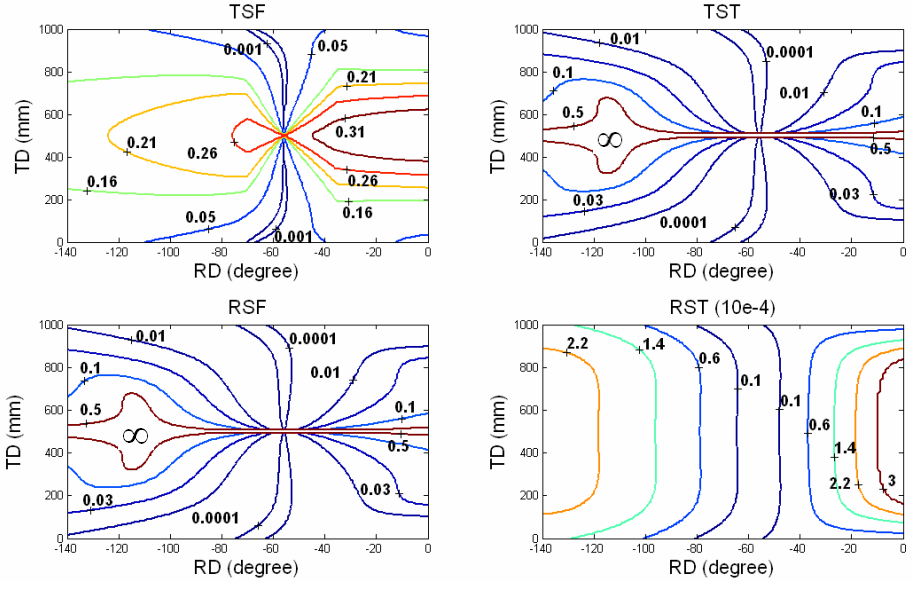
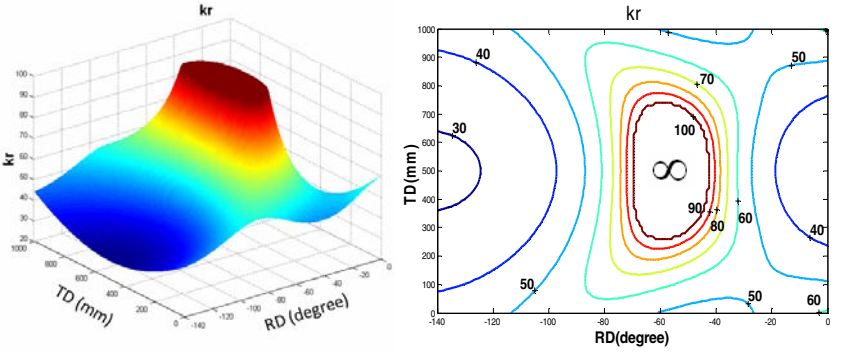


Fig. 4. Contour graph: four types of stiffness

3.3 Stiffness Analysis Based on Combined-Parameter-Separation Method

The following part will apply CPSM on the 4RRR PM. At first step, the *kr* should be calculated on the trace, the result is shown in Fig. 5 with 3D and contour graph.



(a) (b)

Fig. 5. Distribution of *kr*

Note that at about (5,-56) in Fig. 5 (b), the peak of *kr* is almost infinite. Except this particular area around (5,-56), the magnitude of *kr* varies around 50.

After *kr* is gotten, all the matrices in formula (8) can be obtained. Then the CSF and CST on the trace can be calculated, as shown in Fig. 6 and Fig. 7.

From Fig. 6 and Fig. 7, conclusions similar to conclusion 1 and 3 in section 3.2 can be arrived in. Conclusion 2 is blinded because we have combined the RD and TD together.

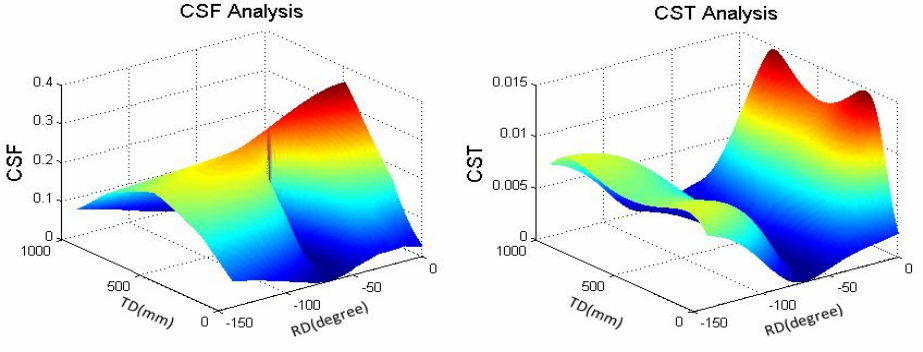


Fig. 6. 3D graph: *CSF* and *CST*

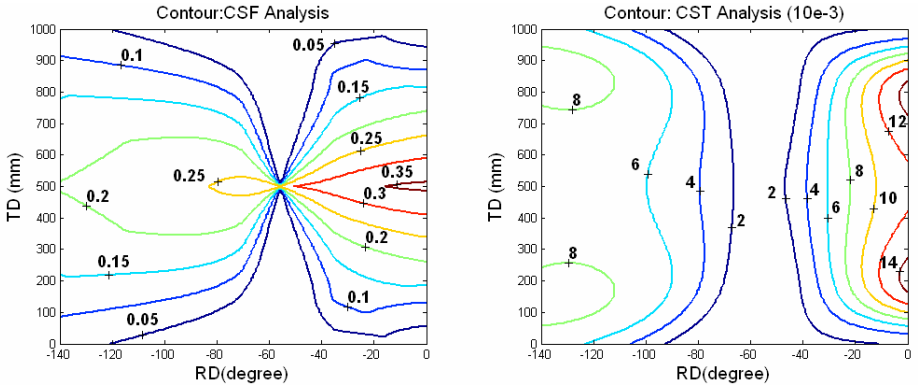


Fig. 7. Contour graph of *CSF* and *CST*

4 Summary

Through theoretical analysis and example illustration, this paper studies a new method of stiffness analysis, called Parameter-Separation Method. This method solves the problem of unit inconsistency by dividing up parameters which may incur such problems and analyzing them separately. Its 4 criteria have evident physical significance and can serve as helpful references in both theoretical and practical analysis of parallel mechanisms.

A derivative method, namely Combined-Parameter-Separation Method, is developed for parallel mechanisms whose inputs are of the same unit. In this method, a weight *kr* is proposed to not only solve the unit inconsistency of RD and TD, but also balance the importance of the RD and TD from the perspective of the need of input, making this method physically meaningful and be able to serve as a practical tool for design and control of parallel mechanisms.

References

1. Pashkevich, A., Chablat, D., Wenger, P.: Stiffness Analysis of 3-d.o.f Over constrained Translation Parallel Manipulators. In: 2008 IEEE International Conference on Robotics and Automation, Pasadena, CA,USA, May 19-23, pp. 1562–1567 (2008)
2. Kim, J.W., Kim, K.W., Kim, H.S., et al.: Stiffness Analysis and Design of a 3-DOF Parallel Robot with One Constraining leg. In: International Conference on Control, Automation and Systems, COEX, Seoul, Korea, October 17-20, pp. 2288–2293 (2007)
3. Liu, H., Ye, C., Wang, H., et al.: Stiffness Analysis and Experiment of a Parallel Kinematic Planer. In: The IEEE International Conference on Automation and Logistics, Jinan, China, August 18-21, pp. 953–958 (2007)
4. Gosselin, C.: Stiffness mapping for parallel manipulators. IEEE Transactions on Robotics and Automation 6, 377–382 (1990)

5. Pond, G., Carretero, J.A.: Formulating Jacobianan matrices for dexterity analysis of parallel manipulators. Mechanism & Machine Theory 41(12), 1505–1519 (2006)
6. Kotlarski, J., Abdellatif, H., Heimann, B.: Improving the pose accuracy of planar 3RRR parallel manipulator using kinematic redundancy and optimized switching patterns. In: 2008 IEEE International Conference on Robotics and Automation, Pasadena, CA, USA, May 19-23, pp. 3863–3868 (2008)

Numerical Methods for Frictional Contact of Multi-rigid-body with Redundant Constraints

Haitao Gao, Zhisheng Zhang, Jun Liu, Guang Lu, and Jinfei Shi

School of Mechanical Engineering, Southeast University, Nanjing, 211189, China
ght@seu.edu.cn, oldbc@seu.edu.cn, chris11@seu.edu.cn

Abstract. To tackle the frictional contact problem of multi-rigid-body with redundant constraints, efficient numerical resolution methods and stabilization algorithm are proposed. Firstly, based on time-stepping method and linear programming theories, a mixed nonlinear complementary model describing frictional contact problem is built. In order to solve the model effectively, the least square method for solving redundant constraints and linearization method are used to change the mixed nonlinear complementary problem into a linear complementary problem that can be easily solved. And then, a direct stabilization algorithm is given to stabilize resolution process of contact forces in advance, which effectively eliminates the drift problem for both equality and inequality constraints. At last, the validity of the numerical resolution methods and stabilization algorithm are verified through a numerical example.

Keywords: Redundant constraints, multi-rigid-body, friction, contact, numerical methods.

1 Introduction

Multibody dynamic systems with friction and contact in which the interacting bodies are nominally rigid are broadly involved in many engineering applications, such as robotic manipulation, graphic simulation, and haptic display. So how to establish frictional contact model and find out efficient numerical algorithm is always essential research content. Contacts in multibody dynamics may be modeled using a local stiffness discretisation according to the contact and a regularization of the friction characteristics. However this method leads to stiff differential equations with high computational times and parameter uncertainties. At present, the popular method is that these unilateral contact problems have been effectively formulated as complementarity problems. Complementarity models fall into two kinds of description methods based on measure differential equations. The first is an acceleration-based approach [1]-[3]. In which contact laws and frictional conditions are described as relationship between acceleration and force respectively, and then complementarity model for frictional contact is built subject to dynamic formulation. However this acceleration-based approach has the disadvantage of that it is necessary to check if the obtained results are compatible with the contact law in solving process, due to acceleration is unable to detect contact state directly. So an event-driven scheme is necessary to detect changes

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 571–579, 2009.

of the constraints (events) such as stick-slip transitions and resolve the exact transition times, which is carried out using the interpolation methods.

The second approach is velocity-based which also called the time-stepping method [4], [5]. In the method, Newton–Euler equation is written into a finite-difference equation based on a time discretisation schemes. As a result, the force-acceleration relationship is replaced by an impulse-velocity relationship. In other words, the integral of the force over each time step is used instead of the instant values of the force. In this method, the frictional contact models are described by velocity and impulse, the whole set of discretized equations and constraints are used to compute the next state of the motion. The advantage of this approach is that there is no need to monitor the contact-state transitions so that this approach is popular in robot simulation and computer animation in recent years.

One major problem in the simulation of mechanical systems with frictional contact is the computation of the contact forces. Appropriate contact forces must be determined among all contacts so that the basic condition of nonpenetration and the relevant contact laws are satisfied. Currently, several approaches have been proposed, such as simplified friction law, the problem can be formulated as a linear complementarity problem by linearizing the friction cone [6]. Then, the resulting LCP can be solved with the well-known Lemke's pivoting scheme or the iterative methods [3], [7].

In summary, the previous studies are mainly concentrated on the frictional contact problem of single rigid body or multi-rigid-body with holonomic constraints, however there is some multi-rigid-body with redundant constraints in practical application which is difficult to solve. Study on frictional contact problem of these systems has not been reported in the related literature. In the paper, we focus on numerical methods and stabilization algorithm for frictional contact problem of multi-rigid-body with redundant constraints.

2 Dynamic Equations of Multi-rigid-body with Frictional Contact

In order to describe contacts problem in multi-rigid-body system, we assume that each contact is isolated point and the shape of contact body is smooth and convex. As a result, it is convenient to define normal and tangential subspace. Let $g_{N_i}(q,t)$ represents normal gap function and $g_{T_i}(q,\dot{q},t)$ represents tangential relative velocities in tangential direction at each contact, where i is the number of contact and $q,\dot{q}$ are generalized coordinates and generalized velocity respectively. Take $W_{N_i} = \partial g_{N_i} / \partial q$, $W_{T_i} = \partial \dot{g}_{T_i} / \partial \dot{q}$, then dynamic equation of multi-rigid-body with multiple frictional contact can be written as:

$$M\ddot{q} + W_E \lambda_E = \sum_{i=1}^{m} (W_{N_i} \lambda_{N_i} + W_{T_i} \lambda_{T_i}) + Q \,. \tag{1}$$

$$g_E(q,t) = 0 \,. \tag{2}$$

Where M is the generalized mass matrix, $W_E = \partial g_E / \partial q$ is constraint Jacobian matrix, and Q is external forces which include the coriolis forces, $\lambda_E, \lambda_{N_i}, \lambda_{T_i}$ are constraint forces, normal and tangential contact forces respectively. The key to solve above equations is to obtain contact forces.

3 Establishment of the Frictional Contact Model

Let h denote a simulation step size, t^l is the current time and t^{l+1} is the next time, the following notations are used in the paper:

$t^{l+1} = t^l + h$, $q^l = q(t^l)$, $\dot{q}^l = u^l = u(t^l)$, $q^{l+1} = q(t^l + h)$, $u^{l+1} = u(t^l + h)$, $g_{N_i}(q^{l+1}, t^{l+1}) = g_{N_i}^{l+1}$, $g_{T_i}(q^{l+1}, u^{l+1}, t^{l+1}) = g_{T_i}^{l+1}$, $\Delta q = u^{l+1} h$, $\Delta u = u^{l+1} - u^l$.

Since the dynamics equations (3) that relate accelerations to forces are not suitable to describe motions with impacts. A more appropriate discrete-time formulation can be achieved by approximating $\frac{\partial u}{\partial t}$ by $\frac{\Delta u}{h}$, $\frac{\partial q}{\partial t}$ similarly:

$$M\Delta u + W_E \lambda_E = \sum_{i=1}^{m} (W_{N_i} S_{N_i} + W_{T_i} S_{T_i}) + Qh \tag{3}$$

Where $S_{N_i} = \lambda_{N_i} \cdot h$ and $S_{T_i} = \lambda_{T_i} \cdot h$ are the unknown generalized contact impulses. It is should noticed that since the contact forces are known only at the end of each time interval, the definitions of the impulses imply that we must view the contact forces as constant over each time interval.

When each pair of bodies comes in contact in multi-rigid-body at t^{l+1}, their contact forces and normal gap function forms a complementary condition:

$$g_{N_i}^{l+1} \geq 0 \,,\ \lambda_{N_i}^{l+1} \geq 0 \,,\ g_{N_i}^{l+1} \cdot \lambda_{N_i}^{l+1} = 0 \,.\ i{=}1 \ldots m. \tag{4}$$

According to above time discretisation method, it is known that $\lambda_{N_i}^{l+1} \geq 0$ is equivalent to $S_{N_i} = \lambda_{N_i} \cdot h \geq 0$. So the normal contact laws also can be written as follows:

$$g_{N_i}^{l+1} \geq 0 \,,\ S_{N_i}^{l+1} \geq 0 \,,\ g_{N_i}^{l+1} \cdot S_{N_i}^{l+1} = 0 \,. \tag{5}$$

Classical coulomb's friction law is used to model the friction force and the tangential forces are limited by a maximal friction force dependent on $\mu\lambda$ in any direction inside the tangential plane, where μ denotes the friction coefficient. Lower forces are concerned with sticking contacts and thus $g_T^{l+1} = 0$. Contrastively, forces reaching the boundary of the friction cone may declare sliding and thus $g_T^{l+1} \neq 0$. Similar to the normal contact conditions, the tangential contact law can be modified for the discrete-time setting by replacing force variables with impulse variables:

$$\left| S_{T_i}^{l+1} \right| \leq \mu_i S_{N_i}^{l+1} \,,\ g_{T_i}^{l+1} \geq 0 \,,\ (\left| S_{T_i}^{l+1} \right| - \mu_i S_{N_i}^{l+1}) \cdot g_{T_i}^{l+1} = 0 \,. \tag{6}$$

Equations (4) and (5) only describe contact and friction laws, which are insufficient to model contact problem in multi-rigid-body. For this reason, contact kinematics of smooth-rigid bodies is needed to supplement. The normal gap formation and the tangential relative velocity are substituted by Taylor series expansions restricted to the first order:

$$g_{N_i}^{l+1} = g_{N_i}^{l} + W_{N_i}^{T} u^{l+1} h + \tilde{W}_{N_i} h \,. \tag{7}$$

$$\dot{g}_{T_i}^{l+1} = \dot{g}_{T_i}^{l} + (W_{T_i}^{T} + \hat{W}_{T_i} h) u^{l+1} - W_{T_i}^{T} u^{l} + \tilde{W}_{T_i} h \,. \tag{8}$$

Where $\tilde{W}_{T_i} = \partial g_{T_i} / \partial t$, $\hat{W}_{T_i} = \partial g_{T_i} / \partial q$.

Contact laws (5)-(8); dynamics equations (3) and bilateral constraint equations (2) constitute a Mixed Nonlinear Complementarity Problem (MNCP) describing friction and contact problem.

4 Resolution of Frictional Contact Model

4.1 Least Squares Resolution of Redundant Constraint

Contact forces and the generalized velocity of the system at t^{l+1} is obtained by resolution of MNCP, which can then be substituted into equation (3) to obtain an estimate of system configuration. A simple method for easy resolution of MNCP is to eliminate bilateral constraint equations (2) so that the MNCP can be converted into a standard NCP. In order to do this, velocity constraint equations are obtained by Taylor series expansion of equations (2).

$$W_E^{T} u^{l+1} = -g_E^{l} / h - \partial g_E^{l} / \partial t \,. \tag{9}$$

Cast equations (3) and (9) into the matrix form:

$$\begin{pmatrix} M & W_E \\ W_E^{T} & 0 \end{pmatrix} \begin{pmatrix} u^{l+1} \\ S_E \end{pmatrix} = \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} \tag{10}$$

Where $d_1 = \sum_{i=1}^{m} (W_{N_i} S_{N_i} + W_{T_i} S_{T_i}) + M u^{l} + Q h$, $d_2 = -g_E^{l} / h - \partial g_E^{l} / \partial t$.

We can solve equations (10) for u^{l+1} and eliminate it to yield a standard NCP. However, when coefficient matrix is irreversible due to redundant constraints in multi-rigid-body, equations (10) is difficult to solve by popular numerical methods. A better method is to search for an approximate solution of equations which make:

$$\left\| \begin{pmatrix} M & W_E \\ W_E^{T} & 0 \end{pmatrix} \begin{pmatrix} u^{l+1} \\ S_E \end{pmatrix} - \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} \right\|_2 = \min$$

The unique approximate solution is also known as minimum norm solution and the resolution method is least-squares method [8]. It is not necessary to assume that the coefficient matrix is full rank in the solving process by this method, so that redundant constraints problem is able to be handled. Based on block matrix and least-squares method, the solution of equation (10) is expressed by:

$$\begin{pmatrix} u^{l+1} \\ S_E \end{pmatrix} = \begin{pmatrix} \hat{M} & \tilde{M} \\ \tilde{M}^T & \bar{M} \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}. \tag{11}$$

Where $\tilde{M} = G_{EE}^{+} W_E^T M^{-1}$, $\hat{M} = M^{-1} + M^{-1} W_E G_{EE}^{+} W_E^T M^{-1}$, $\bar{M} = G_{EE}^{+}$, $G_{EE} = W_E^T M^{-1} W_E$, G_{EE}^{+} is pseudo-inverse.

The result: $u^{l+1} = \hat{M} d_1 + \tilde{M} d_2$ is substituted into equations 7 , 8 and yield the following equations:

$$\begin{pmatrix} g_{N_i}^{l+1} / h \\ \dot{g}_{T_i}^{l+1} \end{pmatrix} = \sum_{j=1}^{m} \begin{pmatrix} G_{N_i N_j} & G_{N_i T_j} \\ \hat{G}_{T_i N_j} & \hat{G}_{T_i T_j} \end{pmatrix} \begin{pmatrix} S_{N_j}^{l+1} \\ S_{T_j}^{l+1} \end{pmatrix} + \begin{pmatrix} b_{N_i} \\ b_{T_i} \end{pmatrix} \tag{12}$$

Where $b_{T_i} = \dot{g}_{T_i}^{l} + (W_{T_i}^T + \hat{W}_{T_i}) \hat{M} (Mu^l + Qh) + (W_{T_i}^T + \hat{W}_{T_i}) \tilde{M} d_2 + \tilde{W}_{T_i} h - W_{T_i}^T u^l$, $G_{\alpha\beta} = W_\alpha^T \hat{M} W_\beta$, $b_{N_i} = g_{N_i}^{l} / h + W_{N_i}^T \hat{M} (Mu^l + Qh) + W_{N_i}^T \tilde{M} d_2 + \tilde{W}_{N_i}$, $\hat{G}_{\alpha\beta} = (W_\alpha^T + \hat{W}_\alpha^T \cdot h) \hat{M} W_\beta, \alpha, \beta = (N, T)$.
Then, equations (4), (5) and (12) constitute a standard NCP.

4.2 Resolution of NCP Equations

In general, NCP equations are not directly solved because of the complex solving process. One way to solve the NCP equations is to transform into a linear complementarity problem (LCP) by approximating friction [9]. In the planar case, it can be done by introducing the following definition of the friction saturation and tangential velocity in positive and negative direction

$$\sigma^+ = \mu S_N + S_T,\ \sigma^- = \mu S_N - S_T,\ k^+ = \tfrac{1}{2}(|\dot{g}_T| + \dot{g}_T),\ k^- = \tfrac{1}{2}(|\dot{g}_T| - \dot{g}_T).$$

For each contact, take $\bar{C}_i^{l+1} = (g_{N_i}^{l+1} / h, k_i^+, \sigma_i^-)^T$, $\bar{S}_i^{l+1} = (S_{N_i}^{l+1}, \sigma_i^+, k_i^-)^T$, $\bar{b}_i = (b_{N_i}, b_{T_i}, 0)$.
The complete contact problem can be written as a linear complementarity problem as follows:

$$\begin{pmatrix} \bar{C}_1^{l+1} \\ \bar{C}_2^{l+1} \\ \vdots \\ \bar{C}_m^{l+1} \end{pmatrix} = \begin{pmatrix} \bar{G}_{11} & \bar{G}_{12} & \cdots & \bar{G}_{1m} \\ \bar{G}_{21} & \bar{G}_{22} & \cdots & \bar{G}_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{G}_{m1} & \bar{G}_{m2} & \cdots & \bar{G}_{mm} \end{pmatrix} \begin{pmatrix} \bar{S}_1^{l+1} \\ \bar{S}_2^{l+1} \\ \vdots \\ \bar{S}_m^{l+1} \end{pmatrix} + \begin{pmatrix} \bar{b}_1 \\ \bar{b}_2 \\ \vdots \\ \bar{b}_m \end{pmatrix}$$

$$\begin{pmatrix} \bar{C}_1^{l+1} \\ \bar{C}_2^{l+1} \\ \vdots \\ \bar{C}_m^{l+1} \end{pmatrix} \geq 0;\ \begin{pmatrix} \bar{S}_1^{l+1} \\ \bar{S}_2^{l+1} \\ \vdots \\ \bar{S}_m^{l+1} \end{pmatrix} \geq 0;\ \begin{pmatrix} \bar{C}_1^{l+1} \\ \bar{C}_2^{l+1} \\ \vdots \\ \bar{C}_m^{l+1} \end{pmatrix}^T \begin{pmatrix} \bar{S}_1^{l+1} \\ \bar{S}_2^{l+1} \\ \vdots \\ \bar{S}_m^{l+1} \end{pmatrix} = 0 \tag{13}$$

Where $\overline{G}_{ij} = \begin{pmatrix} G_{N_iN_j} - G_{N_iN_j}\mu_j & G_{N_iT_j} & 0 \\ \hat{G}_{T_iN_j} - \hat{G}_{T_iT_j}\mu_j & \hat{G}_{T_iT_j} & Z \\ 2\mu_j & -Z & 0 \end{pmatrix}$, $\begin{cases} i = j, Z = 1 \\ i \neq j, Z = 0 \end{cases}$, $i, j = 1...m$.

μ_j is the each frictional coefficients. This model can be simplified as the following abbreviations:

$$y = Ax + b\ ,\ y \geq 0\ ,\ x \geq 0\ ,\ x^T y = 0\ . \tag{14}$$

Equations (16) are a standard LCP and can be solved with a pivoting algorithm like Lemake's method. However this approach has disadvantages of a large amount of calculation and accumulated error. Iterative methods such as splitting projected Gauss-Seidel iteration scheme are a better choice [7], which is easy to implement in general, especially when the sparsity of the problem can be used effectively. Solving these LCP equations is able to yield each contact forces, and then the result can be embedded in differential algebraic equations to yield the motion of next time.

5 Stabilization Method

In this numerical method, velocity constraint equations derived by position constraint equations are used for simplification of contact model, they are not completely equivalent to constraint condition so that this method will inevitably lead to numerical error. Moreover, minimum norm solution which is used to approximate the solution of velocity constraint equations also raises numerical error. These errors can affect accuracy of contact forces and result in drift of motion and must be stabilized to continue to satisfy constraint. Given that the numerical solution of velocity $\tilde{u}^l$ is obtained after *n* integration step. Usually, there is a small error Δu^l making the constraints $\dot{g}(q^l, \tilde{u}^l) \neq 0$ which will be taken into equations (7), (8) and give rise to error of g_N^{l+1} and $\dot{g}_T^{l+1}$. In order to eliminate Δu^l, the generalized velocities $\tilde{u}^l$ can be projected onto the velocity constraint manifold defined by $\dot{g}(q^l, u^l) = 0$ by the following equations [10]:

$$\begin{pmatrix} M & W_E(q^l) \\ W_E^T(q^l) & 0 \end{pmatrix} \begin{pmatrix} \Delta u^l \\ \eta \end{pmatrix} = \begin{pmatrix} 0 \\ -\dot{g}(q^l, \tilde{u}^l) \end{pmatrix} \tag{15}$$

Where η is Lagrange multiplier. Similar to resolution of equations (10), we can find out minimum norm solution of equations (15) which is the best approximate of the exact solutions. To do this, the solution of Δu^l can be obtained:

$$\Delta u^l = \tilde{M}(q^l)\dot{g}(q^l, \tilde{u}^l) \tag{16}$$

Then, $\tilde{u}^l$ can be corrected by the following equations

$$u^l = \tilde{u}^l + \Delta u^l$$

Since the exact solution of position is used for velocity stabilization; the position stabilization must be done in advance of velocity stabilization. Similarly, the generalized position $\tilde{q}^l$ can also be projected onto the position constraint manifold defined by $g(q^l)=0$ by the following equations:

$$\Delta q^l = \tilde{M}(\tilde{q}^l)g(\tilde{q}^l) \tag{17}$$

Equations (17) can be used iteratively until constraint error is less than a user-specified constraint tolerance. In order to improve computer efficiency, only one step of iteration is sufficient for stabilization in application [11]. Δq^l is used to correct $\tilde{q}^l$ by the following equation:

$$q^l = \tilde{q}^l + \Delta q^l$$

Normally, stabilization algorithm can be done after redundant constraints are solved so that the constraint error is avoided into the computation process of contact forces.

6 Numerical Simulation

In order to verify numerical resolution and stabilization methods, as shown in Fig.1, wheels of train moving in the railroad is simulated. Let generalized coordinates and generalized velocities of system be $q=(x_1,y_1,\theta_1,x_2,y_2,\theta_2,x_2,y_2,\theta_2,)$ and $u=(v_{x1},v_{y1},v_{\theta 1},v_{x2},v_{y2},v_{\theta 2},v_{x3},v_{y3},v_{\theta 3})$ respectively. Structural parameters of wheels are $m_1=m_2=m_3=5\text{kg}$; $R_1=R_2=R_3=0.5\text{m}$; $l=3\text{m}$; $l_1=0.4\text{m}$, qualities of other components are ignored. The friction coefficient with railroad: $\mu_1=\mu_2=\mu_2=0.6$. The contact model is built based on above modeling method.

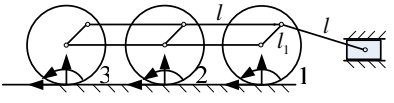


Fig. 1. Wheel model of train

This model is simulated in Matlab where differential algebraic equations are solved by Newton integral algorithm and LCP equations are splitting projected Gauss-Seidel iteration scheme. Simulation step time *h* is set at 0.05s and initial parameters are as follows:

$$q^0=(0,0.5,0,3,0.5,0,6,0.5,0,);\ u^0=(4,0,0,0,0,0,0,0,0,).$$

In Fig.2 and 3, the numerical velocities of wheel 1 and 2 are given respectively. It is known that, from the analytic method, the wheels which are slide should satisfy the momentum balance in the initial motion phase: $m_1 v_{x1}=(m_1+m_2+m_3)v'$. By this formation, we obtain the common velocity of three wheels: $v_{x1}/3$=1.33m/s. Wheels turn rolling from sliding gradually in the following time, due to effects of friction. The

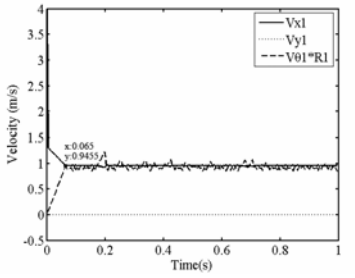


Fig. 2. Numerical velocities of wheel 1

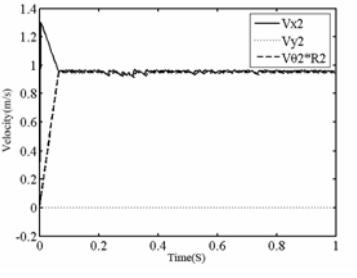


Fig. 3. Numerical velocities of wheel2

linear and angular momentum theorem are satisfied in this process, so we are able to write out and solve linear and angular momentum equations of system, and then obtained the exact transition time is t=0.0648s, after the transition time, the wheel move with a constant velocity: 0.955m/s. from Fig. 2 and 3, it is seen that numerical simulation imitate this process preferably. By the analytic method, the obtained friction forces of wheel: $mg\mu = 29.7\text{N}$, which is consistent with numerical results shown in figure 4. Those contrasts show the validity of numerical methods. But, from figure 4, we can find that numerical simulation causes much error. After stabilization is applied, this error is stabilized and the accuracy of contact force is improved as shown in Fig.5.

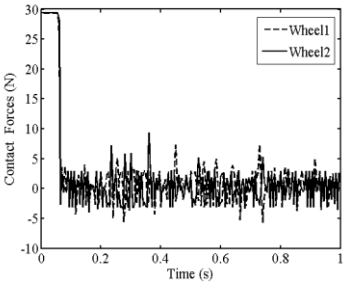


Fig. 4. Numerical contact forces of wheel 1 and 2 without stabilization

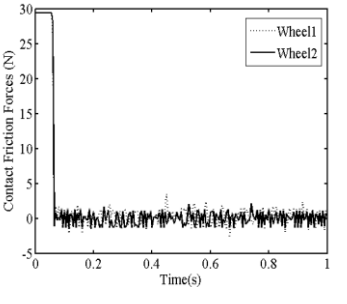


Fig. 5. Numerical contact forces of wheel 1 and 2 after stabilization

7 Conclusion

This paper presents the model and numerical methods for the frictional contact problem of multi-rigid-body with redundant constraints, least square method and linearization method is used for resolution of model and then a direct projection stabilization algorithm is given for reducing numerical errors. These numerical methods are effective in solving frictional contact problem, which can be applied in multi-rigid-body system with not only redundant constraints, but also singular Jacobian matrix. Yet, the

methods also have some shortcoming: such as resolution method of NCP, computational efficiency, finding out the new algorithm for NCP and improving computational efficiency by parallel algorithm is further research content.

Acknowledgments. This work was supported by the Technology Support Foundation of Jiangsu Province (No.BE2008019) and Open Fund Program of the Key State Laboratory (No.MSV-2009-08) and the Hexa-type Elites Peak Program of Jiangsu Province (No.2008144).

References

1. Glocker, C., Pfeiffer, F.: Complementarity Problems in Multibody Systems with Planar Friction. Archive of Applied Mechanics 63(7), 452–463 (1993)
2. Trinkle, J.C., Zeng, D.C.: Prediction of the Quasistatic Planar Motion of a Contacted Rigid Body. IEEE Transactions Robotics Automation 11, 229–246 (1995)
3. Pang, J.S., Trinkle, J.C.: Complementarity Formulation and Existence of Solutions of Dynamic Multi-rigid-body Contact Problem with Coulomb Friction. Mathematical Programming 73(2), 199–266 (1996)
4. Stewart, D.E.: Convergence of a Time-stepping Scheme for Rigid-body Dynamics and Resolution of Painleve's Problem. Archive for Rational Mechanics and Analysis 145(3), 215–260 (1998)
5. Stewart, D.E., Trinkle, J.C.: An Implicit Time-stepping Scheme for Rigid Body Dynamics with Inelastic Coulomb Friction. In: Proceeding of the 2000 IEEE International Conference on Robotics & Automation, vol. 1, pp. 162–169. IEEE Press, San Francisco (2000)
6. Berard, S., Egan, K., Trinkle, J.C.: Contact Modes and Complementary Cones. In: Proceeding of the 2004 IEEE International Conference on Robotics & Automation, vol. 5, pp. 5280–5286. IEEE Press, New Orieans (2004)
7. Forg, M., Pfeiffer, F., Ulbrich, H.: Simulation of Unilateral Constrained Systems with Many Bodies. Multibody System Dynamics 14(2), 137–154 (2005)
8. Zhao, W.J., Pan, Z.K.: Least Square Algorithms and Constraint Stabilization for Euler-Lagrange Equations of Multi-body System Dynamics. Acta Mechanica Sinica 34(2), 594–602 (2002)
9. Glocker, C., Studer, C.: Formulation and Preparation for Numerical Evaluation of Linear Complementarity Systems in Dynamics. Multibody system Dynamics 13(4), 447–463 (2005)
10. Hairer, E., Wanner, G.: Solving ordinary differential equations II: Stiff and differential-algebraic problems, 2nd edn. Springer, Heidelberg (1996)
11. Burgermeister, B., Arnold, M., Esterl, B.: DAE Time Integration for Real-time Applications in Multi-body Dynamics. Journal of Applied Mathematics and Mechanics 86, 759–771 (2006)

Inverse Dynamic Modeling of Two Unsymmetrical 3UPU Parallel Manipulators

Bo Hu[1], Yi Lu[1], and Haixia Mao[2]

[1] College of Mechanical Engineering, Yanshan University
066004, Qinhuangdao, China
hbz0001@yahoo.com.cn
[2] E&A College Of Hebei Normal University of Science & Technology
066004, Qinhuangdao, China

Abstract. The inverse dynamic of two 3UPU parallel manipulators (PM) are analyzed in this paper. The first 3UPU PM has three translational degree of freedom (DOF) and the second 3UPU PM has two translational and one rotational DOF. First, the mobility of the two UPU PMs is analyzed. Second, the constrained forces/torques in UPU-type legs are determined, and the 6×6 Jacobian matrices of the two 3UPU PMs are derived by considering the constrained forces/torques. Third, based on the principle of virtual work, the dynamic is solved with the active forces and the constrained forces/torques derived. Finally, an analytic solved example for the 3UPU PM is given.

Keywords: Parallel manipulator, kinematics, Dynamic.

1 Introduction

Parallel manipulators (PMs) have drawn continuous interest in both academia and industries in recent years [1]. In the kinematics and Dynamic aspect for PMs, Dasgupta [2] proposed Newton-Euler formulation approach for inverse Dynamic. Tsai [3,4] presented a methodology for the Jacobian analysis for limited DOF PMs and solved inverse Dynamic of Stewart-Gough manipulator by using the principle of virtue work. Wang and Gosselin [5] proposed a new approach for the dynamic analysis of PMs. Rico [6,7] studied the kinematic and Dynamic of PMs systemically by using screw theory. Di Gregorio [8,9] studied the Dynamic of some PMs by using Lagrangian equation. Sokolov [10] studied Dynamic analysis of a 3-DOF PM with R–P–S joint structure. Zhang [11] established Kinetostattic modeling of N-DoF PMs with passive constraining leg and prismatic actuators. Lu and Hu [12-14] investigated the kinematics for limited DOF PMs based on geometrical constraint. They derived general simple 6×6 Jacobian, 6×6×6 Hessian matrix for limited DOF PMs with linear active legs and analyzed the kinematic and statics by using CAD variation geometry approach. Hamid and et al. [15] studied the dynamic of a macro-micro redundantly actuated PM.The Dynamic formulae in existing reference only solve the active forces/torques applied on actuators, while the constrained forces/torques existed in subchains are not solved. In general case, the constrained forces/torques exist in limit-DOF PMs, and they should be solved since they greatly affect the accuracy and

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 580–591, 2009.

stiffness of PMs. This paper focuses on solving dynamic of two 3UPU PMs, which have constrained forces/torques in their subchains.

2 Two 3UPU PMs and Their Mobility

2.1 The Geometrical Constraints of Two 3UPU PMs

The 3UPU PMs(see Fig.1 and Fig.2) in this paper have a moving platform, a base B, and three active UPU (universal joint-prismatic joint-universal joint) legs r_i (i=1, 2, 3), Where m is a regular triangle with o as its center and three vertices a_i (i=1, 2, 3). B is a regular triangle with O as its center and three vertices A_i (i=1, 2, 3). Each of UPU legs includes a universal joint U at A_i, one prismatic joint along r_i and one U at a_i. U includes two crossed revolute joints. Let R_{ji} (i=1, 2, 3; j=1, 2, 3, 4) be the j-th revolute joint from B to m in i-th leg. Let $\perp$ and $\parallel$ be perpendicular and parallel constraints, respectively. Then some geometric constraints are satisfied in the first and the third UPU-type limbs for the two PMs.

$$R_{1i} \quad R_{2i}, R_{3i} \quad R_{4i}, R_{2i} \quad r_i, R_{i2} \parallel R_{3i}, R_{1i} \quad B, R_{4i} \quad m \quad (i=1,3) \tag{1}$$

When set the structure constraints in the second limb(R_{12} points to O and R_{42} points to o), a 3UPU PM with three translation DOFs is created (see Fig.1).

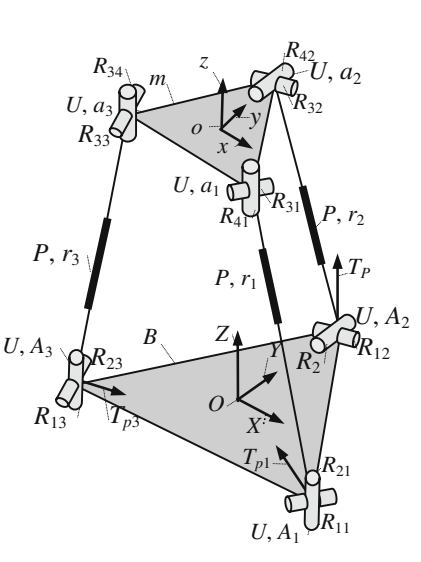


Fig. 1. Sketch of a unsymmetrical pure translational 3UPU parallel manipulator

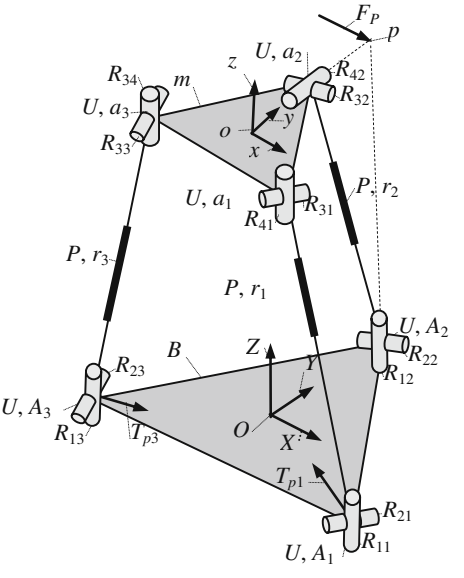


Fig. 2. Sketch of a unsymmetrical 3UPU PM with two translational and one rotational DOF

When set the structure constraints in the second limb(R_{12} B and R_{42} points to o), a 3UPU PM with two translation and one rotational DOFs is created (see Fig.2)

R_{21} and R_{23} lie in B, R_{41} and R_{43} lie in m.By combining with Eq.(1), we obtain

$$B \parallel m \tag{2}$$

Let $\boldsymbol{R}_{ji}(j=1,2,3,4;i=1,2,3)$ be the unit vector of R_{ji}. From Eq.(1), R_{1i}, R_{4i} and r_i are coplanar can be determined, lead to

$$(\boldsymbol{R}_{1i} \times \boldsymbol{R}_{4i}) \cdot \boldsymbol{r}_i = 0 \tag{3}$$

2.2 The Mobility Analysis of the Pure Translational 3UPU PM

In this section,the mobility of pure translational 3UPU PM (see Fig.1) is analyzed**.**For this PM, R_{1i}, R_{4i} and r_i are coplanar, and R_{21} and R_{24} lie in B and m respectively. From Eq.(2), we obtain

$$R_{12} \parallel R_{42} \tag{4}$$

Based on observational methodology used in[14], three constraint torques $\boldsymbol{T}_{pi}$ (i=1, 2, 3) can be found in the three limbs. $\boldsymbol{T}_{pi}$ satisfy: $\boldsymbol{T}_{pi} \perp \boldsymbol{R}_{1i}$, $\boldsymbol{T}_{pi} \perp \boldsymbol{R}_{2i}$, $\boldsymbol{T}_{pi} \perp \boldsymbol{R}_{3i}$, $\boldsymbol{T}_{pi} \perp \boldsymbol{R}_{4i}$. The direction unit vector $\boldsymbol{\tau}_i$ of T_{pi} can be determined by the following equation.

$$\boldsymbol{\tau}_i = \boldsymbol{R}_{1i} \times \boldsymbol{R}_{2i} \tag{5}$$

Let $\boldsymbol{v}$ and $\boldsymbol{\omega}$ be the linear and angular velocity of m. Since $\boldsymbol{T}_{pi}$ (i=1, 2, 3) do no work during the movement of moving platform, there must be

$$T_{pi} \boldsymbol{\tau}_i \cdot \boldsymbol{\omega} = 0 \tag{6a}$$

T_{pi} can be deleted from Eq. (6a), lead to

$$\boldsymbol{\tau}_i \cdot \boldsymbol{\omega} = 0 \tag{6b}$$

Three items of Eq. (6b) can be expressed in matrix form as follow

$$\begin{bmatrix} \boldsymbol{\tau}_1^T \\ \boldsymbol{\tau}_2^T \\ \boldsymbol{\tau}_3^T \end{bmatrix} \cdot \boldsymbol{\omega} = \boldsymbol{0} \tag{6c}$$

$\boldsymbol{\tau}_1$, $\boldsymbol{\tau}_2$ and $\boldsymbol{\tau}_3$ are linearly independent, so the linear equations about $\boldsymbol{\omega}$ must have only trivial solution $\boldsymbol{\omega=0}$, and thus the manipulator has three pure translational DOF.

2.3 The Mobility Analysis of 3UPU PM with Two Translations and One Rotations

In this section,the mobility of 3UPU PM (see Fig.2) with two translational and one rotational DOFs is analyzed**.** Based on observational methodology, two constrained torques $\boldsymbol{T}_{pi}$ (i=1, 3) can be found in the first and third limbs. Eq.(6b) is also appropriate for the two limbs for this PM.

In the second limb, R_{12} and R_{42} are not parallel. From Eq.(3), R_{12} and R_{42} must intersect at one point p. Based on observational methodology, one constrained force $\boldsymbol{F}_p$

which gets across p and parallel with R_{22} can be determined. Let $\boldsymbol{f}$ be the unit vector of $\boldsymbol{F}_p$. As $\boldsymbol{F}_p$ do no work to point o, lead to

$$F_p \boldsymbol{f} \cdot \boldsymbol{v} + (\boldsymbol{d} \times F_p \boldsymbol{f}) \cdot \boldsymbol{\omega} = 0 \,, \boldsymbol{f} = \boldsymbol{R}_{22}, \boldsymbol{d} = \boldsymbol{p} - \boldsymbol{o} \tag{7a}$$

F_p can be deleted from Eq. (7a), lead to

$$\boldsymbol{f} \cdot \boldsymbol{v} + (\boldsymbol{d} \times \boldsymbol{f}) \cdot \boldsymbol{\omega} = 0 \tag{7b}$$

From Eqs.(6b) and (7b), lead to

$$\begin{bmatrix} \boldsymbol{0}_{1\times3} & \boldsymbol{\tau}_1^T \\ \boldsymbol{R}_{22}^T & (\boldsymbol{d} \times \boldsymbol{R}_{22})^T \\ \boldsymbol{0}_{1\times3} & \boldsymbol{\tau}_3^T \end{bmatrix} \boldsymbol{V} = \begin{bmatrix} \boldsymbol{v} \\ \boldsymbol{\omega} \end{bmatrix} = \boldsymbol{0}_{3\times1} \tag{7c}$$

From Eq.(7c), we can see that the number of velocity constrained equation spaces is three. Since $\boldsymbol{\omega} \quad \boldsymbol{\tau}_1$, $\boldsymbol{\omega} \quad \boldsymbol{\tau}_3$, this PM have one rotational and two translational DOF.

3 Two 3UPU PMs and Their Mobility

3.1 Inverse Displacement of the Two 3UPU PMs

For analyzing the kinematics and Dynamic of the two 3UPU PMs, $\{B\}$ is designated as a coordinate O-XYZ with O as its origin fixed on B at O. $\{m\}$ is designated as a coordinate o-xyz with o as its origin fixed on m at o, and some constraints: $X \| A_1A_3$, $Y \perp A_1A_3$, $Z \perp B$, $x \| a_1a_3$, $y \perp a_1a_3$, $z \perp m$ are satisfied (see Fig.1 and Fig.2). The coordinates A_i (i=1, 2, 3) in O-XYZ can be expressed in matrix as follows

$$\boldsymbol{A}_1 = \begin{bmatrix} X_{A1} \\ Y_{A1} \\ Z_{A1} \end{bmatrix} = \begin{bmatrix} aE/2 \\ -E/2 \\ 0 \end{bmatrix}, \boldsymbol{A}_2 = \begin{bmatrix} X_{A2} \\ Y_{A2} \\ Z_{A2} \end{bmatrix} = \begin{bmatrix} 0 \\ E \\ 0 \end{bmatrix}, \boldsymbol{A}_3 = \begin{bmatrix} X_{A3} \\ Y_{A3} \\ Z_{A3} \end{bmatrix} = \begin{bmatrix} -aE/2 \\ -E/2 \\ 0 \end{bmatrix} \tag{8a}$$

The coordinates a_i (i=1, 2, 3) in o-xyz can be expressed in matrix as follows

$${}^m\boldsymbol{a}_1 = \begin{bmatrix} x_{a1} \\ y_{a1} \\ z_{a1} \end{bmatrix} = \begin{bmatrix} ae/2 \\ -e/2 \\ 0 \end{bmatrix}, {}^m\boldsymbol{a}_2 = \begin{bmatrix} x_{a2} \\ y_{a2} \\ z_{a2} \end{bmatrix} = \begin{bmatrix} 0 \\ e \\ 0 \end{bmatrix}, {}^m\boldsymbol{a}_3 = \begin{bmatrix} x_{a3} \\ y_{a3} \\ z_{a3} \end{bmatrix} = \begin{bmatrix} -ae/2 \\ -e/2 \\ 0 \end{bmatrix} \tag{8b}$$

The coordinates a_i (i=1, 2, 3) in O-XYZ can be expressed in matrix as follows

$$\boldsymbol{a_i} = \begin{bmatrix} X_{ai} \\ Y_{ai} \\ Z_{ai} \end{bmatrix} = {}^B_m\mathbf{R} \begin{bmatrix} x_{ai} \\ y_{ai} \\ z_{ai} \end{bmatrix} + \boldsymbol{O}\, {}^B_m, {}^B_m\mathbf{R} = \begin{bmatrix} x_l & y_l & z_l \\ x_m & y_m & z_m \\ x_n & y_n & z_n \end{bmatrix}, \boldsymbol{O} = \begin{bmatrix} X_o \\ Y_o \\ Z_o \end{bmatrix} \tag{8c}$$

Where X_o, Y_o, Z_o are the position of the center of m, ${}^B_m\mathbf{R}$ is a rotation transformation matrix from $\{m\}$ to $\{B\}$. The inverse kinematics can be expressed as

$$r_i = \left| \boldsymbol{a}_i - \boldsymbol{A}_i \right| \qquad (i=1, 2, 3) \tag{9}$$

3.1.1 Inverse Kinematic of 3UPU PM with Three Translational DOF

The rotation transformation matrix of 3UPU PM with three translations can be expressed as

$$ {}^{B}_{m}\mathbf{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{10} $$

From Eqs. (8a) to (10), the inverse displacement of this PM can be derived as

$$ r_1^2 = E^2 + e^2 + X_o^2 + Y_o^2 + Z_o^2 - a(E-e)X_o + (E-e)Y_o - 2eE \tag{11a} $$

$$ r_2^2 = E^2 + e^2 + X_o^2 + Y_o^2 + Z_o^2 - 2(E-e)Y_o - 2eE \tag{11b} $$

$$ r_3^2 = E^2 + e^2 + X_o^2 + Y_o^2 + Z_o^2 + a(E-e)X_o + (E-e)Y_o - 2eE \tag{11c} $$

3.1.2 Inverse Kinematic of 3UPU PM with Two Translations and One Rotations

Let s_α=sinα, c_α=cosα,t_α=$tg\alpha$.The rotation transformation matrix for this 3UPU PM with two translational and one rotational DOF can be expressed as

$$ {}^{B}_{m}\mathbf{R} = \begin{bmatrix} c_\alpha & -s_\alpha & 0 \\ s_\alpha & c_\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{12} $$

Where α denotes the angle of m rotate about Z axis.From Eq.(3),it leads to

$$ \begin{vmatrix} y_l & y_m & y_n \\ ey_l + X_o & ey_m + Y_o - E & ey_n + Z_o \\ 0 & 0 & 1 \end{vmatrix} = 0 \tag{13a} $$

From Eq.(13a),we obtain

$$ y_l(Y_o - E) - X_o y_m = 0 \tag{13b} $$

From Eqs. (12) and (13b),we obtain

$$ X_o = y_l(Y_o - E)/\, y_m = tg_\alpha \cdot (E - Y_o) \tag{14} $$

From Eq. (8a) to (8c),(12) and (14), lead to

$$ \begin{aligned} r_1^2 &= (E-Y_o)^2 t_\alpha^2 + Y_o^2 + Z_o^2 + E^2 + e^2 - aE(E-Y_o)t_\alpha \\ &+ EY_o + aeEs_\alpha - e(Y_o / c_\alpha - Et_\alpha s_\alpha) - 2eEc_\alpha \end{aligned} \tag{15a} $$

$$ r_2^2 = (E-Y_o)^2 t_\alpha^2 + Y_o^2 + Z_o^2 + E^2 + e^2 + 2e(Y_o / c_\alpha - Et_\alpha s_\alpha) - 2E(ec_\alpha + Y_o) \tag{15b} $$

$$ \begin{aligned} r_3^2 &= (E-Y_o)^2 t_\alpha^2 + Y_o^2 + Z_o^2 + E^2 + e^2 + aE(E-Y_o)t_\alpha \\ &+ EY_o - aeEs_\alpha - e(Y_o / c_\alpha - Et_\alpha s_\alpha) - 2eEc_\alpha \end{aligned} \tag{15c} $$

From Eqs.(15a) to (15c),the inverse kinematics of 3UPU PM with two translational and one rotational DOF can be solved

3.2 Solving the 6×6 Inverse Jacobian Matrix

In the two 3UPU PMs, the loop equation of OA_ia_io can be expressed as

$$\boldsymbol{OA}_i + \boldsymbol{A}_i\boldsymbol{a}_i = \boldsymbol{Oo} + \boldsymbol{oa}_i \tag{16}$$

Differentiating both sides of Eq. (16) with respect to time, lead to

$$v_{ri}\boldsymbol{\delta}_i + \boldsymbol{\omega}_{ri}\times r_i\boldsymbol{\delta}_i = \boldsymbol{v} + \boldsymbol{\omega}\times\boldsymbol{e}_i \tag{17}$$

Dot multiplying both side of Eq.(17) by $\boldsymbol{\delta}_i$, lead to

$$v_{ri} = \boldsymbol{v}\cdot\boldsymbol{\delta}_i + (\boldsymbol{\omega}\times e_i)\cdot\boldsymbol{\delta}_i = \begin{bmatrix}\boldsymbol{\delta}_i^T & (\boldsymbol{e}_i\times\boldsymbol{\delta}_i)^T\end{bmatrix}\boldsymbol{V} \tag{18}$$

Eqs.(16) to (18) are fit for the two 3UPU PMs.

3.2.1 Inverse Jacobian for the 3UPU PM with Three Translations

Eq. (6b) can be expressed as

$$\begin{bmatrix}\boldsymbol{0}_{1\times3} & \boldsymbol{\tau}_i^T\end{bmatrix}\boldsymbol{V} = 0 \tag{19}$$

For the 3UPU PM with three translational DOF, there must be

$$\boldsymbol{R}_{11} = \boldsymbol{R}_{41} = \boldsymbol{R}_{13} = \boldsymbol{R}_{43} = \begin{bmatrix}0\\0\\1\end{bmatrix},\ \boldsymbol{R}_{12} = \boldsymbol{R}_{42} = \begin{bmatrix}0\\1\\0\end{bmatrix}, \boldsymbol{R}_{2i} = \boldsymbol{R}_{3i} = \frac{\boldsymbol{R}_{1i}\times\boldsymbol{\delta}_i}{\left|\boldsymbol{R}_{1i}\times\boldsymbol{\delta}_i\right|}$$

From Eqs. (18) and (19), we can obtain

$$\boldsymbol{V}_r = \boldsymbol{J}_{6\times6}\boldsymbol{V}, \boldsymbol{J}_{6\times6} = \begin{bmatrix}\boldsymbol{\delta}_1^T & (\boldsymbol{e}_1\times\boldsymbol{\delta}_1)^T\\ \boldsymbol{\delta}_2^T & (\boldsymbol{e}_2\times\boldsymbol{\delta}_2)^T\\ \boldsymbol{\delta}_3^T & (\boldsymbol{e}_3\times\boldsymbol{\delta}_3)^T\\ \boldsymbol{0}_{1\times3} & \boldsymbol{\tau}_1^T\\ \boldsymbol{0}_{1\times3} & \boldsymbol{\tau}_2^T\\ \boldsymbol{0}_{1\times3} & \boldsymbol{\tau}_3^T\end{bmatrix}, \boldsymbol{V}_r = \begin{bmatrix}v_{r1}\\v_{r2}\\v_{r3}\\0\\0\\0\end{bmatrix} \tag{20}$$

Where, $\boldsymbol{J}_{6\times6}$ is the 6×6 Jacobian for this pure translational 3UPU PM.

3.2.2 Inverse Jacobian for 3UPU PM with Two Translations and One Rotations

For the 3UPU PM with two translational and one rotational DOF, there must be

$$\boldsymbol{R}_{11} = \boldsymbol{R}_{41} = \boldsymbol{R}_{13} = \boldsymbol{R}_{43} = \boldsymbol{R}_{12} = \begin{bmatrix}0\\0\\1\end{bmatrix}, \boldsymbol{R}_{42} = \begin{bmatrix}0\\1\\0\end{bmatrix}, \boldsymbol{R}_{2i} = \boldsymbol{R}_{3i} = \frac{\boldsymbol{R}_{1i}\times\boldsymbol{\delta}_i}{\left|\boldsymbol{R}_{1i}\times\boldsymbol{\delta}_i\right|}$$

From Eqs. (18) and (7c), we obtain

$$\boldsymbol{V}_r = \boldsymbol{J}_{6\times6}\boldsymbol{V}\,,\ \boldsymbol{J}_{6\times6} = \begin{bmatrix} \boldsymbol{\delta}_1^T & (\boldsymbol{e}_1\times\boldsymbol{\delta}_1)^T \\ \boldsymbol{\delta}_2^T & (\boldsymbol{e}_2\times\boldsymbol{\delta}_2)^T \\ \boldsymbol{\delta}_3^T & (\boldsymbol{e}_3\times\boldsymbol{\delta}_3)^T \\ \boldsymbol{0}_{1\times3} & \boldsymbol{\tau}_1^T \\ \boldsymbol{R}_{22}^T & (\boldsymbol{d}\times\boldsymbol{R}_{22})^T \\ \boldsymbol{0}_{1\times3} & \boldsymbol{\tau}_3^T \end{bmatrix} \tag{21}$$

Where $\boldsymbol{J}_{6\times6}$ is the Jacobian for 3UPU PM with two translations and one rotations.

3.3 Inverse Acceleration of Two 3-UPU PMs

Suppose there are two vectors $\boldsymbol{\eta}$ and $\boldsymbol{\varsigma}$. The skew-symmetric matrix of $\boldsymbol{\eta}$ must satisfy following relevant equations [1],

$$\boldsymbol{\eta}=\begin{bmatrix}\eta_x\\ \eta_y\\ \eta_z\end{bmatrix},\boldsymbol{\varsigma}=\begin{bmatrix}\varsigma_x\\ \varsigma_y\\ \varsigma_z\end{bmatrix},\hat{\boldsymbol{\eta}}=\begin{bmatrix}0 & -\eta_z & \eta_y\\ \eta_z & 0 & -\eta_x\\ -\eta_y & \eta_x & 0\end{bmatrix},\ \boldsymbol{\eta}\times\boldsymbol{\varsigma}=\hat{\boldsymbol{\eta}}\boldsymbol{\varsigma},\hat{\boldsymbol{\eta}}^T=-\hat{\boldsymbol{\eta}},-\hat{\boldsymbol{\eta}}^2+\boldsymbol{\eta}\boldsymbol{\eta}^T=\boldsymbol{E}_{3\times3} \tag{22}$$

Where, $\boldsymbol{\eta}$ and $\boldsymbol{\varsigma}$ may be one of $\boldsymbol{\delta}_i$, $\boldsymbol{e}_i$, $\boldsymbol{d}$ and $\boldsymbol{f}_i$, (i=1, 2, 3), $\boldsymbol{E}_{3\times3}$ is the unit matrix.

Let a_{ri}(i=1,2,3) be the acceleration of i-th leg. Let $\boldsymbol{A}$ be the six-dimensional accelera tion vector of m. Differentiating both sides of Eq. (18) and from reference[14],lead to

$$a_{ri}=\begin{bmatrix}\boldsymbol{\delta}_i^T & (\boldsymbol{e}_i\times\boldsymbol{\delta}_i)^T\end{bmatrix}\boldsymbol{A}+\boldsymbol{V}^T\boldsymbol{H}_i\boldsymbol{V},\ \boldsymbol{H}_i=\frac{1}{r_i}\begin{bmatrix}-\hat{\boldsymbol{\delta}}_i^2 & \hat{\boldsymbol{\delta}}_i^2\hat{\boldsymbol{e}}_i\\ -\hat{\boldsymbol{e}}_i\hat{\boldsymbol{\delta}}_i^2 & r_i\hat{\boldsymbol{e}}_i\hat{\boldsymbol{\delta}}_i+\hat{\boldsymbol{e}}_i\hat{\boldsymbol{\delta}}_i^2\hat{\boldsymbol{e}}_i\end{bmatrix}_{6\times6} \tag{23a}$$

Eqs.(23a) can be expressed in matrix form as following

$$\boldsymbol{a}_r=\boldsymbol{J}_\alpha\boldsymbol{A}+\boldsymbol{V}^T\boldsymbol{H}_i\boldsymbol{V},\boldsymbol{a}_r=\begin{bmatrix}a_{r1}\\ a_{r2}\\ a_{r3}\end{bmatrix},\boldsymbol{J}_\alpha=\begin{bmatrix}\boldsymbol{\delta}_1^T & (\boldsymbol{e}_1\times\boldsymbol{\delta}_1)^T\\ \boldsymbol{\delta}_2^T & (\boldsymbol{e}_2\times\boldsymbol{\delta}_2)^T\\ \boldsymbol{\delta}_3^T & (\boldsymbol{e}_3\times\boldsymbol{\delta}_3)^T\end{bmatrix},\boldsymbol{H}=\begin{bmatrix}\boldsymbol{H}_1\\ \boldsymbol{H}_2\\ \boldsymbol{H}_3\end{bmatrix} \tag{23b}$$

Eqs.(22) to (23b) is fit for two 3UPU PMs.

4 Dynamic Modeling

In this part, the dynamic of two 3UPU PMs is solved in a uniform form. The following formulae in this part are fit for the two 3UPU PMs.

Let $\dot{\theta}_{1i}$ and $\dot{\theta}_{2i}$ be the velocity of the first and the second revolute joint in i-th (i=1, 2, 3) limb respectively. The angular velocity $\boldsymbol{\omega}_{ri}$ of each limb r_i can be expressed as

$$\boldsymbol{\omega}_{ri}=\dot{\theta}_{1i}\boldsymbol{R}_{1i}+\dot{\theta}_{2i}\boldsymbol{R}_{2i} \tag{24a}$$

Dot multiplying both side of Eq.(24a) by $\boldsymbol{A}_i\boldsymbol{a}_i$ and combining with Eqs.(17) and (22), lead to

$$\begin{aligned}&\dot{\theta}_{1i}\boldsymbol{R}_{1i}\times\boldsymbol{A}_i\boldsymbol{a}_i+\dot{\theta}_{2i}\boldsymbol{R}_{2i}\times\boldsymbol{A}_i\boldsymbol{a}_i=\boldsymbol{\omega}_{ri}\times\boldsymbol{A}_i\boldsymbol{a}_i=\boldsymbol{v}+\boldsymbol{\omega}\times\boldsymbol{e}_i-v_{ri}\boldsymbol{\delta}_i\\&=\boldsymbol{v}-\boldsymbol{e}_i\times\boldsymbol{\omega}-(\boldsymbol{v}\cdot\boldsymbol{\delta}_i)\boldsymbol{\delta}_i=\boldsymbol{v}-\boldsymbol{\delta}_i(\boldsymbol{\delta}_i^T\boldsymbol{v})-\hat{\boldsymbol{e}}_i\boldsymbol{\omega}=-\hat{\boldsymbol{\delta}}_i^2\boldsymbol{v}-\hat{\boldsymbol{e}}_i\boldsymbol{\omega}\end{aligned}\tag{24b}$$

Dot multiplying both sides of Eq. (24b) by $\boldsymbol{R}_{2i}$, lead to

$$\dot{\theta}_{1i}(\boldsymbol{R}_{1i}\times\boldsymbol{A}_i\boldsymbol{a}_i)\cdot\boldsymbol{R}_{2i}=-\boldsymbol{R}_{2i}^T(\hat{\boldsymbol{\delta}}_i^2\boldsymbol{v}+\hat{\boldsymbol{e}}_i\boldsymbol{\omega})\tag{25a}$$

Dot multiplying both side of Eq.(24b) by $\boldsymbol{R}_{1i}$ and combining with Eq (22), lead to

$$\dot{\theta}_{2i}(\boldsymbol{R}_{2i}\times\boldsymbol{A}_i\boldsymbol{a}_i)\cdot\boldsymbol{R}_{1i}=-\boldsymbol{R}_{1i}^T(\hat{\boldsymbol{\delta}}_i^2\boldsymbol{v}+\hat{\boldsymbol{e}}_i\boldsymbol{\omega})\tag{25b}$$

Since R_{2i} A_ia_i R_{2i} R_{1i}, the following equation can be obtained

$$(\boldsymbol{R}_{1i}\times\boldsymbol{A}_i\boldsymbol{a}_i)\cdot\boldsymbol{R}_{2i}=r_i\tag{26a}$$

$$(\boldsymbol{R}_{2i}\times\boldsymbol{A}_i\boldsymbol{a}_i)\cdot\boldsymbol{R}_{1i}=-r_i\tag{26b}$$

From Eqs.(24a), (25a), (25b), (26a) and (26b), we obtain

$$\dot{\theta}_{1i}=\left[-\boldsymbol{R}_{2i}^T\hat{\boldsymbol{\delta}}_i^2/r_i\quad-\boldsymbol{R}_{2i}^T\hat{\boldsymbol{e}}_i/r_i\right]\boldsymbol{V}\tag{27a}$$

$$\dot{\theta}_{2i}=\left[\boldsymbol{R}_{1i}^T\hat{\boldsymbol{\delta}}_i^2/r_i\quad\boldsymbol{R}_{1i}^T\hat{\boldsymbol{e}}_i/r_i\right]\boldsymbol{V}\tag{27b}$$

$$r_i\dot{\theta}_{1i}=(\boldsymbol{v}-\boldsymbol{e}_i\times\boldsymbol{\omega})\cdot\boldsymbol{R}_{2i}\tag{28a}$$

$$r_i\dot{\theta}_{2i}=-(\boldsymbol{v}-\boldsymbol{e}_i\times\boldsymbol{\omega}-v_{ri}\boldsymbol{\delta}_i)\cdot\boldsymbol{R}_{1i}\tag{28b}$$

As R_{1i} is fixed on B and R_{2i} rotates with R_{1i}, we obtain

$$\dot{\boldsymbol{R}}_{1i}=\boldsymbol{0}_{3\times1},\dot{\boldsymbol{R}}_{2i}=\boldsymbol{R}_{1i}\dot{\theta}_{1i}\tag{29}$$

Differentiating both sides of Eq. (28a) and (28b) with respect to time, lead to

$$\begin{aligned}&r_i\ddot{\theta}_{1i}+v_{ri}\dot{\theta}_{1i}=\left[\boldsymbol{a}-(\boldsymbol{\omega}\times\boldsymbol{e}_i)\times\boldsymbol{\omega}-\boldsymbol{e}_i\times\boldsymbol{\varepsilon}\right]\cdot\boldsymbol{R}_{2i}+(\boldsymbol{v}-\boldsymbol{e}_i\times\boldsymbol{\omega})\cdot(\boldsymbol{\omega}_{R1i}\times\boldsymbol{R}_{2i})\\&=\left[\boldsymbol{a}-(\boldsymbol{\omega}\times\boldsymbol{e}_i)\times\boldsymbol{\omega}-\boldsymbol{e}_i\times\boldsymbol{\varepsilon}\right]\cdot\boldsymbol{R}_{2i}+(\boldsymbol{v}-\boldsymbol{e}_i\times\boldsymbol{\omega})\cdot(\dot{\theta}_{1i}\boldsymbol{R}_{1i}\times\boldsymbol{R}_{2i})\end{aligned}\tag{30a}$$

$$r_i\ddot{\theta}_{2i}+v_{ri}\dot{\theta}_{2i}=-\left[\boldsymbol{a}-(\boldsymbol{\omega}\times\boldsymbol{e}_i)\times\boldsymbol{\omega}-\boldsymbol{e}_i\times\boldsymbol{\varepsilon}-a_{ri}\boldsymbol{\delta}_i-v_{ri}(\boldsymbol{\omega}_{ri}\times\boldsymbol{\delta}_i)\right]\cdot\boldsymbol{R}_{1i}\tag{30b}$$

$$\ddot{\theta}_{1i}=\{\left[\boldsymbol{a}-(\boldsymbol{\omega}\times\boldsymbol{e}_i)\times\boldsymbol{\omega}-\boldsymbol{e}_i\times\boldsymbol{\varepsilon}\right]\cdot\boldsymbol{R}_{2i}+(\boldsymbol{v}-\boldsymbol{e}_i\times\boldsymbol{\omega})\cdot(\dot{\theta}_{1i}\boldsymbol{R}_{1i}\times\boldsymbol{R}_{2i})-v_{ri}\dot{\theta}_{1i}\}/r_i\tag{31a}$$

$$\ddot{\theta}_{2i}=\{-\left[\boldsymbol{a}-(\boldsymbol{\omega}\times\boldsymbol{e}_i)\times\boldsymbol{\omega}-\boldsymbol{e}_i\times\boldsymbol{\varepsilon}-a_{ri}\boldsymbol{\delta}_i-v_{ri}(\boldsymbol{\omega}_{ri}\times\boldsymbol{\delta}_i)\right]-v_{ri}\dot{\theta}_{2i}\}/r_i\tag{32b}$$

Substituting Eqs. (27a) and (27b) into Eq (24a), $\boldsymbol{\omega}_{ri}$ can be expressed as

$$\boldsymbol{\omega}_{ri}=\dot{\theta}_{1i}\boldsymbol{R}_{1i}+\dot{\theta}_{2i}\boldsymbol{R}_{2i}=\left[(\boldsymbol{R}_{2i}\boldsymbol{R}_{1i}^T-\boldsymbol{R}_{1i}\boldsymbol{R}_{2i}^T)\hat{\boldsymbol{\delta}}_i^2\boldsymbol{R}_{1i}^T\hat{\boldsymbol{e}}_i/r_i\quad(\boldsymbol{R}_{2i}\boldsymbol{R}_{1i}^T-\boldsymbol{R}_{1i}\boldsymbol{R}_{2i}^T)\hat{\boldsymbol{e}}_i\boldsymbol{R}_{1i}^T\hat{\boldsymbol{e}}_i/r_i\right]\boldsymbol{V}\tag{33}$$

4.1 The Velocity and Acceleration of the Cylinder

Each of linear active legs r_i (i=1,2,3) is composed of a piston and a cylinder. The piston is connected with m at a_i and the cylinder is connected with B at A_i.

Let r_{si} be the distance from the mass center of the cylinder to the point A_i, r_i be the length of cylinder with r_{si}=r_i/2, respectively.The angular velocity of i-th cylinder $\boldsymbol{\omega}_{si}$ can be obtained as

$$\boldsymbol{\omega}_{si} = \boldsymbol{\omega}_{ri} \tag{34a}$$

The linear velocity of the mass center of *i*-th cylinder $\boldsymbol{v}_{si}$ can be expressed as

$$\boldsymbol{v}_{si} = r_{si}\boldsymbol{\omega}_{ri} \times \boldsymbol{\delta}_i = \left[-r_{si}\hat{\boldsymbol{\delta}}_i(\boldsymbol{R}_{2i}\boldsymbol{R}_{1i}^T - \boldsymbol{R}_{1i}\boldsymbol{R}_{2i}^T)\hat{\boldsymbol{\delta}}_i^2 / r_i \quad -r_{si}\hat{\boldsymbol{\delta}}_i(\boldsymbol{R}_{2i}\boldsymbol{R}_{1i}^T - \boldsymbol{R}_{1i}\boldsymbol{R}_{2i}^T)\hat{\boldsymbol{e}}_i / r_i \right] \boldsymbol{V} \tag{34b}$$

The velocity mapping relation between the *i*-th cylinder and moving platform can be expressed as following

$$\begin{bmatrix} \boldsymbol{v}_{si} \\ \boldsymbol{\omega}_{si} \end{bmatrix} = \boldsymbol{J}_{si}\boldsymbol{V}, \boldsymbol{J}_{si} = \begin{bmatrix} -r_{si}\hat{\boldsymbol{\delta}}_i(\boldsymbol{R}_{2i}\boldsymbol{R}_{1i}^T - \boldsymbol{R}_{1i}\boldsymbol{R}_{2i}^T)\hat{\boldsymbol{\delta}}_i^2 / r_i & -r_{si}\hat{\boldsymbol{\delta}}_i(\boldsymbol{R}_{2i}\boldsymbol{R}_{1i}^T - \boldsymbol{R}_{1i}\boldsymbol{R}_{2i}^T)\hat{\boldsymbol{e}}_i / r_i \\ (\boldsymbol{R}_{2i}\boldsymbol{R}_{1i}^T - \boldsymbol{R}_{1i}\boldsymbol{R}_{2i}^T)\hat{\boldsymbol{\delta}}_i^2 / r_i & (\boldsymbol{R}_{2i}\boldsymbol{R}_{1i}^T - \boldsymbol{R}_{1i}\boldsymbol{R}_{2i}^T)\hat{\boldsymbol{e}}_i / r_i \end{bmatrix} \tag{35}$$

The angular acceleration $\boldsymbol{\varepsilon}_{si}$ of *i*-th cylinder can be expressed as

$$\boldsymbol{\varepsilon}_{si} = \boldsymbol{\varepsilon}_{ri} = \ddot{\theta}_{1i}\boldsymbol{R}_{i1} + \ddot{\theta}_{2i}\boldsymbol{R}_{i2} + \dot{\theta}_{2i}(\boldsymbol{\omega}_{R2i} \times \boldsymbol{R}_{i2}) = \ddot{\theta}_{1i}\boldsymbol{R}_{i1} + \ddot{\theta}_{2i}\boldsymbol{R}_{i2} + \dot{\theta}_{2i}(\dot{\theta}_{1i}\boldsymbol{R}_{i1} \times \boldsymbol{R}_{i2}) \tag{36a}$$

The linear acceleration $\boldsymbol{a}_{si}$ of the mass center of *i*-th cylinder can be expressed as

$$\boldsymbol{a}_{si} = r_{si}\boldsymbol{\varepsilon}_{ri} \times \boldsymbol{\delta}_i + r_{si}\boldsymbol{\omega}_{ri} \times (\boldsymbol{\omega}_{ri} \times \boldsymbol{\delta}_i) \tag{36b}$$

4.2 The Velocity and Acceleration of the Piston

Let r_{ti} be the distance between the mass center of the piston and the point g_i, r_{si} be the length of piston with $r_{ti}=l_i \ r_{si}$.The angular velocity of *i*-th piston $\boldsymbol{\omega}_{ti}$ is

$$\boldsymbol{\omega}_{ti} = \boldsymbol{\omega}_{ri} \tag{37a}$$

The linear velocity of the mass center of *i*-th piston $\boldsymbol{v}_{ti}$ can be expressed as

$$\boldsymbol{v}_{ti} = r_{ti}\boldsymbol{\omega}_{ri} \times \boldsymbol{\delta}_i + v_{ri}\boldsymbol{\delta}_i \boldsymbol{V} = \left[-r_{ti}\hat{\boldsymbol{\delta}}_i(\boldsymbol{R}_{2i}\boldsymbol{R}_{1i}^T - \boldsymbol{R}_{1i}\boldsymbol{R}_{2i}^T\hat{\boldsymbol{\delta}}_i^2) / r_i + \boldsymbol{\delta}_l \ \boldsymbol{\delta}_i^T \quad \boldsymbol{0}_{3\times3} \right] \boldsymbol{V} \tag{37b}$$

The velocity mapping relation between *i-th* piston and moving platform can be expressed as

$$\begin{bmatrix} \boldsymbol{v}_{ti} \\ \boldsymbol{\omega}_{ti} \end{bmatrix} = \boldsymbol{J}_{ti}\boldsymbol{V}, \boldsymbol{J}_{ti} = \begin{bmatrix} -r_{ti}\hat{\boldsymbol{\delta}}_i(\boldsymbol{R}_{2i}\boldsymbol{R}_{1i}^T - \boldsymbol{R}_{1i}\boldsymbol{R}_{2i}^T\hat{\boldsymbol{\delta}}_i^2) / r_i + \boldsymbol{\delta}_i\boldsymbol{\delta}_i^T & \boldsymbol{0}_{3\times3} \\ (\boldsymbol{R}_{2i}\boldsymbol{R}_{1i}^T - \boldsymbol{R}_{1i}\boldsymbol{R}_{2i}^T\hat{\boldsymbol{\delta}}_i^2) / r_i & \boldsymbol{0}_{3\times3} \end{bmatrix} \tag{38}$$

The angular acceleration $\boldsymbol{\varepsilon}_{ti}$ of *i-th* piston can be expressed as

$$\boldsymbol{\varepsilon}_{ti} = \boldsymbol{\varepsilon}_{ri} \tag{39a}$$

The linear acceleration $\boldsymbol{a}_{ti}$ of the mass center of *i-th* piston can be expressed as

$$\boldsymbol{a}_{ti} = r_{ti}\boldsymbol{\varepsilon}_{ri} \times \boldsymbol{\delta}_i + r_{ti}\boldsymbol{\omega}_{ri} \times (\boldsymbol{\omega}_i \times \boldsymbol{\delta}_i) + 2v_{ri}(\boldsymbol{\omega}_{ri} \times \boldsymbol{\delta}_i) + a_{ri}\boldsymbol{\delta}_i \tag{39b}$$

4.3 The Formula of Dynamic

In Dynamic modeling, the friction along the drives and the inertias of all articulated joints are negligible and all the limbs are assumed to be axially symmetrical.Let $\boldsymbol{F}_o$, $\boldsymbol{T}_o$ be the workload applied on the moving platform, $\boldsymbol{f}_o$, $\boldsymbol{n}_o$ be the inertia forces and

torques, m_o and $\boldsymbol{I}_o$ be the mass and inertia moment of the platform with the end effector about point o and $\boldsymbol{G}_o$ be the gravitational force exerted at the center of mass of moving platform.Let $\boldsymbol{f}_{si}$, $\boldsymbol{n}_{si}$(i=1,2,3) be the inertia forces and torques, m_{si} and $\boldsymbol{I}_{si}$ be the mass and inertia moment of the *i-th* cylinder about its mass center and $\boldsymbol{G}_{si}$ be the gravitational force exerted at the center of mass of *i-th* cylinder. Let $\boldsymbol{f}_{ti}$, $\boldsymbol{n}_{ti}$ be the inertia forces and torques, m_{ti} and $\boldsymbol{I}_{ti}$ be the mass and inertia moment of *i-th* piston about its mass center and $\boldsymbol{G}_{ti}$ be the gravitational force exerted at the center of mass of *i-th* piston. Then we can obtain

$$\begin{aligned}
&\boldsymbol{f}_o = -m_o\boldsymbol{a}_o, \boldsymbol{G}_o = m_o\boldsymbol{g}, \boldsymbol{n}_o = -\boldsymbol{I}_o\boldsymbol{\varepsilon} - \boldsymbol{\omega}\times(I_o\boldsymbol{\omega})\\
&\boldsymbol{f}_{si} = -m_{si}\boldsymbol{a}_{si}, \boldsymbol{G}_{si} = m_{si}\boldsymbol{g}, \boldsymbol{n}_{si} = -\boldsymbol{I}_{si}\boldsymbol{\varepsilon}_{si} - \boldsymbol{\omega}_{si}\times(\boldsymbol{I}_{si}\boldsymbol{\omega}_{si})\\
&\boldsymbol{f}_{ti} = -m_{ti}\boldsymbol{a}_{ti}, \boldsymbol{G}_{ti} = m_{ti}\boldsymbol{g}, \boldsymbol{n}_{ti} = -\boldsymbol{I}_{ti}\boldsymbol{\varepsilon}_{ti} - \boldsymbol{\omega}_{ti}\times(\boldsymbol{I}_{ti}\boldsymbol{\omega}_{ti})
\end{aligned} \tag{40}$$

By means of the principle of virtual work, the following equation can be derived

$$\boldsymbol{F}_q^T\boldsymbol{V}_r + \left[\boldsymbol{F}_o^T + \boldsymbol{f}_o^T + \boldsymbol{G}_o^T \quad \boldsymbol{T}_o^T + \boldsymbol{n}_o^T\right]\boldsymbol{V} + \sum_{i=1}^{i=n}\left(\left[\boldsymbol{f}_{si}^T + \boldsymbol{G}_{si}^T \quad \boldsymbol{n}_{si}^T\right]\begin{bmatrix}\boldsymbol{v}_{si}\\ \boldsymbol{\omega}_{ni}\end{bmatrix} + \left[\boldsymbol{f}_{ti}^T + \boldsymbol{G}_{ti}^T \quad \boldsymbol{n}_{ti}^T\right]\begin{bmatrix}\boldsymbol{v}_{ti}\\ \boldsymbol{\omega}_{ni}\end{bmatrix}\right) = 0 \tag{41a}$$

From Eqs. (20) or (21), (35), (38) and (41a), the formula for solving Dynamic are derived as following

$$\boldsymbol{F}_q = -(\boldsymbol{J}_{6\times6}^{-1})^T\begin{bmatrix}\boldsymbol{F}_o + \boldsymbol{f}_o + \boldsymbol{G}_o\\ \boldsymbol{T}_o + \boldsymbol{n}_o\end{bmatrix} - (\boldsymbol{J}_{6\times6}^{-1})^T\sum_{i=1}^{i=n}\left(\boldsymbol{J}_{si}^T\begin{bmatrix}\boldsymbol{f}_{si} + \boldsymbol{G}_{si}\\ \boldsymbol{n}_{si}\end{bmatrix} + \boldsymbol{J}_{ti}^T\begin{bmatrix}\boldsymbol{f}_{ti} + \boldsymbol{G}_{ti}\\ \boldsymbol{n}_{ti}\end{bmatrix}\right) \tag{41b}$$

Where, $\boldsymbol{F}_q$=[F_{a1} F_{a2} F_{a3} T_{p1} T_{p2} T_{p3}] are active forces and constrained torques for pure translational 3UPU PM and $\boldsymbol{F}_q$=[F_{a1} F_{a2} F_{a3} T_{p1} F_p T_{p3}] are the active forces and constrained forces/torques for two translational and one rotational 3UPU PM.

By using the formula (41b), the active force applied on each drive can be solved, and the constrained forces/torques existing in the two 3UPU PMs can also be derived.

5 Analytic Solved Example

In this part, a numerical example for pure translational 3UPU PM is given to illustrate this method. Set E=120/a, e=60/a, $\boldsymbol{F}_o$=[-20 -30 -60]kN, $\boldsymbol{T}_o$=[-30 -30 100] cm·kN, m_o=10 kg, m_{si}=m_{ti}=10kg, g=-9.8 cm/s^2. Give velocity and acceleration of the moving platform as: a_x=3, a_y=6, a_z=5 mm/s^2, v_x=3t, v_y=6t, v_z=5t cm/s· When the moving platform move at its initial position (0, 0 ,120) according to its velocity, and acceleration, the position (X_o, Y_o, Z_o) of moving platform, the length, velocity, and acceleration of active legs (r_1, r_2, r_3) are solved (see Fig. 3b~3d); the active forces F_{ai} and the constrained torque T_{pi} for pure translational 3UPU PM are solved (see Fig. 3e~f).

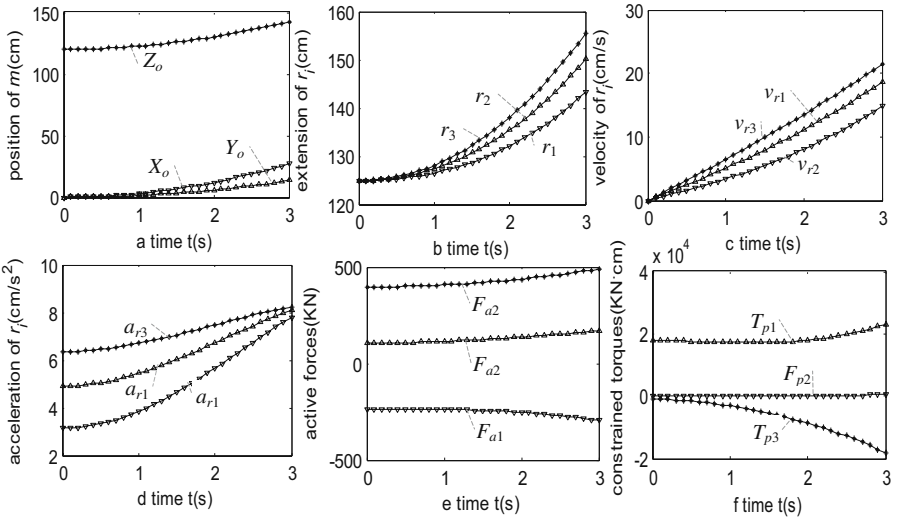


Fig. 3.

6 Conclusions

The kinematical character of a 3UPU PMs is reviewed The two 3UPU PMs analyzed in this paper are identical except the disposal of an R joint in their second UPU limbs. If this R joint points to *O*, the 3UPU PM has three translational DOF. If the R joint is perpendicular with *B*, the 3UPU has two translational and one rotational DOF.

Inverse dynamic formula for the two 3UPU PMs is solved with the active forces and constrained forces/torques derived. In the 3UPU PM with pure translational DOFs, three constraint torques exist in three UPU legs. In the 3UPU PM with two translational and one rotational DOFs, the constrained torques and the constrained force synchronously exist. The active forces and constrained forces/torques are solved in dynamic modeling. a numerical example for pure translational 3UPU PM is given to illustrate this method The two 3UPU PMs have some potential applications for the 3DOF parallel kinematic machines, such as the parallel machine tool, pick-up manipulator for electronic element/food/medicines, package machine, micro surgical manipulator, and so on.

Acknowledgments

This research work is supported by the Natural Sciences Foundation Council of China (NSFC) 50575198, the Doctoral Fund from National Education Ministry No. 20060216006, and the Science Foundation of Yanshan University for the Excellent Ph.D.Students No. YSU200901.

References

1. Merlet, J.-P.: Parallel Robots. Kluwer Academic Publishers, Dordrecht (1999)
2. Dasgupta, B., Mruthyunjaya, T.S.: Closed-Form Dynamic Equations of the General Stewart Platform through the Newton–Euler Approach. Mechanism and Machine Theory 33(7), 993–1012 (1999)
3. Joshi, S.A., Tsai, L.W.: Jacobian Analysis of Limited-DOF Parallel Manipulators. ASME Journal of Mechanical Design 124(2), 254–258 (2002)
4. Tsai, L.W.: Solving the inverse Dynamic of a Stewart-Gough manipulator by the principle of virtual work. Trans. of the ASME, Journal of Mechanical Design 122(1), 3–9 (2000)
5. Wang, J., Gosselin, C.M.: A new approach for the dynamic analysis of parallel manipulators. Multibody System Dynamic 2(3), 317–334 (1998)
6. Rico, J.M., Duffy, J.: Forward and Inverse Acceleration Analyses of In-Parallel Manipulators. Trans. of the ASME, Journal of Mechanical Design 122, 299–303 (2000)
7. Gallardo, J., Rico, J.M., Frisoli, A., Checcacci, D., Bergamasco, M.: Dynamic ofparallel manipulators by means of screw theory. Mechanism and MachineTheory 38(11), 1113–1131
8. Di Gregorio, R., Parenti-Castelli, V.: Dynamic of a class of parallel wrists. Journal of Mechanical Design, Transactions of the ASME 126(3), 436–441 (2004)
9. Di Gregorio, R.: Dynamic model and performances of 2-DOF manipulators. Robotica 24(1), 51–60 (2006)
10. Sokolov, A., Xirouchakis, P.: Dynamic analysis of a 3-DOF parallel manipulator with R–P–S joint structure. Mechanism and Machine Theory 42(5), 541–557 (2007)
11. Zhang, D., Gosselin, C.M.: Kinetostattic modeling of N-DoF parallel mechanisms with passive constraining leg and prismatic actuators. Trans. of the ASME Journal of Mechanical Design 123(3), 375–381 (2001)
12. Yi, L., Bo, H.: Solving driving forces of 2(3-SPR) serial-parallel manipulator by CAD variation geometry approach. Transaction ASME Journal of Mechanical Design 128(6), 1349–1351 (2006)
13. Bo, H., Yi, L.: Analyses of kinematics, statics, and workspace of a 3-RRPRR parallel manipulator and its three isomeric mechanisms. Journal of Mechanical Engineering Science (Proc. IMechE Part C) 222(9), 1829–1837 (2008)
14. Yi, L., Bo, H.: Unification and simplification of velocity/acceleration of limited-dof parallel manipulators with linear active legs. Mechanism and Machine Theory 43(9), 1112–1128 (2008)
15. Taghirad, H.D., Nahon, M.: A Dynamic analysis of a macro-micro redundantly actuated parallel manipulator. Advanced Robotics 22(9), 949–981 (2008)

Kinematic Analysis of the SPKM165, a 5-Axis Serial-Parallel Kinematic Milling Machine

Fugui Xie, Xinjun Liu*, Jinsong Wang, and Liping Wang

Institute of Manufacturing Engineering, Department of Precision Instruments and Mechanology, Tsinghua University, Beijing 100084, China
xinjunliu@mail.tsinghua.edu.cn

Abstract. This paper analyzes the kinematics of a 5-axis serial-parallel kinematic milling machine, i.e., the SPKM 165, which consists of a three-DOF (degree of freedom) parallel module and a two-DOF serial table. In order to achieve 5-axis control, the inverse kinematic equations are derived and corresponding coordinate system is established. The relationship between the NC-code and the operating attitude of the machine is shown and the numerical results are presented.

Keywords: Kinematics, Five-axis mechanism, Parallel kinematics, NC-code.

1 Introduction

Parallel kinematic machines (PKM) have the advantages of high structural rigidity, better payload-weight ratio, high dynamic performances, high velocity and time-saving machining [1-3]. For such reasons, parallel kinematic machines have been studied intensively for more than twenty years and still attract attentions from academe and industry. The most famous parallel mechanism is Stewart platform which has six DOFs (degrees of freedom). And many prototypes and commercial parallel mechanisms with Stewart-Gough parallel kinematic architecture have been developed. However, the Stewart platform's disadvantages of limited workspace and low dexterous manipulation have been exposed during deep research and application [2]. Many industrial applications require less than six DOFs and comparative large workspace, good dexterous manipulation and low cost. The parallel mechanisms with fewer DOFs meet these requirements in some extent and some of them have been developed [4, 5]. Among them, the mechanism with three DOFs is welcome since such a mechanism is easier to be a machining module. As a result, serial-parallel kinematic machines that are usually composed of a 3-DOF parallel mechanism and 2-DOF serial mechanism are becoming more and more popular.

As is well known, the reduction of DOFs can result in coupled motions of the mobile platform, and the kinematic analysis becomes very complicated for some of such kinds of mechanisms, such as the parallel module of the VERNE machine [6]. So, the kinematics of the PKM with fewer DOFs has been investigated by more and more researchers and many of them have focused on the discussion of both analytical and

* Corresponding author.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 592–602, 2009.

numerical methods [7, 8]. And some of them cannot be used in actual application for the complexity of their kinematics.

The machine studied here includes a three-DOF parallel mechanism proposed in [9] by Liu et al., and there are no parasitic motion and no coupling between the rotation and the translation along one direction. What is more, the mobile platform can tilt continuously from -25° to 90°. For such a reason, by using a continuous rotational table in the serial mechanism, the machine is capable of 5-face machining. Thus, this machine has potential application in engineering.

2 Description of the SPKM165 Machine

The SPKM 165 machine shown in Fig. 1 is a 5-axis serial-parallel kinematic milling machine, which consists of a parallel module and a serial table.

Fig. 1. The SPKM165 machine

As shown in Figs. 2 and 3, the moving platform of the parallel module is connected to the base through three legs, namely, legs I, II and III. Legs I and II are in a same plane, i.e., the *o-yz* plane shown in Fig. 3, and have identical chains, each of which consists of a length-fixed link which is connected to the active slider through a passive revolute joint and is connected to the moving platform through a universal joint. Leg III is in the *o-xz* plane, perpendicular to the plane where legs I and II are, and consists of another length-fixed link, which is connected to the active slider through a passive revolute and is attached to the moving platform through a cylinder joint. The three sliders are then linked to the base through three prismatic joints and actuated along the *z*-axis. The constraints and DOFs of the parallel module are shown in Table 1.

Table 1. The constraints and DOFs of the parallel module with prismatic actuators

Single leg			Combination of three legs	
No.	Leg type	Constraints	Constraints	Remained DOFs
I	$P_zR_xU_{xy}$	$\{RO_x, T_x\}$	$\{T_x, RO_x, RO_z\}$	$\{T_y, T_z, RO_y\}$
II	$P_zR_xU_{xy}$	$\{RO_x, T_x\}$		
III	$P_zR_yC_y$	$\{RO_x, RO_z\}$		

where, P-prismatic joint, R-revolute joint, C-cylinder joint, T-translation, RO-rotation, and the subscript stands for the direction of the DOF.

The rotating table IV in the serial table is placed on a slider and rotates about the z-axis, the slider V is connected to the base through a prismatic joint and is actuated along the x-axis.

Then, the parallel module has three spatial DOFs, two translations in the *o-yz* plane and one rotation about the *y*-axis, the serial table has two DOFs, one translation along the x-axis and one rotation about the z-axis. All of the five DOFs are shown in Fig. 3.

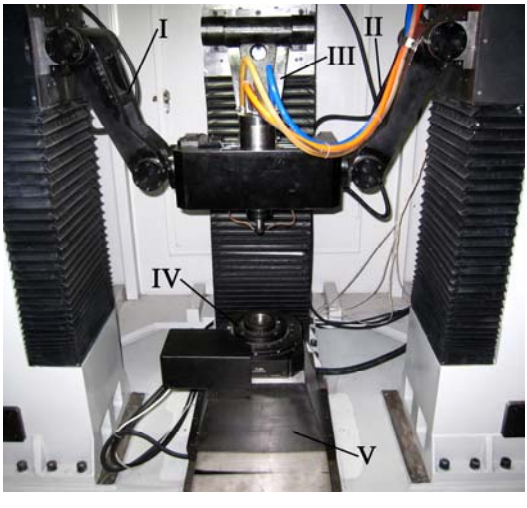


Fig. 2. Parallel module and serial table of the SPKM165 machine

Generally, this milling machine performs as a vertical hybrid machine, its positional workspace is 165 mm×165 mm×165 mm and the rotation angle β about the *y*-axis in the parallel module can be $\beta \in (-25°, 90°)$. The rotating table IV in the serial table can rotate continuously. So, this machine is capable of 5-face machining and can satisfy some special requirements.

3 Inverse and Forward Kinematics of the Parallel Module

According to the specific application requirements, we use the method proposed in [3] to get the parameters of the machine as follow:

$$\begin{aligned}
&\left|T_1T_2\right| = d = 730 \text{ mm}\\
&\left|P_iQ_i\right| = l_i = 300 \text{ mm } (i = 1,2,3)\\
&\left|Q_1Q_2\right| = r = 470 \text{ mm}\\
&\left|oQ_3\right| = r_1 = 180 \text{ mm}\\
&\left|OT_3\right| = r_2 = 240 \text{ mm}\\
&\left|Oo\right| = h = 760 \text{ mm}
\end{aligned}$$

Frame *o-yz* locates in the zero position of the parallel module and the coordinate of point o in frame *O-YZ* is $(0, h)$.

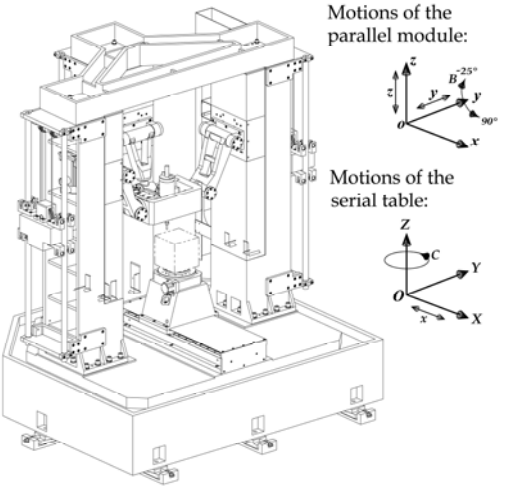


Fig. 3. Motions of the SPKM165 machine

The kinematic scheme of the parallel module is shown in Fig. 4, for a given position (y, z) of the moving platform in frame *o-yz* and rotation angle β, we can get the inputs of the three sliders in frame *O-XYZ* as follows:

$$s_i = \frac{-b_i \pm \sqrt{b_i^2 - 4c_i}}{2} \quad (i = 1,2,3) \tag{1}$$

where

$$b_1 = -2(z + h)$$

$$b_2 = -2(z + h)$$

$$b_3 = 2\left[r_1 \sin\beta - (z+h)\right]$$

$$c_1 = \frac{r^2}{4} - l_1^2 + \left(y + \frac{d}{2}\right)\left(y + \frac{d}{2} - r\right) + (z+h)^2$$

$$c_2 = \frac{r^2}{4} - l_2^2 + \left(y - \frac{d}{2}\right)\left(y - \frac{d}{2} + r\right) + (z+h)^2$$

$$c_3 = r_1^2 - l_3^2 + r_2\left(r_2 - 2r_1\cos\beta\right) + (z+h)\left(z+h-2r_1\sin\beta\right)$$

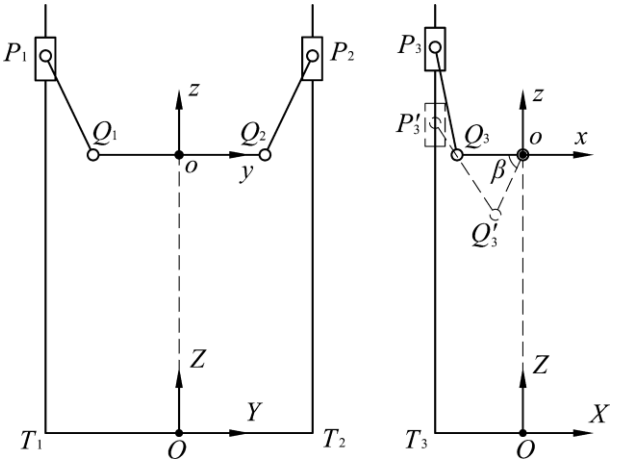


Fig. 4. Kinematic scheme of the parallel module

Then, for a given pose (y, z, β) of the moving platform, there are eight solutions for the parallel module in theory, as shown in Fig. 5.

In practical application, we use the configuration (a). So, we choose "+" in Eq. (1).

The actual positions of the three sliders of the SPKM165 machine are

$$s_i = \frac{-b_i + \sqrt{b_i^2 - 4c_i}}{2} \quad (i = 1, 2, 3) \tag{2}$$

In contrast, when the positions of the three sliders (s_i $(i = 1, 2, 3)$) are given, we can get the position (y, z) of the moving platform in frame *o*-*yz* and rotation angle β as follows:

when $s_1 \neq s_2$

$$y = g \tag{3}$$

$$z = \frac{s_1 + s_2}{2} - h - tg \tag{4}$$

$$\beta = \sin^{-1}(e) \tag{5}$$

where

$$e = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{6}$$

$$a = 4r_1^2\left[r_2^2 + \left(\frac{s_1 + s_2}{2} - tg - s_3\right)^2\right]$$

$$b = -4kr_1\left(\frac{s_1 + s_2}{2} - tg - s_3\right)$$

$$c = k^2 - 4r_1^2 r_2^2$$

$$k = r_1^2 + r_2^2 + \left(\frac{s_1 + s_2}{2} - tg - s_3\right)^2 - l_3^2$$

$$g = \pm\sqrt{\frac{l_1^2}{1+t^2} - \frac{(s_2 - s_1)^2}{4}} \tag{7}$$

$$t = \frac{d - r}{s_2 - s_1}$$

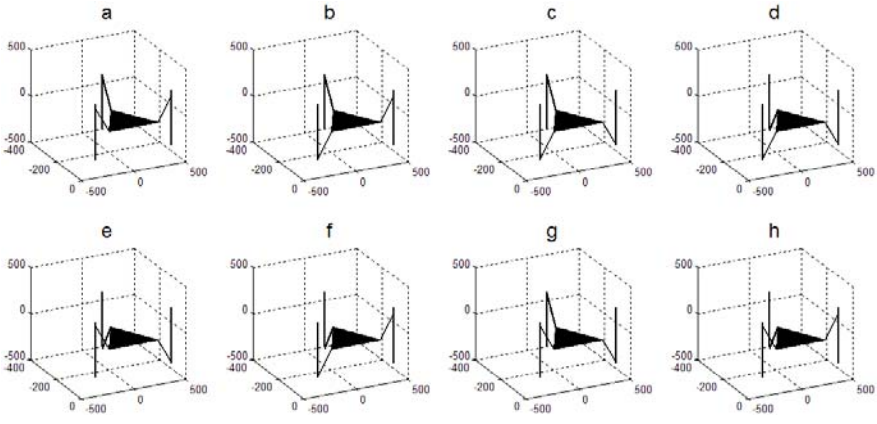


Fig. 5. Available configurations when y=0, z=–20 mm, β=0

when $s_1 = s_2$

$$y = 0 \tag{8}$$

$$z = s_2 \pm \sqrt{l_2^2 - \frac{(d - r)^2}{4}} - h \tag{9}$$

$$\beta = \sin^{-1}(e) \tag{10}$$

where

$$e = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{11}$$

$$a = 4r_1^2\left[r_2^2 + \left(s_2 \pm \sqrt{l_2^2 - \frac{(d-r)^2}{4}} - s_3\right)^2\right] \tag{12}$$

$$b = -4kr_1\left(s_2 \pm \sqrt{l_2^2 - \frac{(d-r)^2}{4}} - s_3\right) \tag{13}$$

$$c = k^2 - 4r_1^2 r_2^2$$

$$k = r_1^2 + r_2^2 + \left(s_2 \pm \sqrt{l_2^2 - \frac{(d-r)^2}{4}} - s_3\right)^2 - l_3^2 \tag{14}$$

So, for the given positions of the three sliders (s_i $(i = 1, 2, 3)$), there are four solutions for the parallel module in theory, as shown in Fig. 6. In practical application, we use the configuration (d). When $s_1 > s_2$, we choose "+" in Eq. (6) and "-" in Eq. (7) ; when $s_1 < s_2$, we choose "+" in Eqs. (6) and (7); when $s_1 = s_2$, we choose "-" in Eqs. (9) and (12-14) and "+" in Eq. (11).

Fig. 6. Available configurations when $s_1 = s_2 = 1010$ mm, $s_3 = 1034$ mm

4 Establishment of the Joint Five-Axis Coordinate System and Inverse Kinematics of the SPKM165 Machine

In order to carry out five-axis control, we will investigate the relationship between the NC-code and the position of the four sliders in this part.

According to the characteristic of the machine, we can get the simplified kinematic scheme as shown in Fig. 7.

In Fig. 7, the tool length $TP = l = 180$ mm and the frame *O-XYZ* is static, the moving frame *o-xyz* is located on the moving table and the point o is coincident with the tool tip P when the parallel module is in zero position.

If NC-code is given, we know exactly the numerical values of x, y, z, B and C. Here, x, y and z are the coordinates of tool tip P in the moving frame *o-xyz*, B is the attitude of the tool and denoted by β, C is the rotational angle of the work piece and denoted by γ.

Then, the ranges of the coordinates are

$$\begin{aligned} &x \in (0, 165) \text{ mm} \\ &y \in (-82.5, 82.5) \text{ mm} \\ &z \in (-165, 0) \text{ mm} \\ &B \in (-25°, 90°) \\ &C \in (0°, 360°) \end{aligned}$$

and the positions of the sliders along the Z-axis are

$$s_i = \frac{-b_i + \sqrt{b_i^2 - 4c_i}}{2} \quad (i = 1, 2, 3) \tag{15}$$

where

$$b_1 = -2(z - l + l\cos\beta + h)$$

$$b_2 = -2(z - l + l\cos\beta + h)$$

$$b_3 = 2[r_1 \sin\beta - (z - l + l\cos\beta + h)]$$

$$c_1 = \frac{r^2}{4} - l_1^2 + \left(y + \frac{d}{2}\right)\left(y + \frac{d}{2} - r\right) + (z - l + l\cos\beta + h)^2$$

$$c_2 = \frac{r^2}{4} - l_2^2 + \left(y - \frac{d}{2}\right)\left(y - \frac{d}{2} + r\right) + (z - l + l\cos\beta + h)^2$$

$$c_3 = r_1^2 - l_3^2 + r_2(r_2 - 2r_1\cos\beta) + (z - l + l\cos\beta + h)(z - l + l\cos\beta + h - 2r_1\sin\beta)$$

The position of the moving table along the X -axis is

$$s_4 = l \sin \beta - x \tag{16}$$

The rotation angle of the moving platform about Z -axis is γ.

By putting all the information of the coordinate system and inverse kinematic equations into the computer numerical control system, we can achieve the five-axis control of the SPKM165 machine.

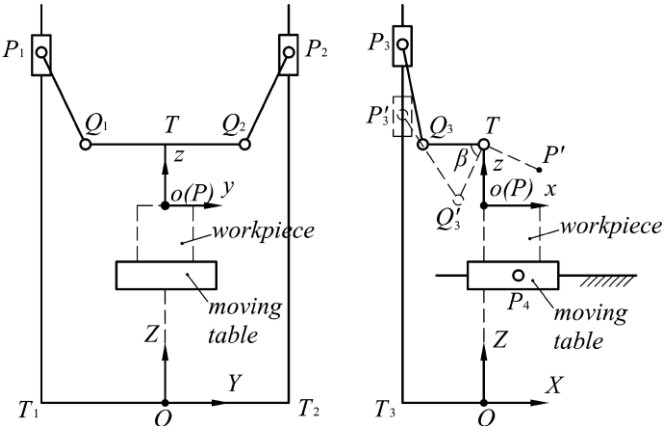


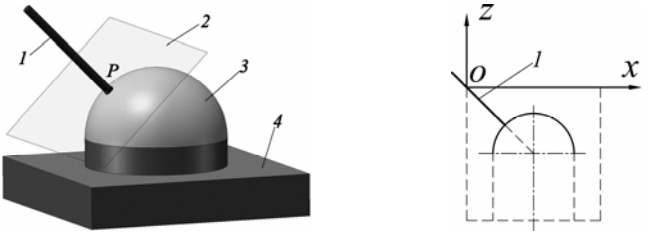
Fig. 7. Kinematic scheme of the full SPKM165 machine

5 Process of Milling a Semi-spherical Work Piece

In order to illustrate the working process, we choose a work piece with typical spherical surface, i.e., a semi-spherical work piece.

Step 1: According to the characteristic of the machine to fix the work piece in the moving table.

As shown in Fig. 8(a), in the milling process, the tool (1) should be perpendicular to the normal plane (2) of the milling point ***P*** in the semi-spherical



(a) (b)

1-tool, 2-normal plane in milling point P, 3-semi-spherical work piece, 4-moving table

Fig. 8. Schematic drawings of the milling process and fixing the work piece

surface (3). So, we fix the work piece in the manner of Fig. 8(b), and, in this case, $y = 0$, $\gamma \in (0, 360°)$.

Step 2: Generate the NC-code of the semi-spherical surface.

We usually get the NC-code of a work piece through a third-part software, such as UG or Pro/E, especially, when the profile is complicate. Here, the profile is circular symmetry about the Z -axis and the generation of the NC-code will be easy, so we just use Matlab to achieve this.

Step 3: Get the numerical result of the five active motors.

Based on the result in Step 2 and the equations in section 4, the result is easy to get, and the numerical result is given in table 2. For the SPKM165 machine, this step is completed automatically by the kinematics module in CNC system.

Table 2. Numerical results of the milling process where the radius is 50 mm and stepping by 10 deg (the step can be changed to make the surface smooth)

x (mm)	z (mm)	B (deg)	δs_1 (mm)	δs_2 (mm)	δs_3 (mm)	δs_4 (mm)
82.50	-115.00	0	-115.00	-115.00	-115.00	82.50
73.82	-115.76	10	-118.49	-118.49	-150.32	42.56
65.40	-118.02	20	-128.87	-128.87	-192.86	3.84
57.50	-121.70	30	-145.81	-145.81	-241.79	-32.50
50.36	-126.70	40	-168.81	-168.81	-296.36	-65.34
44.20	-132.86	50	-197.16	-197.16	-355.95	-93.69
39.20	-140.00	60	-230.00	-230.00	-420.02	-116.69
35.52	-147.90	70	-266.34	-266.34	-488.25	-133.63
33.26	-156.32	80	-305.06	-305.06	-560.80	-144.01
32.50	-165.00	90	-345.00	-345.00	-638.94	-147.50

where, δs_i -drive distance of motor i $(i$=1, 2, 3, 4$)$.

6 Conclusions

This paper introduced a 5-axis milling machine with serial and parallel kinematics, which has the advantages of simple kinematics, 5-face machining and convenient trajectory planning. The inverse kinematics of the parallel module and the full machine were analyzed, respectively, and by putting the information of the coordinate system and inverse kinematic equations into the CNC system, the five-axis control of the machine was achieved. The milling process of a semi-spherical work piece was presented to illustrate the working process of the machine and the numerical result was given at last. In order to satisfy the requirements of precision milling, the kinematic calibration of the machine is in progress.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant 50775118, the National Basic Research Program (973 Program) of China

(No. 2007CB714000), and by Program for New Century Excellent Talents in University of China (No. NCET-08-0323).

References

1. Merlet, J.P.: Parallel Robots. Springer, New York (2005)
2. Tlusty, J., Ziegert, J., Ridgeway, S.: Fundamental comparison of the use of serial and parallel kinematics for machines tools. CIRP Annals - Manufacturing Technology 48, 351–356 (1999)
3. Wang, J.S., Liu, X.J., Wu, C.: Optimal design of a new spatial 3-DOF parallel robot with respect to a frame-free index. Science in China Series E: Technological Sciences 52, 986–999 (2009)
4. Kim, H.S., Tsai, L.-W.: Kinematic synthesis of a spatial 3-RPS parallel manipulator. Journal of Mechanical Design 125, 92–97 (2003)
5. Kanaan, D., Wenger, Ph., Chablat, D.: Workspace analysis of the parallel module of the VERNE machine. Problems of Mechanics 4, 26–42 (2006)
6. Kanaan, D., Wenger, P., Chablat, D.: Kinematic analysis of a serial–parallel machine tool: The VERNE machine. Mechanism and Machine Theory 44, 487–498 (2009)
7. Liu, X.J., Wang, J.S., Gao, F., Wang, L.P.: On the analysis of a new spatial three-degree-of freedom parallel manipulator. IEEE Transactions on Robotics Automation 17, 959–968 (2001)
8. Nair, R., Maddocks, J.H.: On the forward kinematics of parallel manipulators. International Journal of Robotics Research 13, 171–188 (1994)
9. Liu, X.-J., Wang, J.S., Wu, C., Kim, J.: A new family of spatial 3-DOF parallel manipulators with two translational and one rotational DOFs. Robotica 27, 241–247 (2009)

Kinematics Model of Bionic Wheel-Legged Lunar Rover and Its Simulation Analysis

Yongming Wang[1,2], Xiaoliu Yu[1,3], and Wencheng Tang[2]

[1] Anhui University of Technology, Maanshan, Anhui Province, China
[2] Southeast University, Nanjing, Jiangsu Province, China
[3] State Key Laboratory of Robotics, Shenyang, Liaoning Province, China
wym_ahut@163.com

Abstract. Based on the principle of double-half-revolution mechanism, a new-style bionic wheel-legged lunar rover was designed in this paper. Its mobile mechanism is comprised of four identical wheel-legged mechanisms, rover body bracket and steering brackets, etc. On the basis of establishing the coordinate transformation models of the bionic wheel-legged mechanisms, kinematics of its mobile mechanism was analyzed, and the forward kinematics models of the striking rods were established. Based on kinematics simulation in COSMOS Motion software, this paper obtained the trail curves of joints relative to the reference coordinate system. Simulation shows that the fluctuation of the bionic wheel-legged lunar rover is small and it can realize the striding movement smoothly.

Keywords: Lunar Rover, Bionic Wheel-legged Mechanism, Double-half-revolution Mechanism, Kinematics Model, Simulation Analysis.

1 Introduction

As one of the key technology in lunar exploration, the lunar rover is an important tool for realizing the second scientific target of the unmanned lunar exploration program in China. The mobile mechanism of the lunar rover has some fundamental types, such as wheeled-type, legged-type and tracked-type, etc [1-5]. In the existing literatures, their research objects are mostly wheeled-type. But the exploration data shows that the lunar surface is undulate and its lunar soil is soft, the wheeled-type lunar rover may sink into the lunar soil. Compared to the propulsion mode of wheeled or tracked type, wheel-legged mobile mechanism has unique advantages, such as good mobility and adaptability of road surface, and is able to walking on the undulate terrain or soft ground with little less efficiency [6].

Taking a new-style bionic wheel-legged lunar rover for research object, this paper analyzed its kinematics of mechanism, and established the forward kinematics models of the striking rods. Based on kinematics simulation of the lunar rover in COSMOS Motion software, this paper obtained the trail curves of joints relative to the reference coordinate system.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 603–612, 2009.

2 The Principle of the Bionic Wheel-Legged Mechanism

When animals fly in air, swim in water or run on lands, the styles of their motions are different, but they are all nonsymmetrical swing in essence[7]. For example, when the pinions swoop downward or fly upward, the horse' legs kick rearward or stretch forward, their motions are different. Swing is a motion style which is adaptable to the muscles, and the nonsymmetrical motion is a necessary condition for the animals to fly upward or walk forward.

Double-half-revolution mechanism is a new mechanism, which simulate the nonsymmetrical motion of animal organs [8]. It is composed of one first swiveling arm, one second swiveling arm and two striding rods, as shown in Fig.1. At one end of the first swiveling arm there is a spindle connected by a rotary joint, and at the other end there is the second swiveling arm; At each end of the second swiveling arm there is a striding rod which is also connected by a rotary joint, and the two striding rods are always perpendicular to each other. When the first swiveling arm rotates θ around the spindle, the second swiveling arm rotates $-\theta/2$ relative to the first swiveling arm and the striding rods rotate $-\theta/4$ relative to the second swiveling arm.

Based on the Principle of double-half-revolution mechanism, we designed a bionic wheel-legged mechanism, as shown in Fig. 2. The motion between the spindle and the two striding rods is transmitted by two tandem planetary gear trains, in which, the first and second swiveling arm are the planet carriers (also as gearbox casings) of the planetary gear trains [9]. When the motor drives the first swiveling arm to rotate around the spindle, the endpoints of the striding rods contact to the land in turn, thus realize the walking function.

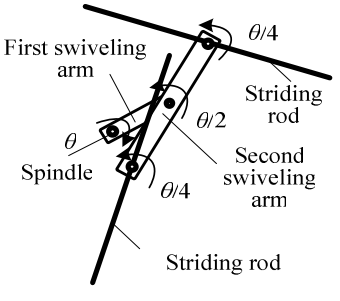


Fig. 1. Double-half-revolution mechanism

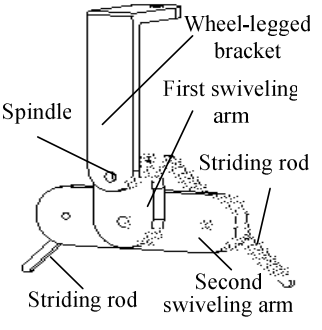


Fig. 2. Bionic wheel-legged mechanism

3 The Principle of the Bionic Wheel-Legged Lunar Rover

The bionic wheel-legged lunar rover is comprised of rover body bracket, four identical wheel-legged mechanisms and steering bracket, the principle of its mobile mechanism is shown in Fig.3. The rover body bracket is used to support the whole lunar rover and the scientific instruments loaded on it. Four identical wheel-legged mechanisms are symmetrically mounted on the corresponding wheel-leg brackets, which are connected to the corresponding mounting brackets with a rotary joint. Among them,

two mounting brackets are mounted on each end of the steering bracket, which is connected to the rover body bracket with a free pivoting joint, and another two mounting brackets are connected to the rover body bracket directly. When the lunar rover climbs the obstacle with one-sided wheel-legged mechanisms, the steering bracket ensures that the four wheel-legged mechanisms could contact the land and have no excessive sidelurch of the rover body, thus is favorable for getting smooth-going and stability of the lunar rover. The walking and steering motions of the bionic wheel-legged lunar rover are driven by the motor independently.

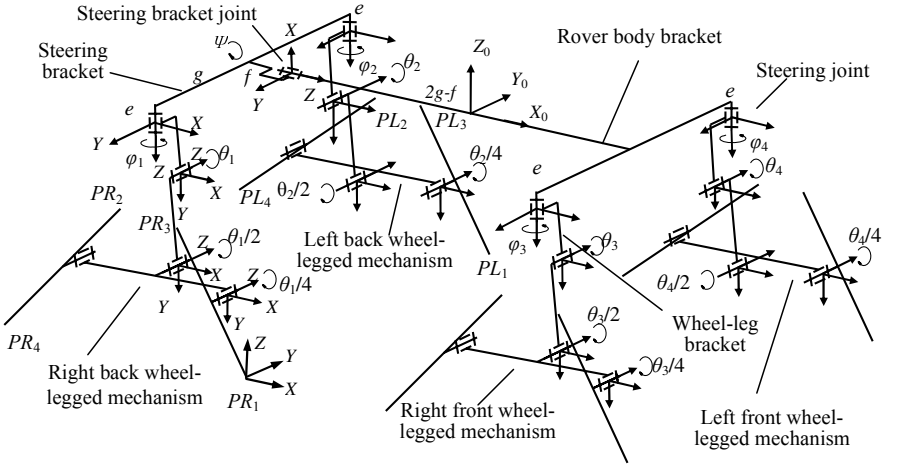


Fig. 3. The principle of the bionic wheel-legged lunar rover

4 Kinematics Analysis of the Bionic Wheel-Legged Lunar Rover

4.1 Definition of the Coordinate Systems and Parameters

The coordinate systems of the lunar rover should be firstly established before kinematics analysis in order to describe their translation and rotation relationship [10], as shown in Fig.3. It is assumed that the centroid of rover body is situated in the center of rover body, and the reference coordinate system ($O_0X_0Y_0Z_0$) is established in there. $\theta_1, \theta_2, \theta_3$ and θ_4 are the rotation angles of the joints of the driving motors, $\varphi_1, \varphi_2, \varphi_3$ and φ_4 are the rotation angles of the joints of the steering motors. ψ is the pivot angle of the steering bracket, which can be measured by the sensor. α, β and γ are the yaw, pitch and roll angles of the rover body.

The parameters of the four wheel-legged mechanisms are identical, and are defined as the following. The length of the striking rod is $2a$, the length of the second swiveling arm is $2b$, the length of the first swiveling arm is c, and the height of the wheel-leg bracket is d. h_1 is the distance between the striking rod and the second swiveling

arm, h_2 is the distance between the first and second swiveling arm, h_3 is the distance between the first swiveling arm and the wheel-leg bracket, h_4 is the distance between the wheel-leg bracket and the steering joint of the wheel-legged mechanism. The vertical interval between the steering joint of the wheel-legged mechanism and the rover body bracket is e, the distance between the right and left steering joints of the wheel-legged mechanisms is $2g$, the distance between the steering bracket joint and the rear beam of the rover body bracket is f, the distance between the steering bracket joint and the front beam of the rover body bracket is $2g$-f.

4.2 Coordinate Transformation Models of the Wheel-Legged Mechanisms

For convenience, we firstly deduce the coordinate transformation matrixes from the steering joints of wheel-legged mechanisms to the endpoints of the striking rods, and then establish the foreword kinematics model of the lunar rover.

Coordinate Transformation Model of the Right Wheel-Legged Mechanisms
Taking the right-back wheel-legged mechanism for example, when the lunar rover walks forward, the endpoints of the two striking rods will contact the land in turn, and the touchdown order is PR_1, PR_2, PR_3, and PR_4, as shown in Fig.3. And when the lunar rover walks backward, the touchdown order is opposite. The coordinate transformation model of the right wheel-legged mechanisms can be established, as shown in Fig.4. So the coordinate transformation from the steering joints of wheel-legged mechanisms to the endpoints of the two striking rods can be expressed as (1).

$$TP_R = T_{R4}T_{R3}T_{R2}T_{R1} \ . \tag{1}$$

Coordinate Transformation Model of the Left Wheel-Legged Mechanisms
Taking the left-back wheel-legged mechanism for example, when the lunar rover walks forward, the endpoints of the two striking rods will contact the land in turn, and the touchdown order is PL_1, PL_2, PL_3, and PL_4, as shown in Fig.3. And when the lunar rover walks backward, the touchdown order is opposite. The coordinate transformation model of the left wheel-legged mechanisms can be established, as shown in Fig.5. So the coordinate transformation from the steering joints of wheel-legged mechanisms to the endpoints of the two striking rods can be expressed as (2).

$$TP_L = T_{L4}T_{L3}T_{L2}T_{L1} \ . \tag{2}$$

4.3 Forward Kinematics Model of Wheel-Legged Mechanisms

Forward kinematics of wheel-legged mechanisms is to solve the endpoints' pose of the striding rods, given the pose of the center of rover body and the displacements of joints. In the reference coordinate system, the coordinate transformation matrix of the center of rover body can be written as (3).

$$T_0 = R(Z,\alpha)R(Y,\beta)R(X,\gamma) \ . \tag{3}$$

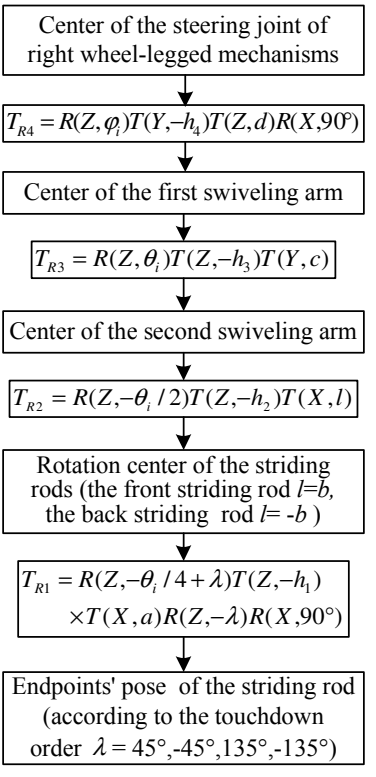


Fig. 4. Coordinate transformation model of the right wheel-legged mechanisms

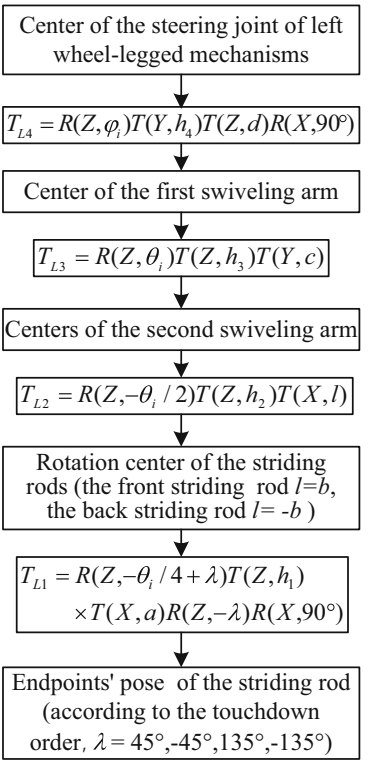


Fig. 5. Coordinate transformation model of the left wheel-legged mechanisms

Then, according to the coordinate systems in Fig.3, we can obtain the transformation matrixes from the center of rover body to the four steering joints of the wheel-legged mechanisms, written as the following.

$$^{o}T_{RB} = T_o T(X, f-g)R(Y,90^{\circ})R(Z,180^{\circ})T(Z,-f)R(Z,\psi)T(Y,g)T(Z,e) \ . \quad (4)$$

$$^{o}T_{LB} = T_o T(Y, f-g)R(Y,90^{\circ})R(Z,180^{\circ})T(Z,-f)R(Z,\psi)T(Y,-g)T(Z,e) \ . \quad (5)$$

$$^{o}T_{RF} = T_o T(X, g)R(X,180^{\circ})R(Z,\psi)T(Y,g)T(Z,e) \ . \quad (6)$$

$$^{o}T_{LF} = T_o T(Y, g)R(X,180^{\circ})R(Z,\psi)T(Y,-g)T(Z,e) \ . \quad (7)$$

Where, subscripts *RB, LB, RF*, and *LF* indicate right-back, left-back, right-front, and right-front.

Therefore, the endpoints' pose of the striding rods can be described in (8).

$$T_j = {}^{o}T_j TP_i \ . \quad (8)$$

Where, *i* represents *L* or *R* *j* represents *RB, LB, RF, LF*.

In the reference coordinate system, we define that T is the endpoints' pose of the striding rods and it is expressed as (9).

$$T = \begin{bmatrix} n & o & a & p \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} n_x & o_x & a_x & p_x \\ n_y & o_y & a_y & p_y \\ n_z & o_z & a_z & p_z \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{9}$$

Where, *n, o, a* are the vectors of direction cosine, *p* is a vector of position.

According to (8) and (9), we can obtain the forward kinematics of the striking rods of each wheel-legged mechanisms relative to the center of rover body, which can be calculated by programming.

5 Simulation Analysis

The mechanism parameters of the lunar rover are given in Table 1. We establish the virtual prototyping model of the lunar rover by setting up constraint relations among the rods of mechanism, such as coincidence, concentricity, parallel and tangency, etc. Taking the lunar rover walks straight on the flat ground for example, in COSMOS Motion software, we define motion coupling relationship of angular speeds of all joints of wheel-legged mechanisms in order to meet transmission characteristics of double- half-revolution mechanism.

Table 1. The mechanism parameters of the bionic wheel-legged lunar rover

Parameters	Values(mm)	Parameters	Values(mm)	Parameters	Values(mm)
a	130	e	20	h_2	28
b	103.5	g	500	h_3	24
c	69.75	f	40	h_4	80
d	245	h_1	28		

Given that the rotational speed of driving motor is 360°/s, and the simulation time is 2 second, we can obtain the centroid position curve of lunar rover body in *X*, *Z* direction, as shown in Fig.6. According to the motion characteristics of double-half-revolution mechanism, the lunar rover body has a certain fluctuation in vertical direction. Simulation data show that the maximum of vertical fluctuation of lunar rover body is 36.7mm.

When lunar rover walks on the flat ground, the motion traces of four wheel-legged mechanisms of lunar rover are similar, so we just take right-back wheel-legged mechanisms for example. Taking the centroid of lunar rover body as the reference coordinate system, we can obtain the trace curves of rotation center of the first swiveling arm, the second swiveling arm and the striding rods, as shown in Fig.7- Fig.10.

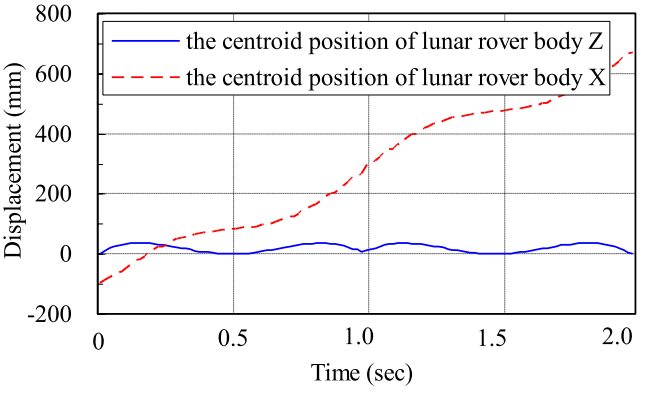


Fig. 6. The centroid position curve of lunar rover body

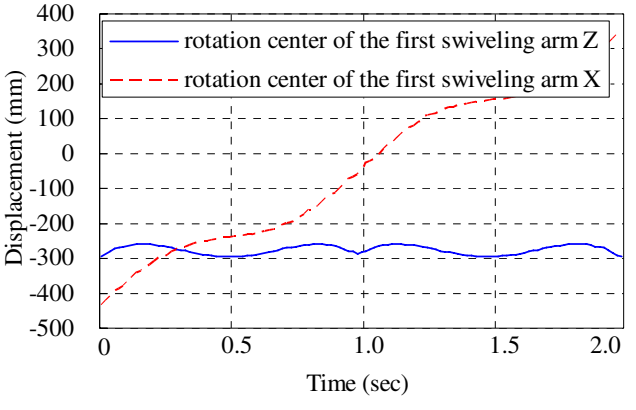


Fig. 7. The trace curves of rotation center of the first swiveling arm

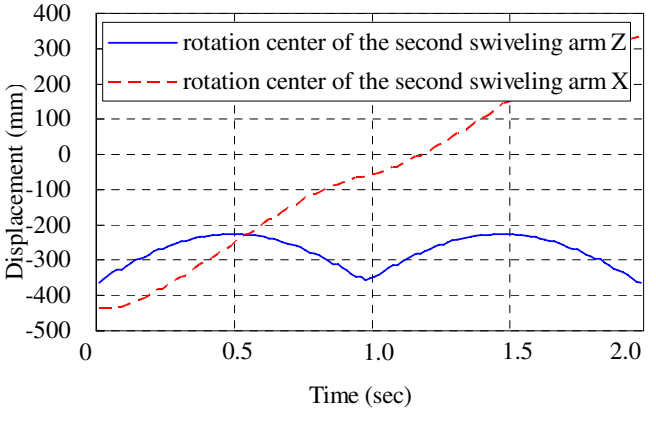


Fig. 8. The trace curves of rotation center of the second swiveling arm

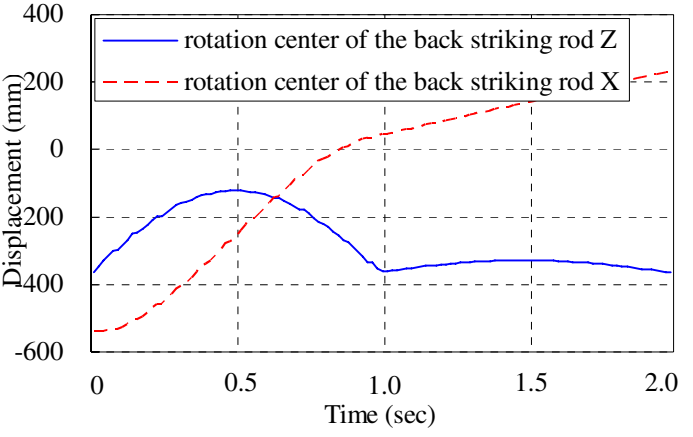


Fig. 9. The trace curves of rotation center of the back striking rod

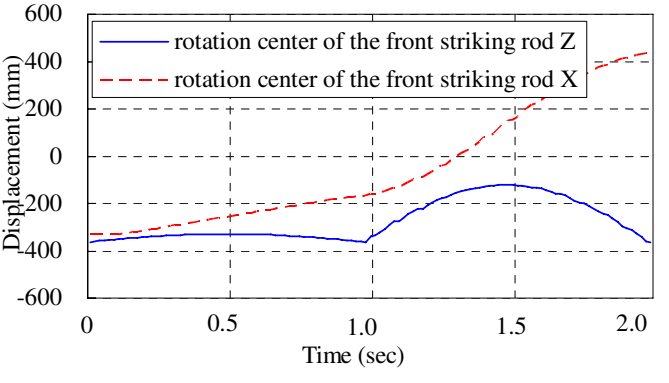


Fig. 10. The trace curves of rotation center of the front striking rod

From Fig.6 and Fig.7, we can conclude that the vertical fluctuation curve of lunar rover body is similar to the trace curve of rotation center of the first swiveling arm in *Z* direction, and essentially they are the vertical fluctuation curve of double-half-revolution mechanism; As shown in Fig.8, the trace curve in *Z* direction is equal to the vertical fluctuation curve of single-half-revolution mechanism, and its vertical fluctuation is 139.6mm. So the vertical fluctuation of the bionic wheel-legged lunar rover with double-half-revolution mechanism is small.

When lunar rover walks on the ground, during a stride action, one striding rod of each wheel-legged mechanism serves as a supporting role, and the other serves as a striding role. Given that the rotational speed of driving motor is 360°/s, and simulation time is 4 second. For the four endpoints (Marked as 1, 2, 3, 4.) of the two striding rods touch the ground in turn, their trace curves also have a horizontal straight lines in turn, as shown in Fig.11.

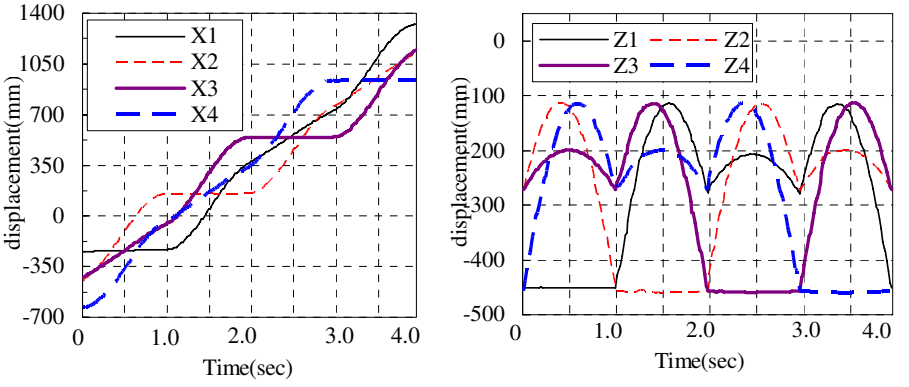


Fig. 11. The trace curves of four endpoints of the two striding rods

6 Conclusions and Future Work

Based on double-half-revolution mechanism, this paper designed a new-type wheel-legged lunar rover and introduced its mechanism principle. Then established the coordinate transformation models of the bionic wheel-legged mechanisms, and analyzed the forward kinematics of the striking rods relative to the center of rover body. Kinematics simulation in COSMOS Motion software shows that the vertical fluctuation of the bionic wheel-legged lunar rover is small and it can realize the striding movement smoothly.

The future work which we are pursuing to do is to improve its mobile performance of the lunar rover. On the basis of innovation design of mechanisms, mechanics analysis and dynamic simulation analysis, etc, the prototype of wheel-legged lunar rover will be developed and its experimental study of mobile performance will be done.

Acknowledgments. This work is supported by the key innovation project granted by the Ministry of Education in China (No.708054), the fund of the Chinese Academy of Sciences State Key Laboratory of Robotics (No.RLO200816) and the key project of Anhui Education Department in China (No.KJ2009A015Z).

References

1. Chacin, M., Yoshida, K.: Evolving legged rovers for minor body exploration missions. In: 1st IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics, Pisa, Italy, pp. 170–175 (2006)
2. Shang, J.Z., Luo, Z.R., Fan, D., Zhao, F.: A six wheeled robot with active self-adaptive suspension for lunar exploration. In: International Technology and Innovation Conference, ITIC 2006, Hangzhou, China, pp. 2035–2038 (2006)
3. Shang, W.Y., Li, S.M., Qiu, F.J., Xin, J.H.: Structure Designing and Analysis of New Style Four Leg Wheeled Lunar Rover. Journal of Wuhan University of Technology (Transportation Science & Engineering) 32(5), 822–825 (2008) (in Chinese)

4. Hou, X.Y., Gao, H.B., Hu, M.: Optimal Design of Suspension's Dimensions for Six-wheeled Rocker Lunar Rover Based on Climbing Obstacle Capability. Journal of Astronautics 29(5), 420–425 (2008) (in Chinese)
5. Ishigami, G., Nagatani, K., Yoshida, K.: Slope traversal experiments with slip compensation control for lunar/planetary exploration rover. In: IEEE International Conference on Robotics and Automation, Pasadena, CA, United States, pp. 2295–2300 (2008)
6. Sun, F., Li, J.: The design of four foot walking mechanism based on Cosmos Motion. Machinery 34(3), 24–25 (2007)
7. Qiu, Z.Z., Xie, N.G.: The Reversed Thinking and Outlet of Animal-motion Bionics. In: Proceedings of the 6th international conference on frontiers of design and manufacturing, Xi'an, China, pp. 477–478 (2004)
8. Wang, Y., Yu, X., Tang, W.: A New-type Wheel-legged Lunar Rover and its Mobile Performance Analysis. In: The 2009 International Conference on Mechanical and Electrical Technology (ICMET 2009), pp. 31–35. IEEE Press, Beijing (2009)
9. Wang, Y., Yu, X., Chi, R., Tang, W.: A New-style Bionic Walking Mechanism and its Motion Simulation. In: The 2009 Second International Conference on Intelligent Computation Technology and Automation (ICICTA 2009), pp. 941–944. IEEE Press, Zhangjiajie (2009)
10. Deng, Z.Q., Hu, M., Gao, H.B.: Kinematics modeling of the Lunar Rover. Chinese Journal of Mechanical Engineering 14(22), 1911–1913 (2003) (in Chinese)

Dynamics and GA-Based Optimization of Rectilinear Snake Robot

Ahmad Ghanbari[1,2], Mir Masoud Seyyed Fakhrabadi[1], and Ali Rostami[2,3]

[1] Department of Mechanical Engineering, University of Tabriz,
P.O. Box 5166616471, Tabriz, Iran
[2] Center of Exellence for Mechatronics, School of Engineering- Emerging Technologies,
University of Tabriz, P.O. Box 5166616471, Tabriz, Iran
[3] Department of Electrical and Computer Engineering, University of Tabriz,
P.O. Box 5166616471, Tabriz, Iran
a-ghanbari@tabrizu.ac.ir, msfakhrabadi@gmail.com,
rostami@tabrizu.ac.ir

Abstract. The goal of this study is analysis and optimization of rectilinear locomotion gait as one of the snake robot motion modes. An overview of kinematics and detailed dynamics and optimization of torque values are presented in this study. Well-known genetic algorithm scheme will be used in this paper.

Keywords: Snake Robot, Dynamics, Genetic Algorithm.

1 Introduction

Mechanics and Biomechanics scientists and researchers always have had great interests about robots inspired of animals like fish, cockroach, spider, snake and others [1-3]. Because robotic prototypes for search and rescue operations tend to imitate crawling organisms such as snake or worms, the crawling motions of snakes and other limbless animals have enormous backgrounds of research for specialists [4-6].

Snakes perform many kinds of movements that are adaptable to a given environment. They are so diversified that exhibit many examples of locomotive specialization, but in which one reptantion mode is always favoured by a particular kind of snakes. When these reptantions are broadly classified, there exist four gliding modes: serpentine, concertina, side-winding, and rectilinear locomotion [7].

Rectilinear gait for modular robot performed like a snake-inspired robot has been studied theoretically [8]. In this described model, progression is achieved by the propagation of an undulatory wave from the rear to the front of the robot.

Detailed kinematics of rectilinear locomotion mode has been studied and developed [4].

In this paper we intend to present its dynamic formulation and optimization using genetic algorithm. For this purpose, we will have a brief overview on broken sub mechanisms and their kinematics, and then dynamics and optimization relations will be developed. Finally required torque values for optimized gait will be calculated and reported.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 613–622, 2009.

2 Overview of Kinematics

Here, we will present description of basic motion pattern of rectilinear snake robot and then develop its kinematics formulation by recursive Newton-Euler method.

It is shown in Fig. 1 that sequences of joint configurations will lead to a forward motion if the robot is driven through them. Then the distance traveled by the robot per complete cycle will be [8]:

$$x = 2\, l\, (1 - \cos \alpha\,). \tag{1}$$

Where α is the angle of risen module and l is the module length.

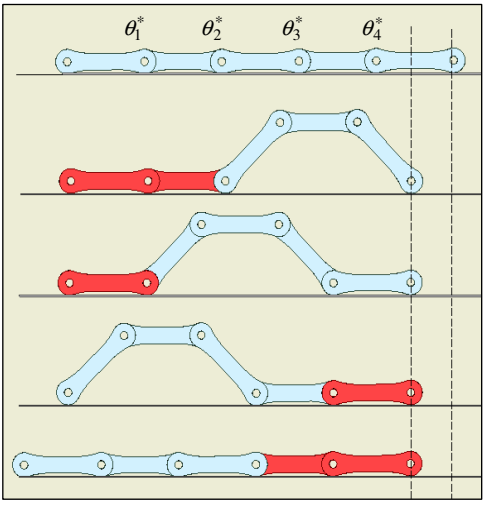


Fig. 1. Inchworm gait sequence

According to the Fig. 1, this motion pattern can be resolved in four sub-mechanisms M_1, M_2, M'_2, and M'_1. The joint angles that should be reached at the end of each step are as values in Table 1.

Table 1. Joint angles at end of each step

Sub mechanism	θ_1^*	θ_2^*	θ_3^*	θ_4^*
start	0	0	0	0
M_1	0	α	- α	- α
M_2	α	- α	- α	α
M'_2	- α	- α	α	0
M'_1	0	0	0	0

We suppose that friction is provided so sufficient that at least one link is in permanent contact with ground and doesn't move. These ground or nonmoving links will be shown by dark color in all figures. Also, with due to attention to Figure 1, we find a symmetric condition in motion sequences. Hence, we can model only two first sub-mechanisms as a manipulator according to Fig. 2 and Fig. 3, then for last two sub-mechanisms, we reverse coordinate system and use opposite trajectory directions.

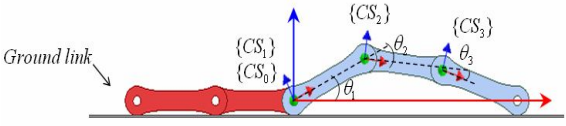


Fig. 2. Sub-mechanism M_1

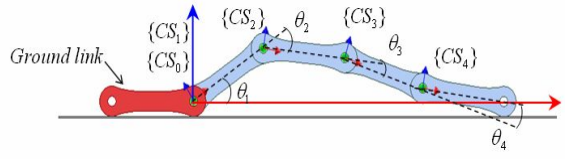


Fig. 3. Sub-mechanism M_2

It is worth to note that, in these models the angles are measured in local joint coordinate systems. Therefore, after solving each sub-problem, we will map them to global coordinate system by using a simple matrix multiplier scheme.

Sub-mechanism M_1 is the first part of gait sequence and it is considered as a 2-DOF mechanism which the tip of the third link remains in constant contact with the ground during the motion. Hence we can write [4]:

$$\sum_{i=1}^{3} \sin \varphi_i = 0. \tag{2}$$

Where φ_i is the i^{th} link angle respect to positive direction of horizontal axes and is given by below formula:

$$\varphi_i = \sum_{j=1}^{i} \theta_j . \tag{3}$$

Then, trajectories will be generated for joints 1 and 2 that were treated as free variables. Also, position, angular velocity and angular acceleration of joint 3 are calculated using Eq.2 and its derivatives [4].

Finally, the kinematic parameters are computed using the recursive Newton-Euler method [9]. The Newton-Euler outward iterative method propagates the kinematics from the base of the manipulator to the tip as the following relations:

$$^{i+1}\omega_{i+1} = {}^{i}\omega_i + \dot{\theta}_{i+1}\,{}^{i+1}\hat{Z}_{i+1} \tag{4}$$

$$^{i+1}\dot{\omega}_{i+1} = {}^{i}\dot{\omega}_i + {}^{i}\omega_i \times \dot{\theta}_{i+1}\,{}^{i+1}\hat{Z}_{i+1} + \ddot{\theta}_{i+1}\,{}^{i+1}\hat{Z}_{i+1} \tag{5}$$

$$^{i+1}\dot{v}_{i+1} = {}^{i+1}_{i}R\left({}^{i}\dot{\omega}_i \times {}^{i}P_{i+1} + {}^{i}\omega_i \times \left({}^{i}\omega_i \times {}^{i}P_{i+1}\right) + {}^{i}\dot{v}_i\right) \tag{6}$$

$$\begin{aligned} {}^{i+1}\dot{v}_{Ci+1} = {}& {}^{i+1}\dot{\omega}_{i+1} \times {}^{i+1}P_{Ci+1} \\ & + {}^{i+1}\omega_{i+1} \times \left({}^{i+1}\omega_{i+1} \times {}^{i+1}P_{Ci+1}\right) + {}^{i+1}\dot{v}_{i+1} \end{aligned} \tag{7}$$

where ${}^{i+1}_{i}R$ is a rotation matrix and ${}^{i}P_{i+1}$ is the vector from the origin of frame *{i+1}* respect to reference frame *{i}*, ${}^{i}\omega_i$ and ${}^{i}v_i$ denote, respectively, the angular velocity of link frame *{i}* and linear velocity of its origin respect to reference frame *{i}* . Also, considering the vertical planar motion $\hat{Z}$ will be $\{0,0,1\}^T$. C is center of gravity parameters. We need ${}^{0}\dot{v}_3$ for friction effect calculations.

$$^{0}\dot{v}_3 = {}^{0}_{1}R\;{}^{1}_{2}R\;{}^{2}_{3}R\;{}^{3}\dot{v}_3 \tag{8}$$

The mechanism M_2 is modeled similarly to the close loop model of M_1 with the difference that M_2 has an extra link. It means that this mechanism has exactly 3-DOF [4]. Position, angular velocity and angular acceleration of joint 4 are calculated like sub mechanism M_1 [4].

3 Dynamics Formulation

The dynamic formulation of gait design involves computation of torque values are needed to achieve the desired joint motion. Rectilinear motion may appear in two modes: open loop and close loop. In open loop model, manipulator has a motion in space and does not contact with ground. Hence the friction effect doesn't exist. Therefore simple recursive Newton-Euler method can be applied. In close loop friction exists and this method can't be applied any more. In this paper we focus on close loop. Free body diagrams for links of sub mechanism M_1 in the close loop configuration are presented in Figs. 4, 5 and 6.

Free body diagrams of M_2 sub mechanism are completely similar to M_1.

Newton-Euler outward and inward recursive dynamic relations are required to complete our analysis about open loop mode. These formulas are as follows:

Outward iteration:

$$^{i+1}F_{i+1} = m_{i+1}\,{}^{i+1}\dot{v}_{Ci+1} \tag{9}$$

$$^{i+1}N_{i+1} = {}^{Ci+1}I_{i+1}\,{}^{i+1}\dot{\omega}_{i+1} + {}^{i+1}\omega_{i+1} \times {}^{Ci+1}I_{i+1}\,{}^{i+1}\omega_{i+1} \tag{10}$$

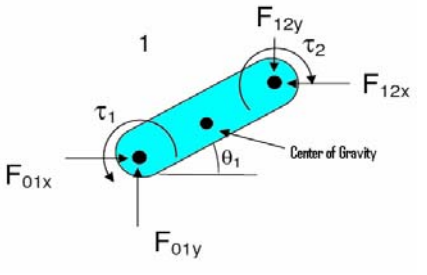


Fig. 4. FBD of first link of sub mechanism M_1

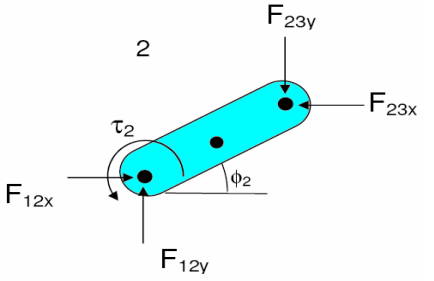


Fig. 5. FBD of second link of sub mechanism M_1

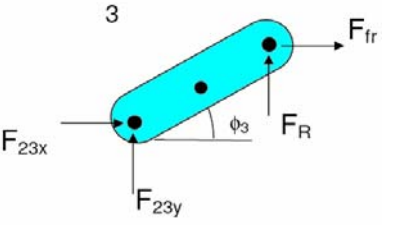


Fig. 6. FBD of third link of sub mechanism M_1

Inward iteration:

$$ {}^{i}f_i = {}^{i+1}_{i+1}R\ {}^{i+1}f_{i+1} + {}^{i}F_i \tag{11} $$

$$ {}^{i}n_i = {}^{i}N_i + {}^{i+1}n_{i+1} + {}^{i+1}P_{Ci+1} \times {}^{i}F_i + {}^{i}P_{i+1} \times_{i+1} {}^{i}R^{i+1}f_{i+1} \tag{12} $$

F,f and *N,n* are forces and torques respectively that act on links. These well- known parameters were described in [9] completely.

The Newton-Euler equations that will be applied to these FBDs (in close loop mode) are as follows:

$$ \Sigma F = m\dot{v} \tag{13} $$

$$ \Sigma M = I\dot{\omega} \tag{14} $$

After applying these equations to sub mechanism M_1 we have:

$$ F_{(i-1)ix} - F_{i(i+1)x} = m\dot{v}_{ix} \tag{15} $$

$$ F_{(i-1)iy} - F_{i(i+1)y} = m\dot{v}_{iy} \tag{16} $$

$$\tau_i - \tau_{i+1} + \frac{1}{2} F_{(i-1)ix} \sin \varphi_i + \frac{1}{2} F_{i(i+1)x} \sin \varphi_i \\ - \frac{1}{2} F_{(i-1)iy} \cos \varphi_i - \frac{1}{2} F_{i(i+1)y} \cos \varphi_i = I \dot{\omega}_i \tag{17}$$

Eqs. (15) to (17) are for any link with the exception of the final link. Eq. (18) describes friction force that acts on tip of final link.

$$F_{fr} = -\mu F_R sign({}^0 v_{3tip}) \tag{18}$$

The governed equations for final link (with *i*=3 for this case) are as below:

$$F_{(i-1)ix} + F_{fr} = m \dot{v}_{ix} \tag{19}$$

$$F_{(i-1)iy} - F_R = m \dot{v}_{iy} \tag{20}$$

$$- \frac{1}{2} F_{(i-1)iy} \cos \varphi_i + \frac{1}{2} F_{(i-1)ix} \sin \varphi_i \\ + \frac{1}{2} F_R \cos \varphi_i - \frac{1}{2} F_{fr} \sin \varphi_i = I \dot{\omega}_i \tag{21}$$

Required torque values are obtained from these equations. Similar formulas can be generalized for sub mechanism M_2.

4 Trajectory Optimization

In this section we will generate trajectories for joints of manipulators by B-Splines and optimize them using genetic algorithm. Detailed description of trajectory generation and GA optimization will be discussed.

4.1 Trajectory Generation

For trajectory planning, we assume that the time history of the joint parameters follows the trajectory that will be described by a cubic B-Spline curve [6]. In addition, the mechanisms should begin and end at rest, thus the derivatives of each curve end points should be zero.

For generating a cubic B-Spline with 3 midpoints we will have:

$$\mathrm{P} = \left\{ \theta_0 , \mathrm{C}_1 , \mathrm{C}_2 , \mathrm{C}_3 , \theta_\mathrm{f} \right\}. \tag{22}$$

Where the parameters θ_i and θ_f are the beginning and end angles of the interval which are always 0, α, or - α and the interior control points, designated as "C_i" are the free variables that change the geometry of the B-Spline curve.

Trajectory of each free joint can be parameterized by a set of control points, as:

$$\theta\left(t, \mathrm{P}\right) = \sum_{i=0}^{p} B_{i,k}\left(t\right) p_i \tag{23}$$

where k is the order of B-spline (for cubic B-spline $k = 4$) and $B_{i,k}(t)$ is the basic function given by the recurrence relations as:

$$B_{i,k}\left(t\right) = \frac{t - t_i}{t_{i+k-1} - t_k} B_{i,k-1}\left(t\right) + \frac{t_{i+k} - t}{t_{i+k} - t_{i+1}} B_{i+1,k-1}\left(t\right) \tag{24}$$

$$B_{i,1}\left(t\right) = \left\{ \begin{array}{c} 1 \ \ if \ \left(t_i \le t \le t_{i+1}\right) \\ , \quad 0 \ \ otherwise \end{array} \right\}$$

4.2 Genetic Algorithm

Genetic algorithm is a well-known optimization method that is extensively used in engineering field [10-11]. This method is a probabilistic search algorithm based on a model of natural evaluation. The algorithm has clearly demonstrated its capability to create good approximate solutions in complex optimization problems.

GA started with a set of random populations or chromosomes. The cost for each chromosome evaluated and sorted from minimum value to maximum or vice versa depending on the type of problem. Using the rate so-called selection rate, some of chromosomes were kept and others deleted.

Mating and crossover operators between kept chromosomes produce new generation of populations. Mutation operator is used as a result of avoiding of convergence in local extremes. Detailed description of this algorithm was presented in some references [11]. Our optimization problem is to optimize a parameter that we name it Effort and is presented as follow:

$$Effort = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{m} \left(\frac{\tau_{i(j+1)} + \tau_{ij}}{2} \right) \frac{t_f}{m} \tag{25}$$

t_f and m are time duration and number of sampling.

5 Results

In this section we will present optimized trajectories of joints (degree vs. time) for sub mechanism M_1 and M_2. Hence required torque values will be obtained. For running the program we use, 45° as the gait angle, 1 second as the gait step time, and a friction coefficient of $\mu = 0.4$. In addition, the length of the link is $l = 0.15$m and its mass is $m = 0.15$ kg. Also, the moment of inertia is calculated by treating the link as a thin rod.

5.1 Results for Sub Mechanisms

Optimized trajectory for joint 1 and 2 of sub mechanism M_1 are as Fig. 7 and 8.

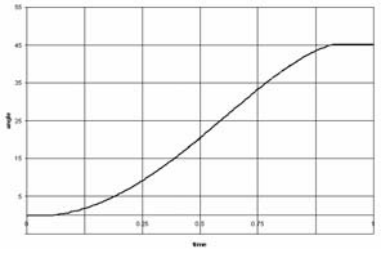


Fig. 7. Optimized angle vs. time for joint 1 of sub mechanism M_1

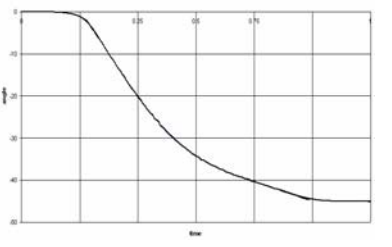


Fig. 8. Optimized angle vs. time for joint 2 of sub mechanism M_1

Torque values (N/m) for joints 1 and 2 of sub mechanism M_1 are as Fig. 9.

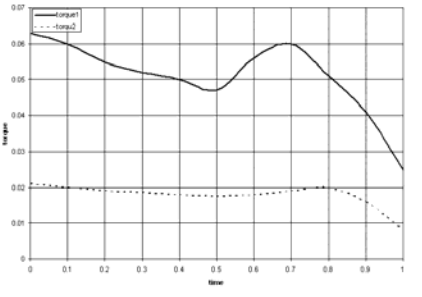


Fig. 9. Optimized torque values vs. time for joints 1 and 2 of sub mechanism M_1

Optimized trajectory for joints 1, 2 and 3 of sub mechanism M_2 are as Figs. 10,11 and 12.

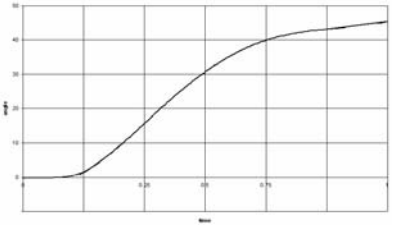


Fig. 10. Optimized angle vs. time for joint 1 of sub mechanism M_2

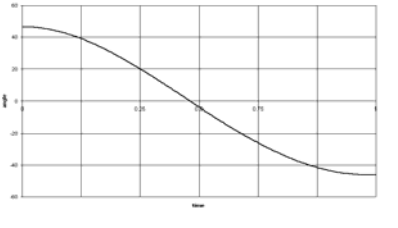


Fig. 11. Optimized angle vs. time for joint 2 of sub mechanism M_2

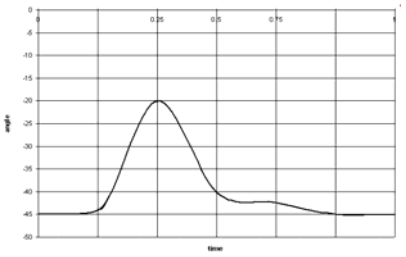


Fig. 13. Optimized angle vs. time for joint 3 of sub mechanism M_2

Torque values (N/m) for joints 1, 2 and 3 of sub mechanism M_2 are as Figure 13.

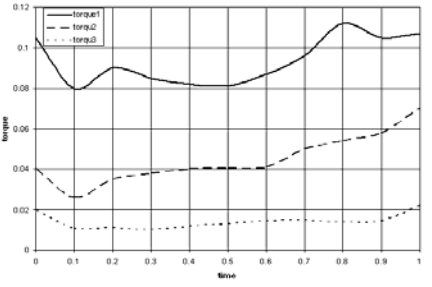


Fig. 14. Optimized torque values vs. time for joints 1 and 2 of sub mechanism M_2

6 Conclusions

In this paper we could present detailed dynamic formulation and optimization of consumed toque values using genetic algorithm.

Also time history and torque values were presented in figures and compared with each other.

References

1. Zhang, D., Hu, D., Shen, L., Xie, H.: Design of an artificial bionic neural network to control fish-robot's locomotion. J. Neurocomputing 71, 648–654 (2008)
2. Sameoto, D., Li, Y., Menon, C.: Multi-Scale Compliant Foot Designs and Fabrication for Use with a Spider-Inspired Climbing Robot. J. Bionic Eng. 5, 189–196 (2008)
3. Bayraktaroglu, Y.: Snake-like locomotion: Experimentations with a biologically inspired wheel-less snake robot. J. Mechanism and Machine Theory 44, 591–602 (2009)
4. Ghanbari, A., Rostami, A., Noorani, S.M., Seyyed Fakhrabadi, M.M.: Modeling and Simulation of Inchworm Mode Locomotion. In: Xiong, C.-H., Liu, H., Huang, Y., Xiong, Y.L. (eds.) ICIRA 2008. LNCS (LNAI), vol. 5314, pp. 617–624. Springer, Heidelberg (2008)

5. Chen, L., Ma, S., Wang, Y., Li, B., Duan, D.: Design and modelling of a snake robot in traveling wave locomotion. J. Mechanism and Machine Theory 42, 1632–1642 (2007)
6. William, B.: Desing, Analysis and Fabrication of a Snake-Inspired Robot With a Rectilinear Gait. J. MS Thesis, Maryland University (2006)
7. Ma, S., Li, W., Wang, Y.: A Simulation to Analyze Creeping Locomotion of a Snake-like Robot. In: Int. Conf. on Robotics & Automation (2001)
8. Merino, C.S., Tosunoglu, S.: Design of a Crawling Gait for a Modular Robot. In: Proc. of The 17th Florida Conf. on Recent Advances in Robotics (2004)
9. Craig, J.J.: Introduction to robotics: mechanics and control. Pearson Education, London (2005)
10. Ugur, A.: Path planning on a cuboid using genetic algorithms. J. Information Sciences 178, 3275–3287 (2008)
11. Haupt, R., Haupt, S.: Practical Genetic Algorithm. Wiley Publication, Chichester (2004)

Optimum Dynamic Modeling of a Wall Climbing Robot for Ship Rust Removal

Xingru Wang[1,2], Zhengyao Yi[1], Yongjun Gong[1], and Zuwen Wang[1]

[1] Institute of Shipping Electromechanics Equipment, Dalian Maritime University
[2] Cosco-Shipyard, China Ocean Shipping Company
116026, Dalian, China
yizhengyao@163.com

Abstract. This paper presents the optimum dynamic model method for wall climbing robot for ship rust removal. The robot includes a frame, two servomotors and reducers, and two crawlers with permanent magnets for walking and absorbing. Its main function is loading the cleaner which can remove the rust on ship surface by water jetting. Because of the water jetting and vacuum recycle rust, the wall climbing robot need load heavy pipelines. The dynamic models of the robot climbing and turning on the ship wall are established and optimized. The climbing driving torque equation is obtained by considering the change of the weight of pipelines load and the center of gravity position. Finally, the simulation analysis shows that the optimum dynamic model and the optimum method are reliable, and the parameters which have main effect on the robot dynamic characteristic are obtained.

Keywords: Wall climbing robot, Ship rust removal, Dynamic model, Model optimization, Simulation analysis.

1 Introduction

In recent years, the ship surface rust removal has become an increasingly popular topic in ship building and ship repairing area. Because water hydraulics has advantages of environmental friendly, cleanness, safety, readily available, inexpensive, and easily disposable, more and more scientific research institutions have begin to study many kinds of wall climbing robot for ship rust removal (WCRSRR) with water jetting, such as JPL of Caltech, DSIE in Cartagena of Spain, and some famous companies like Flow, Kamat, Hammelmann and so on [1-2].

The WCRSRR works with ultra-high pressure water jetting and vacuum recycling pump. There are two important auxiliary systems for the robot removing ship wall surface rust, they are an ultra-high pressure pump unit system (UHPPUS) which is designed with pressure 200-300MPa and power 110-145kW, and a vacuum rust residue recycling displaced system (VRRRDS). The working principle of WCRSRR is shown in Fig.1 [3]. The WCRSRR includes two crawlers with some permanent magnets for suction. Its main function is boarding a rust cleaner, which is a mechanical disc and can remove rust directly. The rust cleaner can flush, scrub, scrape the wall surface and collect sewage automatically [4].

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 623–631, 2009.

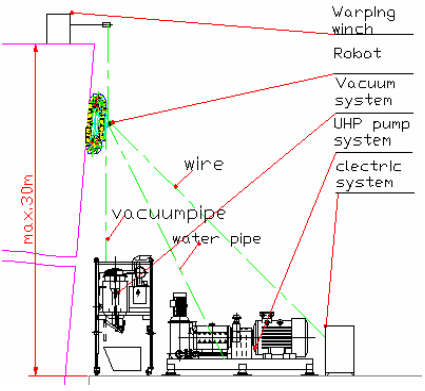


Fig. 1. Ultra-high pressure water jetting complete set of equipment (UHPWJCE)

Because of the UHPPUS and the VRRRDS, the WCRSRR carries loading of ultra-high pressure water pipes and vacuum residue recycle pipes about 30 meters, whose gross weight may reach 80Kg. So the total loading of aerial pipelines and robot body will exceed 150Kg [5-6]. Meanwhile, due to the robot climbs along the ship wall up and down, the carried aerial pipelines length changes momentarily. That is loading weight and loading gravity centre high changes momentarily with robot climbing high. Owing to the loading weight and the loading gravity centre are the main parameters which affect the driving torque, the study on the driving performance is necessary. Furthermore, the ship surface frame is complex. At different heights, the inclination angle of ship wall is different. There are some effects of the inclination angle of ship wall. Since the robot should walk along the ship wall flexibly, the study on of driving ability of WCRSRR is important.

From this argument, the purpose of the present paper is to set up the driving system dynamic model. The relationship of servomotor driving torque, reducer output torque and the mechanism parameters, and optimize the dynamic modeling is analyzed.

The rest of this paper is organized as follows. Section 2 represents the mechanical structure of WCRSRR. Section 3 introduces the process of the dynamic modeling building. In section 4, the optimization model is proposed and analyzed by simulation, and section 5 concludes the paper.

2 Mechanical Structure

2.1 Robot Structure and Prototype

The climbing robot structure is designed as shown in Fig.2. It is driven by two motors that are connected with each reductor. The adsorption mechanism is using two crawlers with 72 permanent magnets. A rust cleaner with some nozzles are set in the centre of the robot. The WCRSRR body consists of four parts: the adsorption mechanism, the walking mechanism, the driving mechanism and the frame. It is centrosymmetric structure, and the overall dimension is about 735×752×280 [mm^3]. The robot body weight is less than 90Kg, and the robot loading ability is about 80Kg.

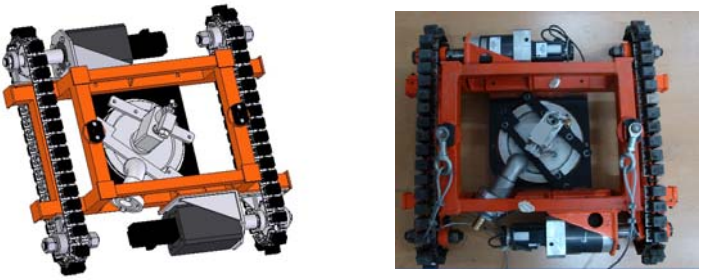


Fig. 2. Structure of WCRSRR

The adsorption mechanism is made of each crawler with 36 permanent magnets, and each crawler has 15 permanent magnets absorbing the ship wall. The walking mechanism contains crawlers, and sprocket shafts. The driving mechanism includes two servo motors and two reducers, and the motors are set as opposite angle.

3 Driving Model

3.1 Robot Climbing Model

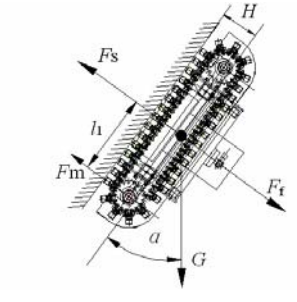


Fig. 3. Force analysis of the robot suction and climbing

In the Fig. 3, the water jetting kick force (F_f) and vacuum suction force (F_s) act on the working robot. Their equations are shown in (1) and (2) [7].

$$F_f = 0.745q \cdot \sqrt{p} \,. \tag{1}$$

$$F_s = \pi \cdot r^2 \cdot P \cdot \tag{2}$$

In (1) and (2), the q is the water flow, the p is the water pressure, the P is the vacuum force, and the r is the water jet radius.

From the Fig.3, when the robot climbs along the ship wall, the output torque from each reducer M_Q should satisfy

$$M_{\mathrm{Q}} - M_{\mathrm{f}} - M_{\mathrm{G}} = 0 \tag{3}$$

Where M_{G} is the torque that arisen from the half of the gravity, as shown in (4).

$$M_{\mathrm{G}} = \frac{GH\cos\alpha}{2} \tag{4}$$

The a is the inclination angle of ship wall. The M_{f} is the torque that arisen from the lowest magnet adsorption force (F_{m}) and the lowest holding power (N_1) in the lowest permanent magnet, as shown in Fig.3 and Fig.4.

$$M_{\mathrm{f}} = N_1 \cdot h \tag{5}$$

Due to the force F_{f}, F_{s} and the torque force (N_{n1}) which arisen from the robot gravity act on the working robot, the value of the holding power (N_1) is not equal to the value of magnet adsorption force (F_{m}). The pressure of the lowest magnet adsorption unit that arisen from robot gravity is shown in (6).

$$N_{\mathrm{n1}} = \frac{\frac{1}{2}G \cdot \cos\alpha \cdot H}{l_1} \tag{6}$$

Where G is the gravity of robot and loading, l_1 is the distance between lowest permanent magnet adscription unit and crawler centre, and H is the distance between loading gravity centre and ship wall. There are n permanent magnets adsorb on the ship wall. Then

$$N_1 = F_{\mathrm{m}} + \frac{F_{\mathrm{s}} - F_{\mathrm{f}} - G\sin\alpha}{n} + \frac{\frac{1}{2}G \cdot \cos\alpha \cdot H}{l_1} \tag{7}$$

$$M_{\mathrm{f}} = (F_{\mathrm{m}} + \frac{F_{\mathrm{s}} - F_{\mathrm{f}} - G\sin\alpha}{n} + \frac{G\cos\alpha H}{2l_1})h \tag{8}$$

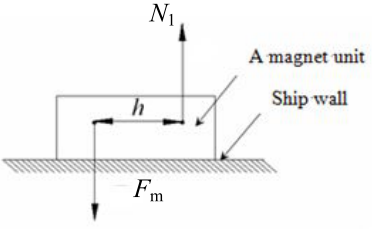


Fig. 4. The F_{mn} and N_{n} in the most lower permanent magnet

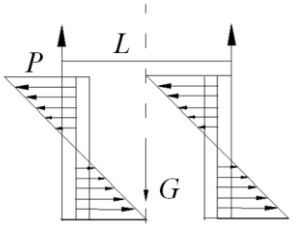


Fig. 5. Force analysis of the robot turning along the ship wall

According to (1), (2), (3) and (5), on the premise of avoiding gliding, the robot climbing driving torque equation of unilateral servomotor can be expressed as

$$M_{\mathrm{Q}} \geq (F_{\mathrm{mn}} + \frac{F_{\mathrm{s}} - F_{\mathrm{f}} - G\sin\alpha}{n} + \frac{GH\cos\alpha}{2l_1})h + \frac{GH\cos\alpha}{2} \quad (9)$$

By transforming formula (9), the driving torque equation of unilateral servomotor may be derived as

$$M_{\mathrm{Q}} \geq (\frac{nF_{\mathrm{m}} + F_{\mathrm{s}} - F_{\mathrm{f}} - G\sin a}{n})h + GH(\frac{1}{2} + \frac{h}{2l_1})\cos\alpha \quad (10)$$

3.2 Robot Turning Model

The robot turning is implemented by the velocity difference of the two crawlers. When the robot turn on the ship wall, it is should satisfy

$$M_{\mathrm{q}} \geq M_{\mathrm{Z}} \quad (11)$$

Where M_{q} is the turning driving moment of the each crawler, and the M_{Z} is the total turning resistance moment that arisen from every permanent magnet.

The analysis of the force which arisen from the two crawlers along the ship wall is shown in Fig. 5. The P is the traction force of every crawler, the L is the distance of the two crawlers. Where

$$M_{\mathrm{q}} = 2PL \quad (12)$$

$$P = M_{\mathrm{Q}} / R \quad (13)$$

The M_{q} can be written as

$$M_{\mathrm{q}} = 2\frac{M_{\mathrm{Q}}}{R}L \quad (14)$$

The R is the sprocket radius. When the robot turn, friction resistance moment M_{Z} can be written as

$$M_{\mathrm{Z}} = 2\sum_{i=1}^{n/2}[(N_i \cdot \mu) \cdot l_i] = 2 \cdot \mu \cdot \sum_{i=1}^{n/2}(N_i l_i) \quad (\mathbf{15})$$

Owing to the l_i is a variable, and its computational complexity is very big, the (5) is easily expressed as

$$N_i = F_{\mathrm{m}i} + \frac{F_{\mathrm{s}} - F_{\mathrm{f}} - G\sin\alpha}{n} \pm N_{\mathrm{n}i} \quad (16)$$

$$\sum_{i=1}^{n/2} N_i l_i = (F_{\mathrm{m}i} + \frac{F_{\mathrm{S}} - F_{\mathrm{f}} - G\sin a}{n}) \cdot \sum_{i=1}^{n/2} l_i \quad (17)$$

Then

$$M_Z = 2 \cdot (F_{mi} + \frac{F_S - F_f - G\sin a}{n}) \cdot \sum_{i=1}^{n/2} l_i \cdot \mu \quad (18)$$

Due to the (11), (14) and (18), the turning driving torque can be given by

$$M_Q \geq \left[(F_{mi} + \frac{F_S - F_f - G\sin a}{n}) \cdot (\sum_{i=1}^{n/2} l_i) \cdot \mu \cdot R \right] / L \quad (19)$$

3.3 Loading and Gravity Center Variable

The process of the robot climbing is very complex. According to the robot climbing driving torque equation of unilateral servomotor (9), we may know that the gravity of robot and loading (G) and the distance between the gravity center (H) are the most important parameters of the equation. When the robot is climbing, the parameters G and H change with the climbing high. Furthermore, the gravity of robot and loading G is very large. So the parameters G and H have some big influence on the climbing driving. The deep analysis of the (9) is necessary, and an optimization modeling basing on (9) should be given. The gravity equation is shown as

$$X_C = \frac{\sum m_i x_i}{m} \quad (20)$$

Where Xc is the whole machine gravity center position along the direction X, x_i is the parts gravity center position along the direction X, the m is whole machine gravity, and m_i is the parts mass. We assume that the ship wall is the gravity center base level, so there is gravity center as

$$H = \frac{\sum G_i H_i}{G} = \frac{G_1 \times H_1 + G_2 \times H_2}{G} \quad (21)$$

Where H_1 is the distance between the robot gravity centre and the ship wall, H_2 is the distance between the loading gravity centre and the ship wall, G_1 is the robot gravity, and G_2 is the loading gravity. Define the G_1, G_2 and G as

$$G_1 = G_{DJ} + G_K \quad (22)$$

$$G_2 = kl \quad (23)$$

$$G = kl + G_{DJ} + G_K \quad (24)$$

Where k is the whole specific gravity of the pipelines, l is the length of the aerial pipelines, G_{DJ} is the total weight of the servomotors and reducers, and the G_K is the total weight of the robot frame.

According to (21), the robot gravity center can be expressed as

$$H = \frac{(G_{DJ} + G_K) \times H_1 + kl \times H_2}{kl + G_{DJ} + G_K} \tag{25}$$

In Fig.6, the H_2 can be obtained as

$$H_2 = \frac{l}{2} \sin a + h_r \tag{26}$$

The h_r is the distance between the ship wall and the joint of loading pipelines on robot. By integrating the (24), (25) and (26) with (10), the robot climbing driving torque equation can be obtained as (27). And the robot turning driving torque equation can be written as (28).

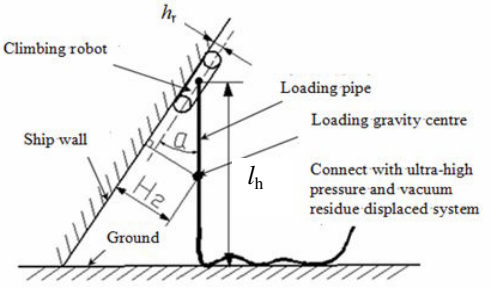


Fig. 6. Force analysis of the robot climbing the ship wall

$$M_Q \geq [nF_m + F_s - F_f - (kl + G_{DJ} + G_K) \sin a] \cdot \frac{h}{n} + [(G_{DJ} + G_K) \cdot H_1 + kl \cdot (\frac{l}{2} \sin a + h_r)] \cdot (\frac{1}{2} + \frac{h}{2l_1}) \cos \alpha \tag{27}$$

$$M_Q \geq \left[(F_m + \frac{F_S - F_f - (kl + G_{DJ} + G_K) \sin a}{n}) \cdot (\sum_{i=1}^{n/2} l_i) \cdot \mu \cdot R \right] / L \tag{28}$$

4 Optimization Model and Simulation Analysis

Generally, the geometry of ship wall may be expressed by hull lines drawing, which represents the every cross section shape from the bow to stern.

The inclination angle changes with ship wall curvature. And in (27) and (28), the G_{DJ} changes with the type of the servomotor and the reducer. According to (20), the H_1 changes with the G_{DJ}. In (27), the parameters F_m F_S F_f G_K h n k and h_r are fixed value, we can optimize the parameters. The G_{DJ} is an independent variable, and the H_1 is a dependent variable. According to the sampling parametric computation of the robot gravity center equation (25), the variation degree of the H_1 affected by G_{DJ} variation is only 0.051. The relative variable weight is small. So we assume the H_1 is a fixed value, which is not affected by G_{DJ} variation.

In (28), the parameters l_i R L μ F_m F_S F_f and n are fixed value, and only the α and its coefficient (kl+G_{DJ}+G_K) are variable. Furthermore, the term sinα is a minuend term. So it can be deduced that, in the range of °to 90°, when the α=0°, the minimum turning driving torque can be expressed as

$$Min(M_Q) \geq \left[(F_m + \frac{F_S - F_f}{n}) \cdot (\sum_{i=1}^{n/2} l_i) \cdot \mu \cdot R \right] / L \quad . \tag{29}$$

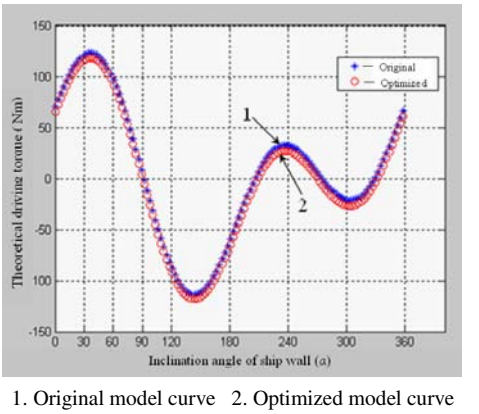
The fixed value in the above are set in (29), the robot turning driving torque that is more than 30N/m is obtained, which meets the requirements. In (27), when the climbing high is set as only 0 meter, the requirements demand the climbing driving torque is at least more than 88N/m. Therefore, comparing the turning driving torque and the climbing driving torque, the impact factor of turning driving torque is less than the climbing driving torque. In the process of studying the dynamic modeling, the climbing driving torque should be prior considered.

In the climbing driving torque equation (27), the parameters h=0.03m and n=30, the ratio is very small. So the dynamic modeling which may evaluate the driving characteristic is easily expressed as optimized driving torque equation (31).

$$M_Q \geq [(G_{DJ} + G_K)H_1 + kl(\frac{l}{2}\sin a + h_r)](\frac{1}{2} + \frac{h}{2l_1})\cos\alpha \quad . \tag{31}$$

Let the original robot climbing driving torque equation (27) compare with the optimized driving torque equation (31). In order to see the whole contrast effect, the range of the inclination angle of ship wall α is enlarged from °- 90°to 0°- 360°.

Define a robot climbing high as 5 meters, the tracking curve between optimized driving torque equation (31) and original robot climbing driving torque equation (27) is shown in Fig.7. The optimized model curve is consistent with the original model curve. The two curves have good overlap degree and small optimization error. It shows that the optimized model is reliable.



1. Original model curve 2. Optimized model curve

Fig. 7. The relationship between climbing high (CH) and the inclination angle of ship wall (α)

According to Fig.7, the magnet adsorption force F_m, the water jetting kick force F_f , vacuum suction force F_s and the number of permanent magnets adsorbs on the ship wall n have little effect on robot driving characteristic. However, the robot body weight $G_{DJ}+G_K$, climbing high (CH) , the distance between the robot gravity centre and the ship wall H_1 and the inclination angle of ship wall (α) have great effect on robot driving characteristic.

5 Conclusions

This paper presents the research on the dynamic modeling.

In the condition of the loading gravity and gravity center variation, the WCRSRR dynamic modeling is established. The driving torque modeling is optimized. The simulation shows that the optimized modeling is reliable, and the magnet adsorption force (F_m), the water jetting kick force (F_f) , vacuum suction force (F_s) and the number of permanent magnets adsorbs on the ship wall (n) have little effect on robot dynamic characteristic. However, the gravity of robot and loading (G), climbing high (CH) , the distance between the robot gravity centre and the ship wall (H_1) and the inclination angle of ship wall (α) have great effect on robot dynamic characteristic.

Acknowledgments

This work was supported in part by National Natural Science Foundation of China under Grant 50805011, the Basic Research Item of Ministry of Transport of the People Republic of China under Grant 225030, and the Major Project of Cosco-Shipyard under Grant G2-07-0060.

References

1. Maniere, E., Simmons, R.: Architecture, the backbone of robotic system. In: 2000 IEEE International Conference on Robotics& Automation, San Francisco, USA, pp. 505–513 (2000)
2. Shen, W.M., Gu, J., Shen, Y.J.: Proposed wall climbing robot with permanent magnetic tracks for inspecting oil tanks. In: 2005 IEEE International Conference on Mechatronics and Automation, vol. 4, pp. 2072–2077 (2005)
3. Wang, X., Yi, Z., Gong, Y., et al.: Ultra-high Pressure Water Jetting Removal Rust Model and Analysis of Removal Rust Capability. In: Proceedings of the 7th, ICFP HANGZHOU, pp. 503–506 (2009)
4. Wang, X., Gong, Y., Yi, Z., et al.: A Study on Ultra-high Pressure Water Jet Wall Climbing Robot for Removal Rust in Vessels. In: Proceedings of the 7th, JFPS TOYAMA, vol. 3(2), pp. 406–409 (2008)
5. Ross, B., Barcs, J.: A semi-autonomous robot for stripping paint from large vessels. J. The International Journal of Robotics 22(7), 617–626 (2003)
6. Cohen, B.Y.: Residue detection for real-time removal of paint from metallic surfaces ship hull maintenance robot. In: The SPIE 8th Annual International Symposium on Smart Structures and Materials, pp. 4335–4320 (2003)
7. Maia, D.: Alternatives to conventional methods of surface preparation for ship repairs. J. Journal of Protective Coatings and Linings 17(5), 31–39 (2000)

Stiffness and Singularity Analysis of 2SPS+2RPS Parallel Manipulator by Using Different Methods

Bo Hu, Yi Lu, and Jiayin Xu

College of Mechanical Engineering, Yanshan University
066004, Qinhuangdao, China
hbz0001@yahoo.com.cn

Abstract. The Stiffness and singularity of a 2SPS+2RPS parallel manipulator is analyzed from different viewpoints in this paper. First, a new method for solving statics is presented, and the generalized force is derived. The 4×4 stiffness matrix which shows the relation of generalized force and generalized coordinates is derived. Second, base on observing the constrained forces in RPS legs, the 6×6 Jacobian matrix is derived and the statics is solved by another method. By using the principle of virtual work, a 6×6 stiffness matrix is derived with the constrained forces considered in stiffness analysis. Third, by locking the four actuators, the singularity is analyzed from different point of view.

Keywords: Parallel manipulator, stiffness, singularity.

1 Introduction

With the development of research of parallel robot, some lower mobility PMs has attracted many scholars [1]. Huang Z. [2], Kong X. [3, 4] synthesized four degree of freedom (DOF) parallel manipulators (PMs) by using screw theory. P. Richard and C. M. Gosselin [5] studied a kinematic analysis and prototyping of a partially decoupled 4-DOF 3T1R PM. M. Zoppi, D. Zlatanov and C. M. Gosselin [6] investigated the analytical kinematics models and special geometries of a class of 4- DOF parallel mechanisms. Lu and Hu [7] studied the 4-DOF 3SPS+UPR PM with passive leg.

In the singularity aspect , J. P. Merlet [8] analyzed the singular configurations of parallel manipulators by using grassman geometry. J. M. Rico et al[9] studied kinematics and singularity analyses of a four degrees of freedom PM by using the theory of screws. Liu C. H. et al [10] studied the direct singular positions of the parallel manipulator Tricept. Kong x. [11] proposed a unified and simplified approach for performing uncertainty singularity analysis of parallel manipulators with 3-XS structure. Zhao J.S. [12] analyzed the singularity of spatial parallel manipulator with terminal constraints. In the aspect of stiffness, Zhang Dan [13, 14] established the kinetostatic modeling and stiffness modeling for some PMs with three to five degrees of freedom. Han S. and Fang y.[15] studied the stiffness of a 4-DOF parallel manipulator by using screw theory.

The 4-DOF PMs are more complex than 3-DOF PMs, and the redundant constraints always exited in 4-DOF PMs especially for symmetrical PMs. The PMs with passive legs can avoid redundant constraints but bring more probabilities of interference. In

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 632–644, 2009.

this paper, a novel 2RPS+2SPS PM with a triangular moving platform and a quadrangular fixed base is present. Compared with symmetrical PMs, this PM has no redundant constraints. It has some merit such as easy mounting, simple structure and simple liner actuators. Compared with PMs with passive legs [7, 14], the PM has less probabilities of interference. The 2RPS+2SPS PM has some potential applications for the 4-DOF PM machine tools, the 4-DOF sensor, the surgical manipulator, the tunnel borer, the barbette of war ship, and the satellite surveillance platform. This paper is focus on solving the stiffness and singularity of 2SPS+2RPS PM.

2 Description of 2SPS+2RPS PM

A 2SPS+2RPS PM is composed of a moving platform m, a fixed base B, two SPS type legs and two RPS type legs (see Fig. 1).Where, m is an equilateral triangle $a_1a_2a_3$ with the sides $l_i = l(i=1,2,3)$ and a central point o; B is an equilateral quadrangle $A_1A_2A_3A_4$ with the sides $L_i=L(i=1,2,3,4)$ and a central point O.

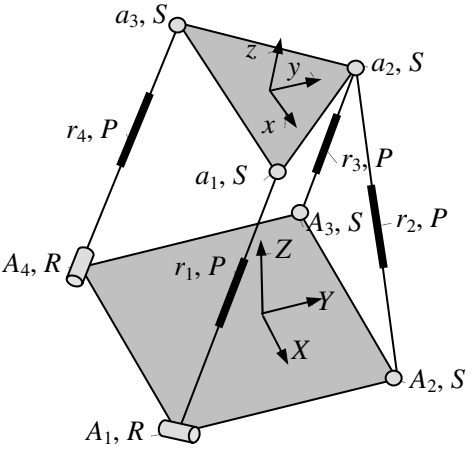


Fig. 1. The sketch of 2SPS+2RPS PM

Each of the RPS-type active leg connects B with m by a S joint at point $A_i(i=1,4)$, a S joint at point a_j(j=1,3), an active leg r_i (i=1,4) with prismatic joint P_i. The R joints are located in the Base. Each of the SPS-type active leg connects B with m by a S joint at point $A_i(i=2,3)$, a S joint at point a_2 and an active leg r_i (i=2,3) with prismatic joint $P_i(i=2,3)$. $\{m\}$ is the coordinate frame o-xyz fixed onto m at o and $\{B\}$ is the coordinate frame O-XYZ fixed onto B at o.

The coordinates A_i (i=1, 2, 3, 4), o in O-XYZ and a_i (i=1, 2, 3, 4) in o-xyz can be expressed in matrix as follows

$$\boldsymbol{A}_1 = \begin{bmatrix} X_{A1} \\ Y_{A1} \\ Z_{A1} \end{bmatrix} = \begin{bmatrix} E \\ -E \\ 0 \end{bmatrix}, \boldsymbol{A}_2 = \begin{bmatrix} X_{A2} \\ Y_{A2} \\ Z_{A2} \end{bmatrix} = \begin{bmatrix} E \\ E \\ 0 \end{bmatrix}, \boldsymbol{A}_3 = \begin{bmatrix} X_{A3} \\ Y_{A3} \\ Z_{A3} \end{bmatrix} = \begin{bmatrix} -E \\ E \\ 0 \end{bmatrix}, \boldsymbol{A}_4 = \begin{bmatrix} X_{A4} \\ Y_{A4} \\ Z_{A4} \end{bmatrix} = \begin{bmatrix} -E \\ -E \\ 0 \end{bmatrix} \quad (1a)$$

The coordinates a_i in *O-xyz* can be expressed in matrix as

$$ {}^{m}\boldsymbol{a}_1=\begin{bmatrix}x_{a1}\\ y_{a1}\\ z_{a1}\end{bmatrix}=\begin{bmatrix}qe/2\\ -q/2\\ 0\end{bmatrix},{}^{m}\boldsymbol{a}_2=\begin{bmatrix}x_{a2}\\ y_{a2}\\ z_{a2}\end{bmatrix}=\begin{bmatrix}0\\ e\\ 0\end{bmatrix},{}^{m}\boldsymbol{a}_3=\begin{bmatrix}x_{a3}\\ y_{a3}\\ z_{a3}\end{bmatrix}=\begin{bmatrix}-qe/2\\ -e/2\\ 0\end{bmatrix} \tag{1b} $$

The coordinates a_i in *O-XYZ*can be expressed in matrix as

$$ \boldsymbol{a}_i={}^{B}_{m}\mathbf{R}\,{}^{m}\boldsymbol{a}_i+\boldsymbol{o} \tag{2} $$

Where $E=L/2$ $q=3^{1/2}$. e is the length of vector $\boldsymbol{oa}_i$ (i=1,2,3). X_o, Y_o, Z_o are the position of the center of m, ${}^{B}_{m}\mathbf{R}$ is a rotation transformation matrix.

$$ {}^{B}\boldsymbol{o}=[X_o \quad Y_o \quad Z_o]^T,{}^{B}_{m}\mathbf{R}=[\boldsymbol{x} \quad \boldsymbol{y} \quad \boldsymbol{z}] $$

$$ \boldsymbol{x}=[x_l \quad x_m \quad x_n]^T,\boldsymbol{y}=[y_l \quad y_m \quad y_n]^T,\boldsymbol{z}=[z_l \quad z_m \quad z_n]^T $$

$\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ are the vector of the *x*-,*y*-,and *z*-axes of the moved reference frame.

Let R_1 and R_2 be the R joints in the first and fourth limb respectively, θ_1 be the angle between R_1 and X and θ_2 be the angle between R_2 and X. The unit vector of R_1 and R_2 can be expressed as:

$$ \boldsymbol{R}_1=[c_{\theta1}\ s_{\theta1}\ 0]^T,\ \boldsymbol{R}_2=[c_{\theta2}\ s_{\theta2}\ 0]^T \tag{3} $$

As r_1 $\boldsymbol{R}_1$, r_4 $\boldsymbol{R}_2$, lead to

$$ (aex_l/2-ey_l/2+X_o-E)c_{\theta1}+(aex_m/2-ey_m/2+Y_o+E)s_{\theta1}=0 \tag{4} $$

$$ (-aex_l/2-ey_l/2+X_o+E)c_{\theta2}+(-aex_m/2-ey_m/2+Y_o+E)s_{\theta2}=0 \tag{5} $$

From Eqs.(4) and (5), lead to

$$ X_o=\left[-qex_l s_{(\theta1+\theta2)}/2+ey_l s_{(\theta2-\theta1)}/2+Es_{(\theta1+\theta2)}-qex_m s_{\theta1}s_{\theta2}\right]/s_{(\theta2-\theta1)} \tag{6} $$

$$ Y_o=\left[qex_m s_{(\theta2+\theta1)}/2-ey_m s_{(\theta1-\theta2)}/2+Es_{(\theta1-\theta2)}+(qex_l-2E)c_{\theta1}c_{\theta2}\right]/s_{(\theta2-\theta1)} \tag{7} $$

Let three Euler angles α, β, λ rotate about the *y*-, *z*-, and *x*-axis of the moved reference frame. Thus, the rotational transformation matrix can be expressed as below.

$$ {}^{B}_{m}\mathbf{R}=\begin{bmatrix}c_\alpha c_\beta & -c_\alpha s_\beta c_\lambda+s_\alpha s_\lambda & c_\alpha s_\beta s_\lambda+s_\alpha c_\lambda\\ s_\beta & c_\beta c_\lambda & -c_\beta s_\lambda\\ -s_\alpha c_\beta & s_\alpha s_\beta c_\lambda+c_\alpha s_\lambda & -s_\alpha s_\beta s_\lambda+c_\alpha c_\lambda\end{bmatrix} \tag{8} $$

From Eqs.(6),(7) and (8), lead to

$$ X_o=\left[-qes_{(\theta1+\theta2)}c_\alpha c_\beta+es_{(\theta2-\theta1)}(-c_\alpha s_\beta c_\lambda+s_\alpha s_\lambda)+2Es_{(\theta1+\theta2)}-2qes_{\theta1}s_{\theta2}s_\beta\right]/2s_{(\theta2-\theta1)} \tag{9} $$

$$Y_o = \left| qes_{(\theta 2+\theta 1)} s_\beta - es_{(\theta 1-\theta 2)} c_\beta c_\lambda + Es_{(\theta 1-\theta 2)} + 2(qec_\alpha c_\beta - 2E) c_{\theta 1} c_{\theta 2} \right| / 2s_{(\theta 2-\theta 1)} \quad (10)$$

The inverse kinematics can be expressed as

$$r_i^{\,2} = \left| {}^{B}\boldsymbol{a}_i - {}^{B}\boldsymbol{A}_i \right|^2 \quad (11)$$

When given the independent parameters (α, β,λ, Z_o), X_o and Y_o can be solved from (9) and (10), Then from Eq.(11),the inverse kinematics can be solved.

3 Stiffness Analysis

3.1 Jacobian Matrix Analysis

3.1.1 4×4 Jacobian Matrix Analysis

Let $\boldsymbol{v}$ and $\boldsymbol{\omega}$ be the linear velocity and angular velocity of moving platform, v_{ri}(i=1,2,3,4) be the velocity of r_i.

The loop equation of each leg can expressed as:

$$\boldsymbol{OA}_i + \boldsymbol{A}_i\boldsymbol{a}_i = \boldsymbol{Oo} + \boldsymbol{oa}_i \quad (12)$$

Taking the derivative of both sides of Eq.(12) by time, we obtain:

$$v_{ri}\boldsymbol{\delta}_i + \boldsymbol{\omega}_{ri} \times r_i\boldsymbol{\delta}_i = \boldsymbol{v} + \boldsymbol{\omega} \times \boldsymbol{e}_i \quad (13)$$

Where $\boldsymbol{\omega}_{ri}$ be the angle velocity of r_i

$$\boldsymbol{\delta}_i = (\boldsymbol{a}_i - \boldsymbol{A}_i) / \left| \boldsymbol{a}_i - \boldsymbol{A}_i \right|, \boldsymbol{e}_i = \boldsymbol{a}_i - \boldsymbol{o}$$

Taking the dot product of both sides of Eq.(13), the velocity along active leg r_i (i=1,2,3,4) can be derived as

$$v_{ri} = \boldsymbol{\delta}_i \cdot \boldsymbol{v} + (\boldsymbol{e}_i \times \boldsymbol{\delta}_i)\boldsymbol{\omega} = [\boldsymbol{\delta}_i^T \quad \boldsymbol{e}_i^T \times \boldsymbol{\delta}_i^T] \begin{bmatrix} \boldsymbol{v} \\ \boldsymbol{\omega} \end{bmatrix} \quad (14)$$

Eq (14) can expressed in a matrix form as follows

$$\boldsymbol{v}_r = \mathbf{J}_\alpha \begin{bmatrix} \boldsymbol{v} \\ \boldsymbol{\omega} \end{bmatrix} \quad \boldsymbol{v}_r = [v_{r1}\ v_{r2}\ v_{r3}\ v_{r4}]^T \quad (15)$$

$$\mathbf{J}_\alpha = \begin{bmatrix} \boldsymbol{\delta}_1^T & \boldsymbol{e}_1^T \times \boldsymbol{\delta}_1^T \\ \boldsymbol{\delta}_2^T & \boldsymbol{e}_2^T \times \boldsymbol{\delta}_2^T \\ \boldsymbol{\delta}_3^T & \boldsymbol{e}_3^T \times \boldsymbol{\delta}_3^T \\ \boldsymbol{\delta}_4^T & \boldsymbol{e}_4^T \times \boldsymbol{\delta}_4^T \end{bmatrix} \quad (16)$$

Taking the derivative of Eqs.(9) and (10) with respect to time, leads to

$$v_x = \dot{X}_o = \frac{\partial X_o}{\partial \alpha}\dot{\alpha} + \frac{\partial X_o}{\partial \beta}\dot{\beta} + \frac{\partial X_o}{\partial \lambda}\dot{\lambda} + \frac{\partial X_o}{\partial Z_o}\dot{Z}_o \quad (17)$$

$$v_y = \dot{X}_o = \frac{\partial Y_o}{\partial \alpha}\dot{\alpha} + \frac{\partial Y_o}{\partial \beta}\dot{\beta} + \frac{\partial Y_o}{\partial \lambda}\dot{\lambda} + \frac{\partial Y_o}{\partial Z_o}\dot{Z}_o \tag{18}$$

From Eqs.(17), (18),lead to

$$\begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix} = \mathbf{J}_{o1}\begin{bmatrix} \dot{\alpha} \\ \dot{\beta} \\ \dot{\lambda} \\ \dot{Z}_o \end{bmatrix}, \mathbf{J}_{o1} = \begin{bmatrix} \partial X_o/\partial\alpha & \partial X_o/\partial\beta & \partial X_o/\partial\lambda & \partial X_o/\partial Z_o \\ \partial Y_o/\partial\alpha & \partial Y_o/\partial\beta & \partial Y_o/\partial\lambda & \partial Y_o/\partial Z_o \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{19}$$

The angle velocity can be written as follows:

$$\begin{bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{bmatrix} = \boldsymbol{R}_\alpha\dot{\alpha} + \boldsymbol{R}_\beta\dot{\beta} + \boldsymbol{R}_\lambda\dot{\lambda} = \mathbf{J}_{o1}\begin{bmatrix} \dot{\alpha} \\ \dot{\beta} \\ \dot{\lambda} \\ \dot{Z}_o \end{bmatrix}, \mathbf{J}_{o2} = \begin{bmatrix} 0 & s_\alpha & c_\alpha c_\beta & 0 \\ 1 & 0 & s_\beta & 0 \\ 0 & c_\alpha & -s_\alpha c_\beta & 0 \end{bmatrix} \tag{20}$$

Where, $\boldsymbol{R}_\alpha, \boldsymbol{R}_\beta, \boldsymbol{R}_\gamma$ are the axes of Euler angle α, β, λ

$$\boldsymbol{R}_\alpha = [0 \quad 1 \quad 0]^T \tag{21a}$$

$$\boldsymbol{R}_\beta = \begin{bmatrix} c_\alpha & 0 & s_\alpha \\ 0 & 1 & 0 \\ -s_\alpha & 0 & c_\alpha \end{bmatrix}\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} s_\alpha \\ 0 \\ c_\alpha \end{bmatrix} \tag{21b}$$

$$\boldsymbol{R}_\lambda = \begin{bmatrix} c_\alpha & 0 & s_\alpha \\ 0 & 1 & 0 \\ -s_\alpha & 0 & c_\alpha \end{bmatrix}\begin{bmatrix} c_\beta & -s_\beta & 0 \\ s_\beta & c_\beta & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} c_\alpha c_\beta \\ s_\beta \\ -s_\alpha c_\beta \end{bmatrix} \tag{21c}$$

By combining Eq.(19) and (20), the velocity of moving platform can be obtain as

$$\boldsymbol{V} = \mathbf{J}_o\dot{\boldsymbol{q}}, \mathbf{J}_o = \begin{bmatrix} \mathbf{J}_{o1} \\ \mathbf{J}_{o2} \end{bmatrix} \tag{22}$$

Where, $\boldsymbol{V}=[v_x\ v_y\ v_z\ \omega_x\ \omega_y\ \omega_z]^T$, $\dot{\boldsymbol{q}} = \begin{bmatrix} \dot{\alpha} & \dot{\beta} & \dot{\lambda} & \dot{Z}_o \end{bmatrix}^T$

From Eqs.(15) and (22),lead to

$$\boldsymbol{v}_r = \mathbf{J}_{4\times4}\dot{\boldsymbol{q}}, \quad \dot{\boldsymbol{q}} = \mathbf{J}^{-1}_{4\times4}\boldsymbol{v}_r \tag{23}$$

$$\mathbf{J}_{4\times4} = \mathbf{J}_\alpha\mathbf{J}_\mathbf{o} \tag{24}$$

$\mathbf{J}_{4\times4}$ is the 4×4 form Jacobian matrix.

3.1.2 6×6 Jacobian Matrix Analysis

In the 2SPS+2RPS parallel manipulator, the workloads can be simplified as a wrench $\boldsymbol{F}$ applied onto m at o. Where, $\boldsymbol{F}$ is a six dimension vector which has three force elements and three torque elements.

Based on observe method for finding constrained force/torque [25], two constrained forces $\boldsymbol{F}_{\mathrm{p1}}$ and $\boldsymbol{F}_{\mathrm{p2}}$ exerted on r_1 at a_1 and on r_4 at a_3 respectively can be found. From the geometric constraints, the unit vectors $\boldsymbol{f}_j$ of $\boldsymbol{F}_{pj}$ (j=1, 2) are determined as

$$\boldsymbol{f}_1=\boldsymbol{R}_1,\ \boldsymbol{f}_2=\boldsymbol{R}_2 \tag{25}$$

Since the constrained forces/torques do no work to the point o, lead to

$$\begin{aligned} &F_{Pi}\boldsymbol{f}_i\cdot\boldsymbol{v}+(F_{Pi}\boldsymbol{d}_i\times\boldsymbol{f}_i)\cdot\boldsymbol{\omega}=0\\ &\boldsymbol{f}_i\cdot\boldsymbol{v}+(\boldsymbol{d}_i\times\boldsymbol{f}_i)\cdot\boldsymbol{\omega}=0 \end{aligned} \tag{26}$$

Where, $\boldsymbol{d}_1=\boldsymbol{o}-\boldsymbol{a}_1$, $\boldsymbol{d}_2=\boldsymbol{o}-\boldsymbol{a}_3$.

The inverse/forward velocities can be derived from Eqs.(23) and (26) as follows:

$$\boldsymbol{V}_r=\mathbf{J}_{6\times6}\boldsymbol{V},\ \boldsymbol{V}=\mathbf{J}_{6\times6}^{-1}\boldsymbol{V}_r,\ \mathbf{J}_{6\times6}=\begin{bmatrix}\mathbf{J}_\alpha\\ \mathbf{J}_v\end{bmatrix},\mathbf{J}_v=\begin{bmatrix}\boldsymbol{f}_1^T & (\boldsymbol{d}_1\times\boldsymbol{f}_1)^T\\ \boldsymbol{f}_2^T & (\boldsymbol{d}_2\times\boldsymbol{f}_2)^T\end{bmatrix}. \tag{27}$$

Where, $\mathbf{J}_{6\times6}$ is a 6×6 Jacobian matrix. $\boldsymbol{V}_r=[v_{r1}\ v_{r2}\ v_{r3}\ v_{r4}\ 0\ 0]^T$

3.2 Stiffness Medeling

3.2.1 First Method

Using the principle of virtual work, we can obtain

$$\boldsymbol{F}_a^r\boldsymbol{v}_r=-\boldsymbol{F}^T\boldsymbol{V} \tag{28}$$

Where, $\boldsymbol{F}_a=[F_{a1},\ F_{a2},\ F_{a3},\ F_{a4}]^T$, with F_{ai} (i=1,2,3,4) denotes the active force of actuators. $\boldsymbol{F}$ is the loads on the center of platform $\boldsymbol{F}=[F_x,\ F_y\ \ F_z\ \ T_x,\ T_y\ T_z]^T$.

Substituting Eqs. (22), (23) into (28), lead to

$$\boldsymbol{F}_a^T\boldsymbol{v}_r=-\boldsymbol{F}^T\boldsymbol{V}=-\boldsymbol{F}^T\ \mathbf{J}_o\,\dot{\boldsymbol{q}}=-\boldsymbol{F}^T\ \mathbf{J}_o\,\mathbf{J}_{4\times4}^{-1}\boldsymbol{v}_r \tag{29}$$

Then the formula for solving statics can be derived

$$\boldsymbol{F}_a=-(\mathbf{J}_o\,\mathbf{J}_{4\times4}^{-1})^T\boldsymbol{F} \tag{30}$$

Eq.(30) is a novel formulae for statics analysis for this 2SPS+2RPS PM, $(\mathbf{J}_o\,\mathbf{J}_{4\times4}^{-1})^T$ is a 4×6 form matrix, and this methodology is fit for other lower mobility PMs. In another aspect, from Eqs.(22) and (29), lead to

$$\boldsymbol{F}_a^{\mathrm{r}}\boldsymbol{v}_r=-\begin{bmatrix}F_x\\F_y\\F_z\\T_x\\T_y\\T_z\end{bmatrix}^T\begin{bmatrix}\dot{X}_o\\\dot{Y}_o\\\dot{Z}_o\\\dot{\psi}_x\\\dot{\psi}_y\\\dot{\psi}_z\end{bmatrix}=-\begin{bmatrix}F_x\frac{\partial X_o}{\partial\alpha}+F_y\frac{\partial Y_o}{\partial\alpha}+F_z\frac{\partial Z_o}{\partial\alpha}+T_x\frac{\partial\psi_x}{\partial\alpha}+T_y\frac{\partial\psi_y}{\partial\alpha}+T_z\frac{\partial\psi_z}{\partial\alpha}\\F_x\frac{\partial X_o}{\partial\beta}+F_y\frac{\partial Y_o}{\partial\beta}+F_x\frac{\partial Z_o}{\partial\beta}+T_x\frac{\partial\psi_x}{\partial\beta}+T_y\frac{\partial\psi_y}{\partial\beta}+T_z\frac{\partial\psi_z}{\partial\beta}\\F_x\frac{\partial X_o}{\partial\lambda}+F_y\frac{\partial Y_o}{\partial\lambda}+F_x\frac{\partial Z_o}{\partial\lambda}+T_x\frac{\partial\psi_x}{\partial\lambda}+T_y\frac{\partial\psi_y}{\partial\lambda}+T_z\frac{\partial\psi_z}{\partial\lambda}\\F_x\frac{\partial X_o}{\partial Z_o}+F_y\frac{\partial Y_o}{\partial Z_o}+F_z\frac{\partial Z_o}{\partial Z_o}+T_x\frac{\partial\psi_x}{\partial Z_o}+T_y\frac{\partial\psi_y}{\partial Z_o}+T_z\frac{\partial\psi_z}{\partial Z_o}\end{bmatrix}\begin{bmatrix}\dot{\alpha}\\\dot{\beta}\\\dot{\lambda}\\\dot{Z}_o\end{bmatrix}=-\begin{bmatrix}F_x\\F_y\\F_z\\T_x\\T_y\\T_z\end{bmatrix}^T\mathbf{J}_o\begin{bmatrix}\dot{\alpha}\\\dot{\beta}\\\dot{\lambda}\\\dot{Z}_o\end{bmatrix}\tag{31}$$

From Eqs.(23) and (31), the following equation can be got

$$\boldsymbol{F}_a=-(\mathbf{J}_{4\times4}^{-1})^T\boldsymbol{Q}\tag{32a}$$

Where $\boldsymbol{Q}=[Q_1\ Q_2\ Q_3\ Q_4]^T$. $Q_i(i=1,2,3,4)$ are generalized forces and so α, β ,λ and Z_o ar e generalized coordinates.

$$\boldsymbol{Q}=\begin{bmatrix}F_x\frac{\partial X_o}{\partial\alpha}+F_y\frac{\partial Y_o}{\partial\alpha}+F_z\frac{\partial Z_o}{\partial\alpha}+T_x\frac{\partial\psi_x}{\partial\alpha}+T_y\frac{\partial\psi_y}{\partial\alpha}+T_z\frac{\partial\psi_z}{\partial\alpha}\\F_x\frac{\partial X_o}{\partial\beta}+F_y\frac{\partial Y_o}{\partial\beta}+F_x\frac{\partial Z_o}{\partial\beta}+T_x\frac{\partial\psi_x}{\partial\beta}+T_y\frac{\partial\psi_y}{\partial\beta}+T_z\frac{\partial\psi_z}{\partial\beta}\\F_x\frac{\partial X_o}{\partial\lambda}+F_y\frac{\partial Y_o}{\partial\lambda}+F_x\frac{\partial Z_o}{\partial\lambda}+T_x\frac{\partial\psi_x}{\partial\lambda}+T_y\frac{\partial\psi_y}{\partial\lambda}+T_z\frac{\partial\psi_z}{\partial\lambda}\\F_x\frac{\partial X_o}{\partial Z_o}+F_y\frac{\partial Y_o}{\partial Z_o}+F_z\frac{\partial Z_o}{\partial Z_o}+T_x\frac{\partial\psi_x}{\partial Z_o}+T_y\frac{\partial\psi_y}{\partial Z_o}+T_z\frac{\partial\psi_z}{\partial Z_o}\end{bmatrix}=\mathbf{J}_o^T\boldsymbol{F}\tag{32b}$$

Suppose that the rigid platform *m* is elastically suspended by four elastic active legs.

Let $\delta r_i(i=1,\ 2,\ 3,\ 4)$ be the axial deformation along $r_i(i=1,2,3,4)$ due to the active $F_{ai}(i=1,\ 2,3,\ 4)$, lead to.

$$F_{ai}=k_i\delta r_i(1,2,3,4)\tag{33a}$$

$$k_i=\frac{MS_i}{r_i}\tag{33b}$$

Where, *M* is the Young's modulus and S_i is the area of the *i*-th leg.

From Eq.(33a), we obtain

$$\boldsymbol{F}_a=\mathbf{K}_a\boldsymbol{\delta r}\ ,\ \mathrm{K}_a=diag[k_1,k_2,k_3,k_4]\tag{34}$$

From Eq.(23),lead to

$$\boldsymbol{\delta r}=\mathbf{J}_{4\times4}\ \boldsymbol{\delta q},\boldsymbol{\delta q}=\mathbf{J}_{4\times4}^{-1}\ \boldsymbol{\delta r}\tag{35}$$

Where $\delta\mathbf{r}=[\delta r_1\ \delta r_2\ \delta r_3\ \delta r_4]^{\mathrm{T}}$, $\delta\mathbf{q}=[\delta\alpha\ \delta\beta\ \delta\lambda\ \delta Z_o]^{\mathrm{T}}$.

The stiffness matrix of this parallel manipulator can be expressed as

$$\mathbf{K}_m = -\mathbf{J}_{4\times4}^T \mathbf{K}_a \mathbf{J}_{4\times4} \tag{36}$$

$\mathbf{K}_m$ is the 4 4 stiffness matrix which shows the relation between the generalized forces and the deformation of the generalized coordinate.

3.2.2 Second Method

The first method doesn't take into account the constrained forces. But the constrained forces exit in the lower-mobility PM is a universal phenomenon. They may produce dominating elastic deformations and affect the accuracy of parallel manipulator sometimes. So in this part, this factor will be considered and another method of stiffness model for this PM will be established.

Based on the principle of virtual work, the formula for solving the statics can be derived as follows:

$$\boldsymbol{F}_r^T \boldsymbol{V}_r + \boldsymbol{F}^T \boldsymbol{V} = 0,\ \boldsymbol{F}_r = -(\mathbf{J}_{6\times6}^{-1})^T \boldsymbol{F} = -(\mathbf{J}_{6\times6}^T)^{-1} \boldsymbol{F} \tag{37}$$

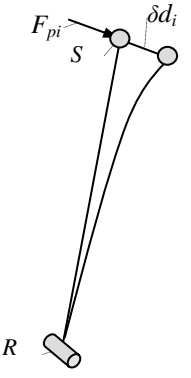


Fig. 2. The flexibility deformation of RPS leg

Where, $\boldsymbol{F}_r$=[F_{a1} F_{a2} F_{a3} F_{a4} F_{p1} F_{p2}]^T .The constrained forces in r_i(i=1,2) produce flexibility deformation. Let δd_i (i=1,2) be the flexibility deformation due to the constrained forces F_{pi} (i=1,2). The direction of this deformation can be considered along F_{pi}.(see Fig.2).

The relation between F_{pi} and δd_i can be expressed as

$$F_{pi} = s_i \delta d_i, (1,2) \tag{38}$$

$$s_i = \frac{3MI}{r_i^3} \tag{39}$$

Where, I is the area moment of leg's cross section.

From Eqs. (33) and (38), lead to

$$\boldsymbol{F}_r = \mathbf{K}_p \begin{bmatrix} \delta \boldsymbol{r} \\ \delta \boldsymbol{l} \end{bmatrix} \tag{40}$$

$$\mathbf{K}_p = diag[k_1, k_2, k_3, k_4, k_5, k_6] \tag{41}$$

Where, $\delta \boldsymbol{d}=[\delta d_1\ \delta d_2]^T$, with $\delta d_i (i=1,2,3)$denotes the bending deformation produced by constrained forces $F_{pi} (i=1,2)$.

Let $\delta \boldsymbol{x}=[\delta x\ \delta y\ \delta z]^T$, $\delta \boldsymbol{\theta}=[\delta\theta_x\ \ \delta\theta_y\ \delta\theta_z]^T$ be the linear and angle deformations of moving platform. From the principle of virtue work, lead to

$$\boldsymbol{F}_r^T \begin{bmatrix} \delta \boldsymbol{r} \\ \delta \boldsymbol{l} \end{bmatrix} + \boldsymbol{F}^T \begin{bmatrix} \delta \boldsymbol{x} \\ \delta \boldsymbol{\theta} \end{bmatrix} = 0 \tag{42}$$

From Eqs.(37), (40),(42),lead to

$$\begin{bmatrix} \boldsymbol{F} \\ \boldsymbol{T} \end{bmatrix} = -\mathrm{J}_{6\times6}^T \mathrm{K}_p \mathrm{J}_{6\times6} \begin{bmatrix} \delta \boldsymbol{r} \\ \delta \boldsymbol{l} \end{bmatrix} = -\mathrm{K}_{6\times6} \begin{bmatrix} \delta \boldsymbol{r} \\ \delta \boldsymbol{l} \end{bmatrix} \tag{43}$$

$\mathbf{K}_{6\times6} = -\mathbf{J}_{6\times6}^T \mathbf{K}_p \mathbf{J}_{6\times6}$ is the stiffness matrix both considering active forces and constrained forces.

4 Singularity Analysis

When the actuators are locked, the 2SPS+2RPS PM became a 2SS+2RS structure. Based on the observe method for finding constrained force/torque [16], some constrained forces can be found in this PM In the limb r_1, one constrained force $\boldsymbol{F}_{s1}$ which is along r_1 and one constrained $\boldsymbol{F}_{m1}$ which is parallel with R_1 and get across S joint can be found. In the limb r_2 and r_3, two constrained forces $\boldsymbol{F}_{s2}$ and $\boldsymbol{F}_{s3}$ along r_2 and r_3 respectively can be found. In the limb r_4, one constrained force $\boldsymbol{F}_{s4}$ which is along r_4 and one constrained $\boldsymbol{F}_{m2}$ which is parallel with R_2 and get across S joint can be found.

The center point of moving platform o can be seen as the terminal of each subchain, so the constrained forces do no work to the point o, lead to

$$\mathbf{G}_{6\times6} \boldsymbol{V} = \boldsymbol{0}_{6\times1} \tag{44a}$$

$$\mathbf{G}_{6\times6} = \begin{bmatrix} \boldsymbol{\delta}_1^T & (\boldsymbol{e}_1 \times \boldsymbol{\delta}_1)^T \\ \boldsymbol{\delta}_2^T & (\boldsymbol{e}_2 \times \boldsymbol{\delta}_2)^T \\ \boldsymbol{\delta}_3^T & (\boldsymbol{e}_3 \times \boldsymbol{\delta}_3)^T \\ \boldsymbol{\delta}_4^T & (\boldsymbol{e}_4 \times \boldsymbol{\delta}_4)^T \\ \boldsymbol{f}_1^T & (\boldsymbol{d}_1 \times \boldsymbol{f}_1)^T \\ \boldsymbol{f}_2^T & (\boldsymbol{d}_2 \times \boldsymbol{f}_2)^T \end{bmatrix} \tag{44b}$$

$\mathbf{G}_{6\times6}=\mathbf{J}_{6\times6}$ can be confirmed for this PM by comparing the two matrix. If Eq. (44a) has nontrivial solutions about $\boldsymbol{V}$, it means that this PM can move when actuators are locked. In this case, the singularity will appears, with the singularity condition is $|\mathbf{G}|=0$. The six row of matrix $\mathbf{G}$ represent six line vectors respectively, with four line vectors along r_i(i=1,2,3,4) and two line vectors which get across S joint and parallel with R joint in r_j(j=1,4).Then the singularity can be analyzed by Grassmann line geometry[8].But for this parallel manipulator, another method for singularity analysis proposed by Kong X.[11] is more aptitude. In another aspect, when the four actuators are locked, r_i can be seen as a rigid body. Let $\boldsymbol{v}_{ai}$(i=1,2,3) be the velocity vector of $\boldsymbol{a}_i$(i=1,2,3), v_{ai}(i=1,2,3) be the magnitude of $\boldsymbol{a}_i$(i=1,2,3). As a_1 and a_3 rotate with R_1 and R_2, a_2 rotate with A_2A_3. $\boldsymbol{v}_{ai}$(i=1,2,3) can be expressed as following

$$\boldsymbol{R}_1 v_{a1} = \boldsymbol{v}_{a1},\ \ \boldsymbol{R}_2 v_{a3} = \boldsymbol{v}_{a3} \tag{45a}$$

$$(\boldsymbol{\delta}_2 \times \boldsymbol{\delta}_3) v_{a2} / \left|(\boldsymbol{\delta}_2 \times \boldsymbol{\delta}_3)\right| = \boldsymbol{v}_{a2} \tag{45b}$$

As the moving platform is an equilateral triangle $a_1a_2a_3$, lead to

$$\left|\boldsymbol{a}_2 - \boldsymbol{a}_1\right| = \left|\boldsymbol{a}_3 - \boldsymbol{a}_1\right| = \left|\boldsymbol{a}_2 - \boldsymbol{a}_3\right| = qe \tag{46}$$

Then, we obtain

$$(\boldsymbol{a}_2 - \boldsymbol{a}_1)^T (\boldsymbol{a}_2 - \boldsymbol{a}_1) = 3e^2 \tag{47a}$$

$$(\boldsymbol{a}_3 - \boldsymbol{a}_1)^T (\boldsymbol{a}_3 - \boldsymbol{a}_1) = 3e^2 \tag{47b}$$

$$(\boldsymbol{a}_2 - \boldsymbol{a}_3)^T (\boldsymbol{a}_2 - \boldsymbol{a}_3) = 3e^2 \tag{47c}$$

Taking the derivative of both sides of Eq.(47a) to (47c) by time, lead to

$$[\dot{\boldsymbol{a}}_2 - \dot{\boldsymbol{a}}_1]^T (\boldsymbol{a}_2 - \boldsymbol{a}_1) = 0 \tag{48a}$$

$$[\dot{\boldsymbol{a}}_3 - \dot{\boldsymbol{a}}_1]^T (\boldsymbol{a}_3 - \boldsymbol{a}_1) = 0 \tag{48b}$$

$$[\dot{\boldsymbol{a}}_2 - \dot{\boldsymbol{a}}_3]^T (\boldsymbol{a}_2 - \boldsymbol{a}_3) = 0 \tag{48c}$$

From Eqs.(45a) to (45c) and Eqs. (48a) to (48c), lead to

$$[(\boldsymbol{\delta}_2 \times \boldsymbol{\delta}_3) v_{a2} / \left|(\boldsymbol{\delta}_2 \times \boldsymbol{\delta}_3)\right| - \boldsymbol{R}_1 v_{a1}]^T (\boldsymbol{a}_2 - \boldsymbol{a}_1) = 0 \tag{49a}$$

$$(\boldsymbol{R}_2 v_{a3} - \boldsymbol{R}_1 v_{a1})^T (\boldsymbol{a}_3 - \boldsymbol{a}_1) = 0 \tag{49b}$$

$$[(\boldsymbol{\delta}_2 \times \boldsymbol{\delta}_3) / \left|(\boldsymbol{\delta}_2 \times \boldsymbol{\delta}_3)\right| - \boldsymbol{R}_2 v_{a3}]^T (\boldsymbol{a}_2 - \boldsymbol{a}_3) = 0 \tag{49c}$$

Eqs.(49a) and Eq.(49c) can be written in matrix form as following

$$\begin{bmatrix} -\boldsymbol{R}_1 \cdot (\boldsymbol{a}_2 - \boldsymbol{a}_1) & (\boldsymbol{\delta}_2 \times \boldsymbol{\delta}_3) \cdot (\boldsymbol{a}_2 - \boldsymbol{a}_1) / \left|(\boldsymbol{\delta}_2 \times \boldsymbol{\delta}_3)\right| & 0 \\ -\boldsymbol{R}_1 \cdot (\boldsymbol{a}_3 - \boldsymbol{a}_1) & 0 & \boldsymbol{R}_2 \cdot (\boldsymbol{a}_3 - \boldsymbol{a}_1) \\ 0 & (\boldsymbol{\delta}_2 \times \boldsymbol{\delta}_3)(\boldsymbol{a}_2 - \boldsymbol{a}_3) / \left|(\boldsymbol{\delta}_2 \times \boldsymbol{\delta}_3)\right| & -\boldsymbol{R}_2 \cdot (\boldsymbol{a}_2 - \boldsymbol{a}_3) \end{bmatrix} \begin{bmatrix} v_{a1} \\ v_{a2} \\ v_{a3} \end{bmatrix} = 0 \tag{50}$$

When the actuators are locked, v_{ai} (i=1, 2, 3) must equals 0. If Eq.(50) have nontrivial solutions about v_{ai}, the singularity configuration will appears and the following condition must satisfied.

$$\begin{vmatrix} -\boldsymbol{R}_1\cdot(\boldsymbol{a}_2-\boldsymbol{a}_1) & (\boldsymbol{\delta}_2\times\boldsymbol{\delta}_3)\cdot(\boldsymbol{a}_2-\boldsymbol{a}_1)/|(\boldsymbol{\delta}_2\times\boldsymbol{\delta}_3)| & 0 \\ -\boldsymbol{R}_1\cdot(\boldsymbol{a}_3-\boldsymbol{a}_1) & 0 & \boldsymbol{R}_2\cdot(\boldsymbol{a}_3-\boldsymbol{a}_1) \\ 0 & (\boldsymbol{\delta}_2\times\boldsymbol{\delta}_3)(\boldsymbol{a}_2-\boldsymbol{a}_3)/|(\boldsymbol{\delta}_2\times\boldsymbol{\delta}_3)| & -\boldsymbol{R}_2\cdot(\boldsymbol{a}_2-\boldsymbol{a}_3) \end{vmatrix}=0 \tag{51}$$

From Eq.(51),the singularity condition can be derived as following

$$\begin{aligned} &[\boldsymbol{R}_1\cdot(\boldsymbol{a}_2-\boldsymbol{a}_1)][(\boldsymbol{R}_2\cdot(\boldsymbol{a}_3-\boldsymbol{a}_1)][(\boldsymbol{\delta}_2\times\boldsymbol{\delta}_3)\cdot(\boldsymbol{a}_3-\boldsymbol{a}_2)] \\ &-[(\boldsymbol{\delta}_2\times\boldsymbol{\delta}_3)\cdot(\boldsymbol{a}_2-\boldsymbol{a}_1)][\boldsymbol{R}_1\cdot(\boldsymbol{a}_3-\boldsymbol{a}_1)][\boldsymbol{R}_2\cdot(\boldsymbol{a}_2-\boldsymbol{a}_3)]=0 \end{aligned} \tag{52}$$

5 Analytic Solved Example

Set θ_1=45 and θ_2=135 . From Eqs. (9) and (10),lead to,

$$X_o = e(y_l - qx_m)/2 = e(-c_\alpha s_\beta c_\lambda + s_\alpha s_\lambda - qs_\beta)/2$$
$$Y_o = e(y_m - qx_l)/2 = ec_\beta(c_\lambda - qc_\alpha)/2$$

Set E=0.5m, e=0.6/qm, F_x=-20,F_y=-30, F_z=-30,T_x=T_y=-30, T_z=0KN.m. The position and orientation of the moving platform are given as Z_o=1.2m, α ,β=8 ,γ=15 .

The length of r_i(i=1,2,3,4) is solved as:

$$\boldsymbol{r}=(1.1553 \quad 1.4535 \quad 1.3923 \quad 1.2260)^T m$$

The statics of this parallel manipulator can be solved as

$$\boldsymbol{F_r}=[-59.4587\ 15.5287\ 81.0946\ 27.7766\ \ 24.8142\ \ 45.1086]N$$

Some physical parameters are given as:M=2.11×10^{10}N/m^2,S_1=S_2=S_3=S_4 =0.0013m^2, I=1.2560×10^{-7}m^4

The deformation produced by active forces of this parallel manipulator is

$$\delta\boldsymbol{r}=10^{-7}(-2.5920\ \ 0.85170\ \ 4.2603\ \ 1.2850)^T m$$

The deformation produced by constrained forces is

$$\delta\boldsymbol{d}=(0.0005\ \ 0.0010)^T \text{cm}$$

The deformation of generalized condinate of moving platform solved by first method is

$$\delta\boldsymbol{q}=10^{-6}(0.5354\text{rad} \quad -0.3976\text{rad} \quad 0.7503\ \text{rad} \quad 0.0552\text{m})^T$$

The deformation of the moving platform of this parallel manipulator is

$$\delta\boldsymbol{x}=(0.00094071 \quad 0.0010452 \quad -0.00006796)^T m$$
$$\delta\boldsymbol{\theta}=(0.000815 \quad -0.000031069 \quad 0.0028286)^T rad$$

The 4×4 stiffness matrix of this parallel manipulator is

$$\mathbf{K}_{4\times4}=10^8\begin{bmatrix} 0.3554 & -0.1911 & 0.0180 & 0.0542 \\ -0.1911 & 0.3733 & 0.0600 & 0.3542 \\ 0.0180 & 0.0600 & 0.5036 & 0.4139 \\ 0.0542 & 0.3542 & 0.4139 & 7.2639 \end{bmatrix}$$

The 6×6 stiffness matrix of this parallel manipulator is

$$\mathbf{K}_{6\times6} = 10^7 \begin{bmatrix} 6.12 & -0.37 & -3.16 & -0.58 & 2.55 & -1.77 \\ -0.37 & 3.08 & 0.51 & -3.57 & -0.73 & 0.02 \\ -3.16 & 0.51 & 72.64 & 4.52 & 0.67 & 1.36 \\ -0.58 & -3.57 & 4.52 & 4.99 & -0.32 & 0.18 \\ 2.55 & -0.73 & 0.67 & -0.32 & 3.36 & -0.43 \\ -1.78 & 0.02 & 1.32 & 0.18 & -0.43 & 0.56 \end{bmatrix}$$

6 Conclusions

The first stiffness model established for this 2SPS+2RPS PM with the constrained forces/torques unconsidered is fit for common lower-mobility, with the stiffness matrix expressed in a unified form. By using the second method, the active and constrained forces are solved, and the deformations produced by them are derived. Based on the solving result, we can see that the deformation produced by constrained forces hold dominating station for this parallel manipulator, so the second method with the constrained forces considered is more aptitude for this 2SPS+2RPS PM. The singularity of this PM is analyzed by locking the four actuators. The first approach is fit for common lower-mobility PMs for singularity analysis, but analyzing a 6×6 matrix is a complex thing. By using the second method, the singularity of this PM can be expressed in a explicit form. This PM has some potential applications for the 4-DOF parallel machine tool, micro surgical manipulator, pick-up manipulator and so on for its simple structure.

Acknowledgments

This research work is supported by the Natural Sciences Foundation Council of China (NSFC) 50575198, the Doctoral Fund from National Education Ministry No. 20060216006, and the Science Foundation of Yanshan University for the Excellent Ph.D.Students No. YSU200901.

References

1. Merlet, J.-P.: Parallel robots. Kluwer Academic Publishers, London (2000)
2. Li, Q., Huang, Z.: Type synthesis of 4-DOF parallel manipulators. In: IEEE ICRA 2003, pp. 755–760 (2003)
3. Kong, X., Gosselin, C.M.: Type synthesis of 4-DOF SP-equivalent parallel manipulators: A virtual chain approach. Mechanism and Machine Theory 41(11), 1306–1319 (2006)
4. Kong, X., Gosselin, C.M.: Type synthesis of 3T1R 4-DOF parallel manipulators based on screw theory. IEEE Transactions on Robotics and Automation 20(2), 181–190 (2004)

5. Richard, P.-L., Gosselin, C.M., Kong, X.: Kinematic analysis and prototyping of a partially decoupled 4-DOF 3T1R parallel manipulator. ASME Journal of Mechanical Design 129(6), 611–616 (2007)
6. Zoppi, M., Zlatanov, D., Gosselin, C.M.: Analytical kinematics models and special geometries of a class of 4-dof parallel mechanisms. IEEE Transactions on Robotics 21(6), 1046–1055 (2005)
7. Lu, Y., Hu, B.: Analyzing kinematics and solving active/constrained forces of a 3SPU+UPR parallel manipulator. Machine and Mechanism Theory 42(10), 1298–1313 (2007)
8. Merlet, J.P.: Singular configurations of parallel manipulators and grassman geometry. International Journal of Robotics Research 8, 45–56 (1989)
9. Gallardo-Alvarado, J., Rico-Martinez, J.M., et al.: Kinematics and singularity analyses of a 4-dof parallel manipulator using screw theory. Mechanism and Machine Theory 41(9), 1048–1061 (2006)
10. Liu, C.H., Hsu, F.-K.: Direct singular positions of the parallel manipulator Tricept. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science 221(1), 109–117 (2007)
11. Xianwen, K., Gosselin, C.M.: Uncertainty singularity analysis of parallel manipulators based on the instability analysis of structures. International Journal of Robotics Research 20(11), 847–856 (2001)
12. Zhao, J.-S., Feng, Z.-J., Zhou, K., Dong, J.-X.: Analysis of the singularity of spatial parallel manipulator with terminal constraints. Mechanism and Machine Theory 40(3), 275–284 (2005)
13. Zhang, D., Lang Sherman, Y.T.: Stiffness modeling for a class of reconfigurable PKMs with three to five degrees of freedom. Journal of Manufacturing Systems 23(4), 316–327 (2004)
14. Zhang, D., Gosselin, C.M.: Kinetostatic modeling of parallel mechanisms with a passive constraining leg and revolute actuators. Mechanism and Machine Theory 37(6), 599–617 (2002)
15. Han, S., Yuefa, F., Huai, C.: Stiffness analysis of four degrees parallel manipulator. Chinese Journal of Mechanical Engineering 42, 31–34 (2006)
16. Lu, Y., Hu, B.: Unification and simplification of velocity/acceleration of limited-dof parallel manipulators with linear active legs. Mechanism and Machine Theory 43(9), 1112–1128 (2008)

A Simulation Model to Evaluate and Verify Functions of Autonomous Vehicle Based on Simulink®

Jie Zhou*, Hui Chen, and Caijing Xiu

Tongji University, School of Automotive Studies, Cao An road. 4800, 201804 Shanghai, China
zhoujie19831011@163.com,
hui-chen@mail.tongji.edu.cn,
xcj2_11@163.com

Abstract. The paper presents a simulation model of autonomous vehicle based on Matlab/Simulink®, which is simple and efficient. Concerning about the relationship among autonomous vehicle, road, and obstacle, it forms a 'Vehicle—Sensor—Controller' closed-loop control system. The model provides a platform to design and validate the control logic of Local Path Planning, and to design post-process of the raw data from sensors. The paper also presents the simulation results of a PID steering controller for lane following, Artificial Potential Field (APF) method for collision avoidance, and the dynamic planning for overtaking when the obstacle is moving. The simulation model is verified to be effective by a lane change experiment.

Keywords: Autonomous Vehicle, Local Path Planning, Simulation, raw data post-process.

1 Introduction

The development of autonomous vehicles becomes a hot topic due to the demand of the intelligent transportation systems (ITS) and the military applications [1]. Local path planning plays an important role in the navigation of the autonomous vehicle, because it collects the information from sensors, and decides next move of the vehicle. In most conditions, autonomous vehicle should drive follow the road line and overtake an obstacle. Different methods of local path planning have been developed for meeting the requirements of lane following and collision avoiding. Another point is when the camera and radar detect the environment around the vehicle; the raw data from these sensors should be post process and be translated to different information which can be understood by the local path planner.

Many research institutes have developed their own simulator or simulation model for the function development, such as a C++ based platform for testing Path Planning algorithms [2], a Visual C++ 6.0 based AGV path planning system [3], and the Zhao's

* Supported by Tongji University, Shanghai 200092, China, TRW Automotive R&D (Shanghai) Co., Ltd. Science and Technology Commission of Shanghai Municipality(07dz05901) and SLG Chassis Electronic Fund Chairs.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 645–656, 2009.

simulation system based on VC++6.0 [4]. These simulation platforms are vividly and intuitively but the developer have to cost a lot of time to design and create the simulation environment and vehicle's performance. It is not an easy job for those who are not good at C code programming. Therefore, a simple and efficient model would be expected.

Some papers have introduced some simulation platform for parts of the autonomous vehicle, such as Simulation of control system [5], Simulation of Local Path Planning [3]. Some simulation models based on a completed and precise map generate the ideal path, and the simulation will run after the map established, such as the simulation introduced by Shi Jian [6], the road data should be pre-loaded to the model, which can not evaluate the path planning control logic as a closed-loop control system.

For these reasons, a model based on Matlab/Simulink® is presented and most of the aspects of autonomous vehicle functions can be simulated as a closed-loop control system, including sensor (as camera and radar in this paper) raw data post process, local path planning, trajectory following, car body control, and so on. The model can be used to design and validate the control logic of local path planning, and to simulate the post process of raw data from sensors according to the various path planning methods. Since the environment information is detected and fed back, the model can simulate the control algorithm when the obstacle is moving.

The presented simulation model consists of three modules, as Fig. 1 shows, in which 2 DOF (Degree Of Freedom) vehicle dynamics module receives the steering angle and outputs the current coordinate XY (global coordinate) of the vehicle. The perception module receives the vehicle's coordinate, compares it with the environment information which is defined in configuration file, then outputs the required information for path planning, and the post process of raw data from environment perception sensors is included in this module. The rest module is local path planner which outputs the steering angle to control the vehicle according to the environment information, and the algorithm of local path planning is in this module.

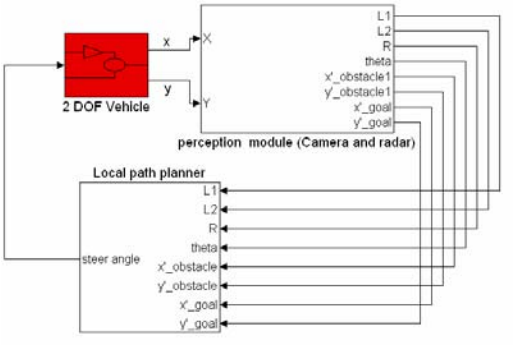


Fig. 1. Simulation model

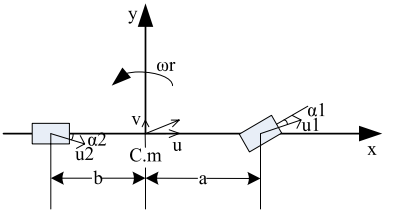


Fig. 2. 2 DOF vehicle dynamics module

The model configuration will be described in section 2, the three modules will be detail described in section 3, 4, 5, and examples of lane following, collision avoidance and overtaking will be given in section 6.

2 Model Configuration

The configuration file is written in a M file and is pre-loade before the simulation running, which contains the definition of the necessary parameters.

A. Since the 2DOF vehicle model is used, the vehicle parameters as follow should be defined. [7]

- The speed of the vehicle u, which is supposed to be a constant value.
- The mass of the vehicle m.
- The moment of Inertia around Z axis Iz.
- The front and rear tire cornering stiffness k1, k2.
- The length between center of gravity and front axle a, center of gravity and rear axle b.

B. The position of the road. The road is a continuous curve which is fit by several points.

C. The position of the obstacle. The obstacle's position in this model is defined as one point, which means the nearest point of the obstacle from the vehicle. The obstacle's position is measured from the radar, and the point will be calculated and chosen from the radar's output in real situation. The obstacle's position can be defined as other ways, like the shape of the obstacle or the occupied area of it, which depends on the navigation algorithm used in Local path planner.

Both of the position of road and obstacle is defined in the global XY coordinate, and they will be transformed to vehicle's local coordinate in the perception module. After loading the parameters, the World Map containing road and obstacle can be drawn like Fig. .

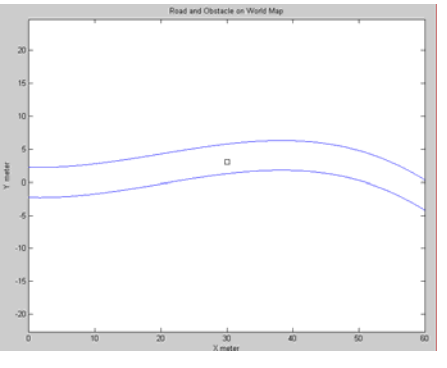


Fig. 3. Road and Obstacle on World Map

3 2DOF Vehicle Dynamic Module

The 2DOF vehicle dynamic module, which is also named as bicycle model, is responsible for calculating the dynamic coordinate of the vehicle. The input of the module is steering angle of front wheel, and output is XY coordinate of the vehicle. [7]

$$\begin{cases} (k_1+k_2)\beta+\dfrac{1}{u}(ak_1-bk_2)\omega_\gamma-k_1\delta=m\left(\upsilon'+u\omega_\gamma\right) \\ (ak_1-bk_2)\beta+\dfrac{1}{u}\left(a^2k_1+b^2k_2\right)\omega_\gamma-ak_1\delta=Iz\omega_\gamma' \end{cases} \tag{1}$$

In (1), ν means lateral velocity, ωr means vehicle yaw rate, δ means steering angle of front wheel, β means vehicle side slip angle, and $\beta=\dfrac{\upsilon}{u}$.

$$\begin{cases} X=\left(\int_0^t\cos\left(\beta+\int_0^t\omega_\gamma dt\right)dt\right)*\sqrt{u^2+v^2} \\ Y=\left(\int_0^t\sin\left(\beta+\int_0^t\omega_\gamma dt\right)dt\right)*\sqrt{u^2+v^2} \end{cases} \tag{2}$$

The dynamic XY coordinate of the vehicle is calculated by (2), which can be compared with the position of road.

4 Perception Module

The Perception module is responsible for the post process of raw data from camera and radar, and expressing the relationship among autonomous vehicle, road, and obstacle as well.

4.1 Camera Data

In most situations, camera is used to detect the road line for an autonomous vehicle, and it will lead the vehicle to follow a lane. We choose the Lane Detector product of Vislab, University of Parma, IT [8], which will send out several coordinate of the points on the road line as raw data. See Fig. 4. [8]

In order to control the vehicle to follow the lane, and moved in a smooth, steady trajectory, the local path planner need the detail information of the road line. See Fig. 5.

1: Express the two road lines with two 3-order polynomials.
2: L1 means the distance between the center line of the vehicle and the left road line. L2 means the distance between the center line of the vehicle and the right road line.
3: The angle θ means the angle between the center line of the vehicle and the center line of the road.
4: The radius R of the road.

Such information can be calculated by following post process methods and be expressed in the xy coordinate, which is vehicle's local coordinate:

1: Use Matlab function 'polyfit' to fitting the points.
X=[0 5 10 15]; Y=[0 0.5 -1 0.5]; coff=polyfit(X,Y,3);
2: L1 and L2 are calculated by the Y value of road line and vehicle.
3: Angle θ is calculated by the gradient of road line and vehicle.

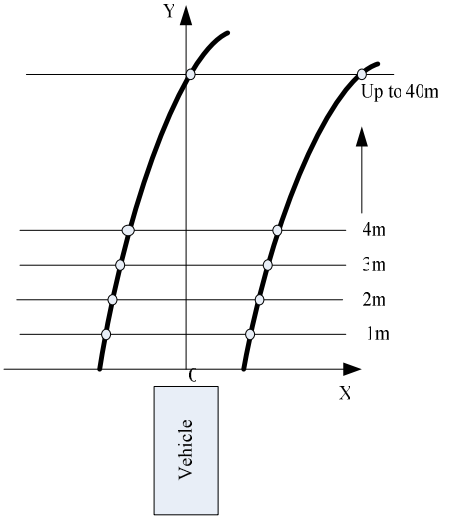


Fig. 4. Points of road line (raw data)

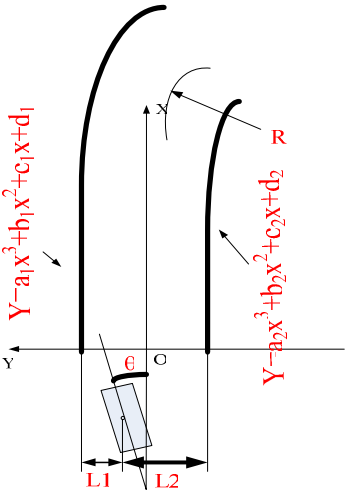


Fig. 5. Required data of path planner

$$\theta = \arctan\left(\frac{y_{t1}^{vehicle} - y_{t0}^{vehicle}}{x_{t1}^{vehicle} - x_{t0}^{vehicle}}\right) - \arctan k_{x_{t1}}^{road} \tag{3}$$

$k_{x_{t1}}^{road}$ means the gradient of the road line on the vehicle's current position.

4: R is calculated by the road lines 3-order polynomials. [9]

$$R = \frac{\left(1 + (y')^2\right)^{\frac{3}{2}}}{|y''|} \tag{4}$$

y', y'' mean the 1st and 2nd derivative of road line.

Since all these equations are based on dynamic vehicle's xy coordinate, the four values (L1, L2, θ, R) can be directly used by the Local path planner. See Fig. 6

4.2 Radar Data

In most situations, radar is used to detect the obstacle for an autonomous vehicle, and it will lead the vehicle to avoid a collision. The LMS291 Laser Measurement Systems is chose as our sensor [10], which will send out the position of the obstacle in polar coordinates and they are transformed to be a XY coordinate which is more convenient for us to do further process. See Fig. 7. [10]

The max scan field can be configured as 180 degree, and the resolution as 0.5 degree, so the max number of measured values is 361.

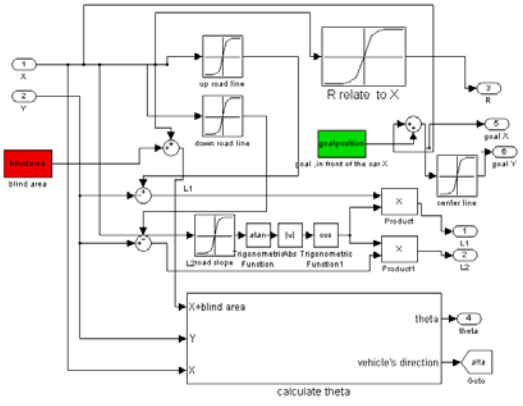


Fig. 6. Camera raw data post process

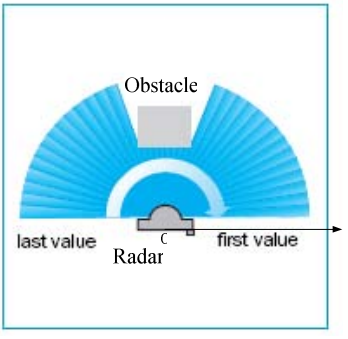


Fig. 7. Radar scan map

When the radar is mounted on the vehicle, the output data is the obstacle's position in the vehicle's local coordinate. But when the perception module is used to simulate the output of radar, the data should be transformed into vehicle's local coordinate just like they were measured by the radar mounted on the vehicle, because the obstacle's position is defined with a global coordinate in the configuration file.

The coordinate transformation matrix is as follow

$$\begin{cases} x' = (x - X)\cos(\text{alfa}) + (y - Y)\sin(\text{alfa}) \\ y' = -(x - X)\sin(\text{alfa}) + (y - Y)\cos(\text{alfa}) \end{cases} \tag{5}$$

In (5), x and y mean the global coordinate of the obstacle, X and Y mean the dynamic global coordinate of the vehicle, and alfa means the direction of the vehicle.

$$\text{alfa} = \arctan\left(\frac{y_{t1}^{vehicle} - y_{t0}^{vehicle}}{x_{t1}^{vehicle} - x_{t0}^{vehicle}}\right) \tag{6}$$

Here x' and y' mean the obstacle's position expressed in vehicle's local coordinate, which can be directly used by the Local path planner.

5 Local Path Planner Module

The Local path planner module is responsible for calculating the steering angle of the front wheel according to the information provided by Perception module. Variety of local path planning methods can be developed in this module. Parameters of the local path planning algorithms can be tuned, designed and validated by running the whole model. According to different methods, such as Vector Pursuit Path Tracking [11], Optimal Feedback [6], Preview Follower Theory [12], the Perception module should be changed slightly to meet the requirement.

In this paper, a PID steering controller is used to keep L1 equals L2, so that the vehicle will drive in the middle of the lane. See Fig. .The angle θ and the R will be also considered in future work.

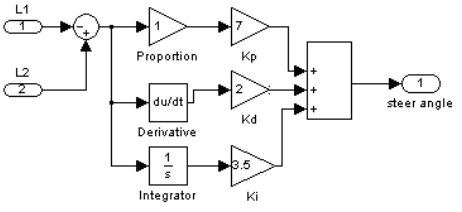


Fig. 8. PID controller for lane keeping

In this paper, the Artificial Potential Field method is introduced for collision avoidance. In the traditional artificial potential field methods, an obstacle is considered as a point of highest potential, and a goal as a point of lowest potential [13]. The attractive force towards the goal (F_a) and the repulsive force (F_r) from an obstacle are defined respectively with a function of the distance of them. And the vehicle should drive in the direction of the resultant force (F_{art}), so that it can avoid a collision while go toward the goal. See Fig. 9 and 10.

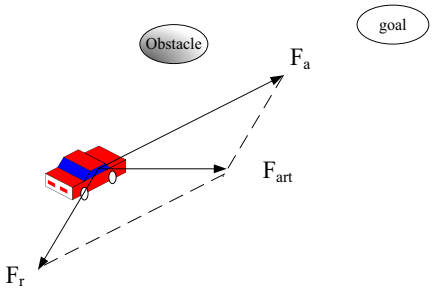


Fig. 9. Artificial Potential Field Method

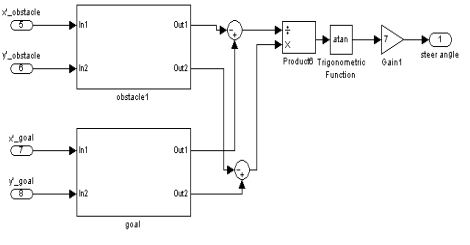


Fig. 10. APF for collision avoidance

In this paper, the aim point's angle error is used to control the vehicle to perform overtaking. Suppose the speed of the obstacle is lower than that of the vehicle, when the vehicle is D1 behind the moving obstacle, the road line of the adjacent lane which is provided by camera, will guide the vehicle to move into that lane, and the original lane's information will guide the vehicle drive back when it is far enough (D2) in front of the moving obstacle. See Fig. 11 and 12.

As Fig. 13 shows, the steering angle is proportional to the angle θ, which keeps the vehicle driving towards the aim point.

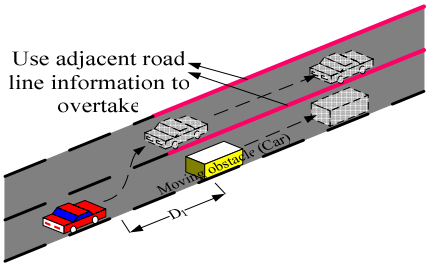


Fig. 11. Overtaking

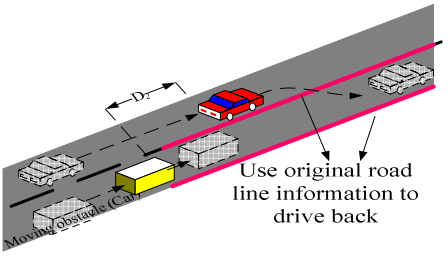


Fig. 12. Drive back

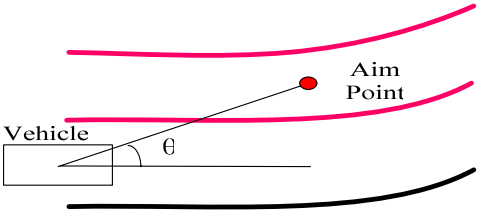


Fig. 13. Steering control logic for overtaking

6 Simulation Results

As Fig. 14, 15 shows, the blue line means the road line, and the square in Fig. 15 is an obstacle, the red line is the trajectory of the vehicle. The performance of Lane Following and Collision Avoidance are acceptable. The parameter of the vehicle is as follow: m=1359.8kg, Iz=1992.54 kg*m2, a=1.06282m, b=1.48518m, k1= 52480 N/rad, k2=88416N/rad, and vehicle speed is u=5km/h.

Notice that the center point of the road which is 10m ahead of the dynamic vehicle's position is chose as goal on each simulation step.

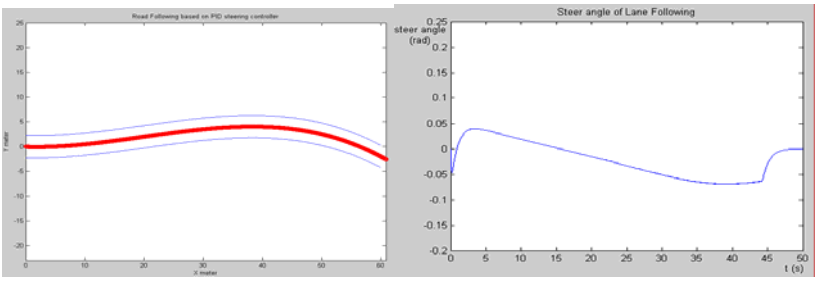


Fig. 14. Lane following and the steering angle output

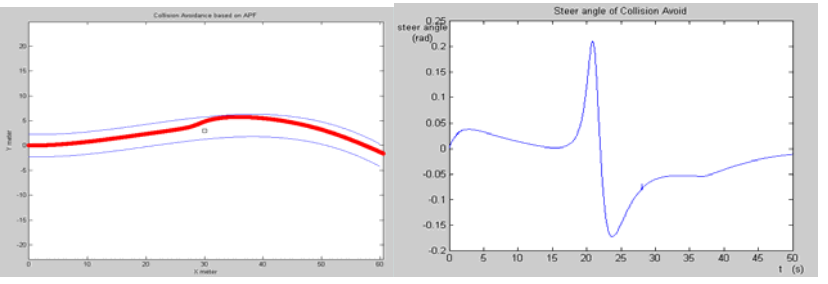
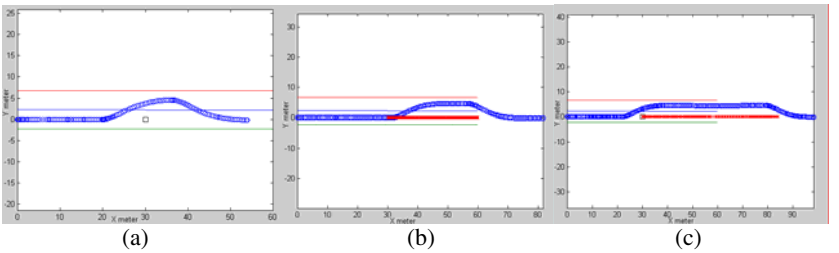


Fig. 15. Collision Avoid and the steering angle output

When performing a simulation of overtaking, as Fig. 1 shows, the vehicle can dynamically plan the suitable path to go. The speed of the vehicle is 5.5m/s, shown in blue and the obstacle is moving with 2m/s (b), 3m/s (c), shown in red. In (a), the obstacle is not moving.



(a) (b) (c)

Fig. 16. (a) Overtaking, ( ) Overtaking, (c) Overtaking

The distance of the vehicle runs in the adjacent lane is obviously different according to the difference of the speed of the obstacle. The simulation of dynamic planning is based on the closed-loop of the control system and the dynamic detection of the virtual sensors.

7 Experiment Result

In order to verify the usage of the simulation model, a lane change simulation and the vehicle experiment is conducted. A sine function steering angle like A*sin(wt) is sent to the chassis system when the obstacle is within a certain distance in front of the vehicle. The parameters A and w are decided by road width (ydes) and the parameters of vehicle, such as a, b, m, k1, k2 described in Chapter 2.

The relationship of them is described as follow equation (7)

$$\frac{ydes[(a+b)^2 + mu^2(\frac{a}{k_r} - \frac{b}{k_f})]}{u^2(a+b)} = \frac{2\pi A}{\omega^2} \tag{7}$$

In which, w is firstly determined according to the vehicle speed and the distance from the obstacle, and then A is calculated by equation (7) assuming that u is a constant value during the lane change period. In this paper, the ydes is set to 4m and 5m, u is set to 2m/s, and while the trigger condition is that the obstacle is within 15m in front of the vehicle. The w is chose as 0.6283rad/s, and A is 0.1508 and 0.1885. The parameters and trigger condition are set same both in simulation and experiment, and the control logic is completed by Simlink/Stateflow.

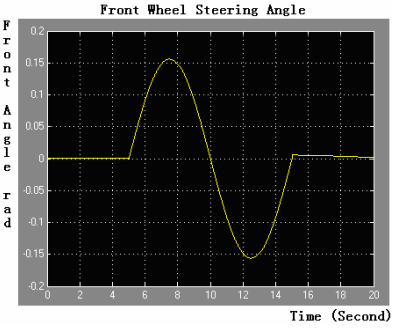


Fig. 17. Steering Angle Command

Fig. 18. Experimental Vehicle

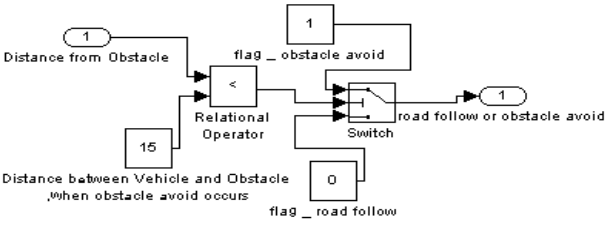


Fig. 19. Trigger control

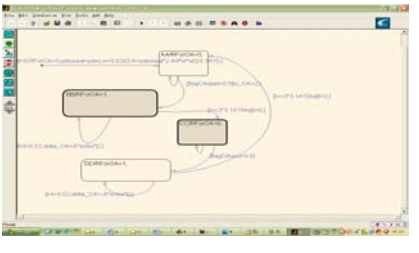


Fig. 20. Control logic

We use GPS/INS RT3050 to record the trajectory of the vehicle in red line and compare it with the simulation result in blue line, as Figure 21, Figure 22 show, the obstacle is at position (35, 0)/(20,0), the vehicle is triggered at (20, 0)/(5,0), and send the sine signal to the steering system. The two lines match well enough to verify the effectiveness of the simulation model.

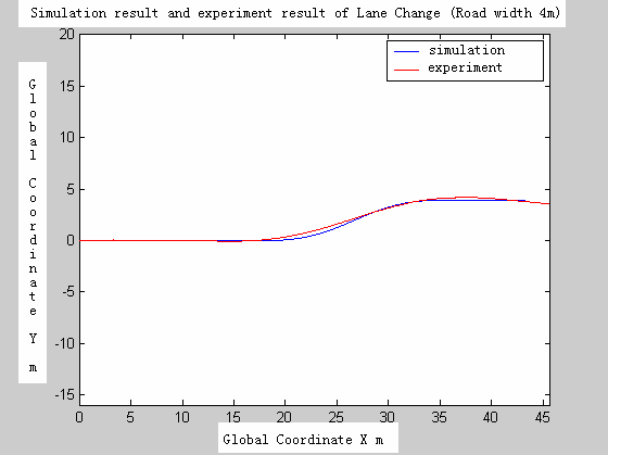


Fig. 21. 4m Lane Change Simulation and Experiment

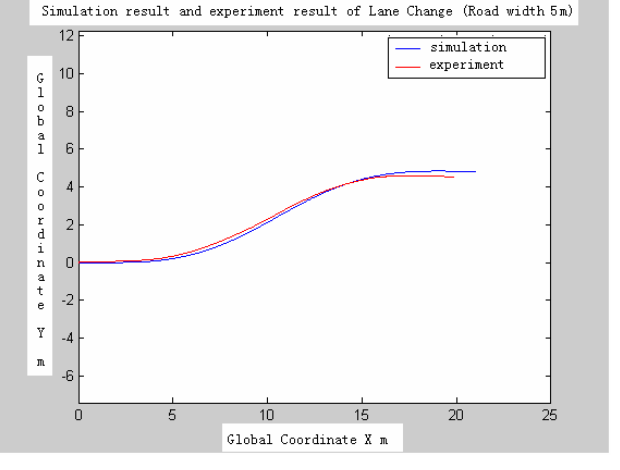


Fig. 22. 5m Lane Change Simulation and Experiment

8 Conclusion

This paper presents a simulation model which could provides a platform to design and validate the control logic of Local Path Planning, and to design post-process of the raw

data from sensors, which has been proved to be effective through a lane change experiment. Other applications, like Automatic Parking [14], Adaptive Cruise Control (ACC) [14], the method of local path planning and sensor data post-process can also be developed by this model. The environment can be edited and detected dynamically. This model will be complemented more details, including the blind area of the camera, the delays of the actuators or sensors, to approach a much more real situation. Also, some visual software like Virtual Reality Toolbox ™ [15] will be used to make the simulation results more intuitive in future.

References

1. Xie, M., Chen, H., Zhang, X.F., Guo, X., Yu, Z.P.: Development of Navigation System for Autonomous Vehicle to Meet the DARPA Urban Grand Challenge. In: Proceedings of the 2007 IEEE Intelligent Transportation Systems Conference, Seattle, WA, USA, September 30- October 3 (2007)
2. Rubin, R.: DARPA Urban Challenge, a C++ based platform for testing Path Planning Algorithms: An application of Game Theory and Neural Networks, http://mpra.ub.uni-muenchen.de/4603/
3. Minghua, Z.: Research and Simulation of Path Plan Algorithm for AGV of Automated Storage and Retrieva1 System, Graduate thesis, Jiangsu University (2006)
4. Chunxia, Z., Zhenmin, T., Jianfeng, L., Jingyu, Y., Chuancai, L.: The Simulation System for the Local Path Planning of the Autonomous Vehicle. Journal of Nanjing University of Science and Technology 26(6) (December 2002)
5. WenFeng, Y.: Research on Intelligent Vehicle's Hierarchical Architecture and Route Following Algorithm, Graduate thesis, Nanjin Tech. and Industry College (2001)
6. Jian, S., Bin, Z.: Simulation of Automatic Driving Vehicle. Automotive Engineering 22(2) (2000)
7. Zhisheng, Y.: Automobile Theory. China Machine Press (2004) ISBN 7-111-02076-6
8. VisLab Embedded Lane Detector Technical Information Version 1.1.0, Vislab.it, Parma, Italy, http://vislab.it/
9. Xuefei, Z., Zhuoping, Y., Hui, C., Ming, X.: A Method of Local Path Planning and Trajectory Generation for ALV. In: 2007 SAE-China, Tianjin, China (November 2007)
10. LMS200/211/221/291Laser Measurement Systems Technical description, http://www.sick.com
11. Wit, J.S.: Vector pursuit path tracking for autonomous ground vehicle, Ph.D. Dissertation, Department of Mechanical Engineering, University of Florida (2000)
12. Zhenhai, G., Xin, G., Konghui, G.: Application of preview follower theory and driver model in the research of vehicle intelligent handling. Journal of Traffic and Transportation Engineering 2(2), 64–66 (2002)
13. Vadakkepat, P., Tan, K.C., Wang, M.L.: Evolutionary artificial potential fields and their application in real time robot path planning. In: Congress on Evolutionary Computation, Proceedings of the 2000, vol. 1, pp. 256–263 (2000)
14. Bishop, R.: Intelligent Vehicle Technology and Trends. Artech House Publishers, London (2005)
15. http://www.mathworks.cn/products/virtualreality/

Experimental Evaluation on Dynamic Characteristics of Coaxial Magnetic Couplings for Vacuum Robot

Jun Wu, Pinkuan Liu, and Yulin Wang

State Key Lab. of Mechanical System and Vibration,
School of Mechanical Engineering, Shanghai Jiao Tong University,
800 Dong Chuan Road, Shanghai 200240, P.R. China
wujun_sjtu@sjtu.org, {pkliu,wyl_sjtu}@sjtu.edu.cn

Abstract. Permanent magnetic couplings have the ability of transmitting torque from a primary driver to a follower without mechanical contact. In this paper, a pair of coaxial couplings are applied to vacuum robot design and fabrication. The tracking performance at different velocities without misalignment is presented first. Then the effects of radial offset and axial shift on the tracking errors are investigated. And, the cross coupling effects of two sets of couplings are simulated by 3D FEA, based on which a new method to minimize the cross coupling is proposed. Finally, experiments are carried out and results prove that magnetic couplings are feasible and reliable for further application.

Keywords: Magnetic couplings, tracking error, misalignment effects, cross coupling effects.

1 Introduction

With the semiconductor industry advancing production to the 65nm node and developments at the 32nm node, it is increasingly critical that equipments and process reproducibility be controlled to the tolerances which are never required in semiconductor industry previously. Many processing techniques such as etch, deposition, ion implantation and so on require ultrahigh cleanness and vacuum or special gases which are hazardous for human. Vacuum-compatible transfer robot is used to handle wafer among different processing modules in the completely hermetic volume. With a vacuum environment, airborne molecular contamination problems are solved, but new challenges regarding the survival of mechanical components in the vacuum are induced. Although motions can be created by placing motors in the vacuum directly [1], it is not appropriate to execute in high vacuum because lubricants will vaporize to result in bearing failures or contaminate other components. Moreover, it is very difficult for the heat generated by the motor to dissipate in high vacuum environment.

In order to prevent vacuum robot from generating particles pollution by the actuator, motors are installed in atmosphere and torques are transmitted to vacuum environment. Since the permanent magnetic couplings (PMC) have the ability of transmitting torque from a primary drive to the follower through magnetic field

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 657–665, 2009.

without mechanical contact. They are appropriate for torque transmission from outside atmosphere to robotic shaft in vacuum across a thin vacuum containment barrier. The sealing achieved with PMC provides thoroughly isolation by static seal instead of dynamic one. Robotic arms operating in the vacuum processing chamber are magnetically coupled to direct drive (DD) motors outside the vacuum area. As the exterior shaft is driven by a DD motor, the interior component moves accordingly by magnetic force. The transmitted torque is determined by the rotational lag and the magnetic flux density. If the maximum coupling torque and the maximum torsion angle are exceeded, the power transmission is interrupted and thus PMC act as an overload protection function of the robot.

The PMC possess dominant placement for vacuum robot mechanical design, fabrication and assembly, so accurate analysis and optimal design of PMC are essential to robot performance. Various methods for analyzing the magnetic field and the transmitted torque of PMC have been proposed, including analytical methods and finite element analysis (FEA) methods. Elles and Lemarquand [2] presented the analytical formulas of the force between two magnets, which can be used to compute and optimize the torque of coaxial synchronous magnetic coupling, as well as to calculate the radial force and the radial stiffness of the cylindrical air-gap couplings. He [3] provided a method based on a cylindrical current sheet model for the analysis and design of induction-type coilgun. Wallace [4] proposed that the power losses were proportional to the clearance and offset between primary and secondary rotors by carrying out an experiment on the misalignment of an eddy current coupling. Huang [5] theoretically analyzed the effects of misalignment on the transmission characteristics of axial PMC, derived the formulas of the torque and force interaction characteristics, and then proposed that the existence of misalignment would reduce the transmitted torque and induce extra transversal force. The static characteristics of PMC with misalignment in the limited range are similar to those for well-aligned couplings. In addition, the efficiency ratio was employed to verify dynamic characteristics of PMC. Elles and Lemarquand [6] analytically researched the radial stability of synchronous couplings and presented a new structure for unstable couplings as mechanical filter. In addition, FEA methods have been widely employed for design and verification of PMC. Fereira [7] analyzed a radial type coupling using both 2D and 3D FEA methods. Although 2D FEA is much easier to program and quicker to compute, 3D FEA method is proved to be more accurate and effective for the analysis and optimal design of PMC because it takes the end leakage effects into account. Wang [8] demonstrated a new 3D FEA optimal design method, which combines the orthogonal test and the method of exhaustion in small range to find the optimum geometry of multi sets of coaxial PMC. Kim [9] studied the dynamic characteristics of a machine tool spindle system with magnet couplings; simulated and analyzed the transverse and torsional vibrations; and found that the deflection at a main spindle was much smaller than at excitation part.

However, these researches mostly focus on the torque calculation and the optimization of PMC or effects of misalignment on the static transmission characteristics. There are few reports on dynamic characteristics and the effects of misalignment on tracking performance of synchronous coaxial PMC. The tracking performance is

essential for vacuum robot to transfer wafer smoothly and precisely. This paper describes the dynamic characteristics of coaxial PMC, and verifies the tracking performance under well aligned situation at different velocities. The influences of radial offset and axial shift misalignments on tracking errors are discussed. The cross coupling effects of two sets of couplings are minimized by 3D FEA simulation. Finally, experiments show excellent performance of PMC and prove that they are feasible and reliable for the application.

2 Dynamic Characteristics of PMC

The dynamic model of synchronous coaxial PMC is shown in Fig.1. In the stationary states the north and south poles of the magnets are opposite to each other and the magnetic field is symmetric. When the external shaft is twisted, the magnetic field lines are moved. Hence the torque is transmitted through the air gap and a synchronous rotation with a constant torsion angle is produced. Transmitted torque depends on the relative angular displacements of magnets attached to couplings of the driver and the follower. The relation was proposed by Yonnet [10] as a sine function in practice, it can be simplified like equation (1):

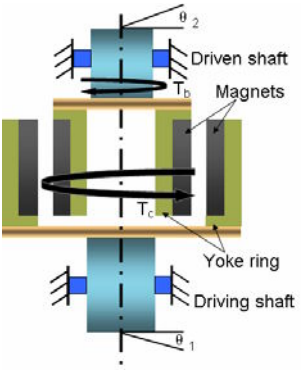


Fig. 1. Dynamic model of coaxial PMC

$$T_c = K \sin \varphi_c \tag{1}$$

Where, T_c is the transmitted torque between the PMC, K is the equivalent torsional stiffness of the PMC, φ_c is angle difference. When the relative angular displacement is in a small range, the relation between transmitted torque and the angular offset can be written as:

$$T_c = K(\theta_1 - \theta_2) \tag{2}$$

where, θ_1 and θ_2 are the angular displacements of the driving and driven couplings shaft respectively

2.1 Tracking Performance without Misalignment

An experimental apparatus is constructed to evaluate the dynamic characteristics of synchronous coaxial PMC. Fig.2 shows the structure of the test apparatus which consist of the following principal elements: X, Y and Z motion tables which can pre-align the PMC initially and provide both radial (X, Y) and axial (Z) movements for misalignment experiments. A DD motor (190ST2M-111K010H, ALXION) is physically fixed on the primary driving shaft with an optical encoder (RESM20USA075, RENISHAW) placed on the secondary shaft. In order to evaluate the tracking performance accurately, the PMC are adjusted to be stationary and well coaxial.

The PMC adopted in this paper are cylindrical couplings with iron yokes and magnets located separately. The magnetic material is Nd-Fe-B and yoke material is steel 45. The yoke has high permeability for the inner and outer magnet rings to avoid the loss of magnetic leakage. The air-gap between the couplings equals 5mm, enabling the vacuum partition wall to have sufficient thickness and structural stiffness. The designed torque is 75 N.m and the actual measured maximum transmitted torque is 72.2 N.m.

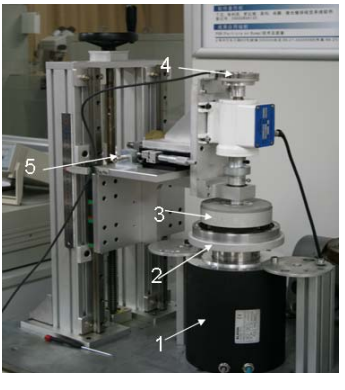
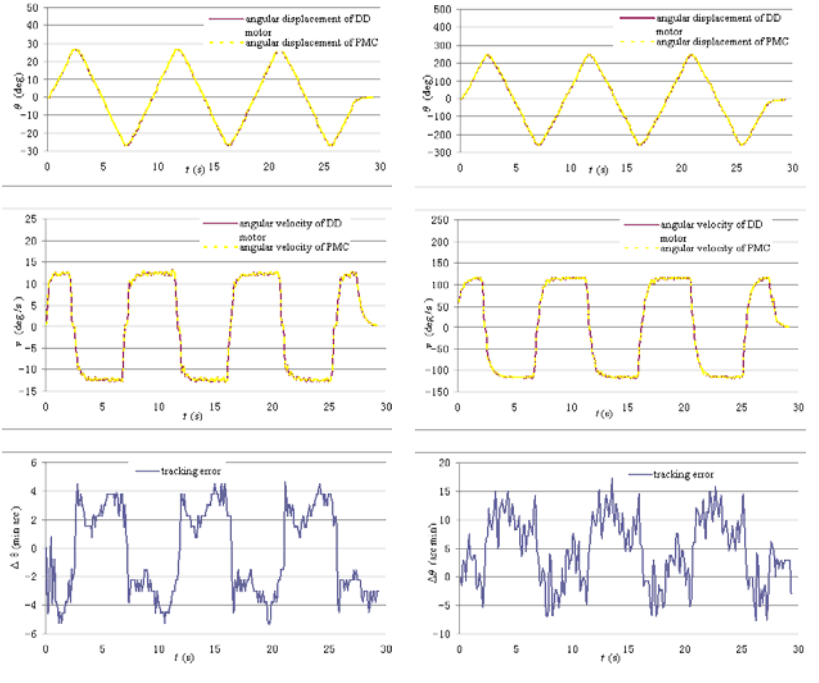


Fig. 2. Photo of the test apparatus. 1) Direct drive motor 2) Outer PMC shaft 3) Inner PMC shaft 4) Optical encoder 5) X, Y and Z motion table.

In this study, two groups of prescribed angular motions shown in Fig.3 are taken as the input. The DD motor drives the active shaft of PMC with triquetrous shape functions at different velocities. The passive shaft rotates as quickly as the active shaft by the magnetic force between the inner and outer magnetic rings. The solid lines indicate the angular displacement and velocity profiles of DD motor and the dotted lines indicate the angular displacement and velocity profiles of the passive shaft respectively. The tracking error is defined by $\Delta\theta=\theta_1 - \theta_2$ Because the PMC are optimal

designed with high torsional stiffness, the driven shaft rotates nearly synchronous with the driving shaft.

Some conclusions can be drawn from the graphs that the tracking error increases when the driving shaft starts to accelerate, and the response time is less than 0.01s. The tracking error oscillates when the angular velocity is maximal. When the motor changes rotation direction suddenly, the tracking error decreases rapidly and then oscillates in the other side. Comparing the tracking errors at different velocities, the maximum values increases from near 4.8' to 17.2' with angular velocities changing from 12deg/s to 120deg/s, and the tracking error oscillates more drastically at high velocity than low velocity. The result proves that the tracking error is closely related to the velocity profile of the input due to the magnetic viscous friction.



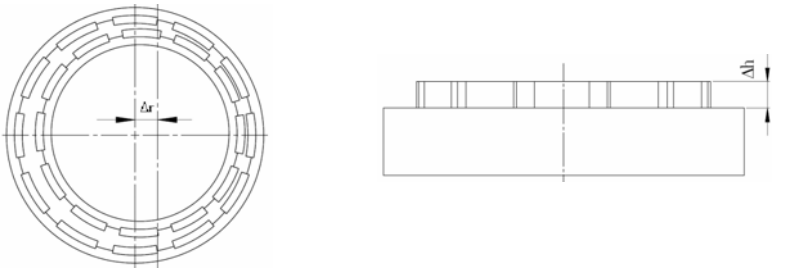
(a) (b)

Fig. 3. Tracking performance without misalignment

2.2 Tracking Performance with Misalignment

The PMC are the critical components in the vacuum robot design and fabrication. In practice, misalignment may happen due to assembly errors or other component's failure. Two typical types of misalignments are shown in Fig.4. In Fig.4 (a) radial offset may decrease the distance of magnets between inner and outer rings. As the

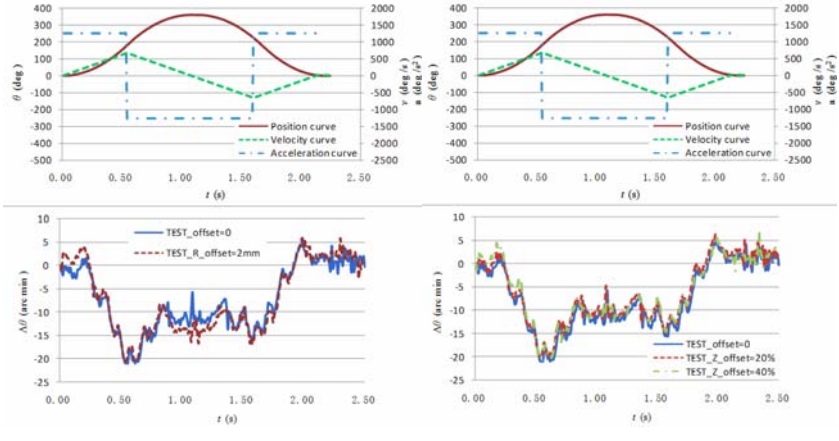
induction strength is inversely proportional to the square of the distance from the magnet surface, the radial force increases more and more rapidly [6]. Fig.4 (b) shows the axial shift may reduce the average distance between the magnetic faces, therefore the induction and the torque decreases too. The existence of misalignment causes torque changes which will reduce the life cycle of couplings and transmission system.



(a) Radial offset (b) Axial shift

Fig. 4. Shaft misalignment conditions

In this study, an S-shape angular motion is considered to be the input for both experiments. The effect of tracking error with radial offset 2mm comparing to well centered situation is shown in Fig.5 (a) And the effects of tracking error with axial shift 6mm and 12mm comparing to well centered situation are shown in Fig.5 (b). The solid lines indicate tracking error well aligned and the dashed line means the tracking error with misalignments. The maximum misalignment in both situations is 40% compared to the initially aligned condition.



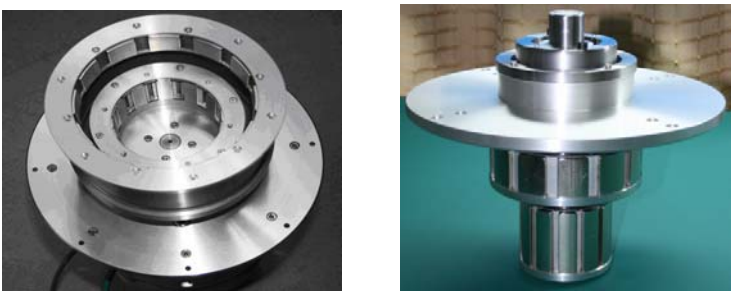
(a) radial offset (b) axial shift

Fig. 5. Effects of misalignment on tracking error

Ideally, the active and passive rings rotate synchronously without mechanical friction. In the misalignment dynamic experiments, it was found that the magnetic coupling continues to rotate synchronously despite the existence of axial shift or radial offset. Results show that shaft misalignment does not affect the tracking performance to a large extent. Neither axial shift nor radial offset changes maximum tracking error significantly because of high designed tortional stiffness. However, misalignment will break the symmetry of PMC structure, result in unbalanced magnetic force and cause extra vibration. Comparing the measurements of two type misalignments, axial shift does not change the uniform and periodic distribution of magnetic field, the variations of the maximum tracking error is less than radial offset.

3 Cross Coupling Effects of Two Sets of PMC

In applications, a coaxial mechanism is usually incorporated into the design of a compact wafer transfer robot. The robot should be able to perform radial linear extending and retracting movement, rotation movement and vertically up-down movement. Both the radial linear and rotation movements are transmitted by PMC. With two sets of PMC being used, they are crossed coupled. That, is, as one pair of couplings rotates, it may make the other pair to move accordingly. The effects of cross coupling results in the positioning error in motion control. Fig.6 (a) shows the structure of out rings and (b) shows the inner rings of PMC assembled separately.

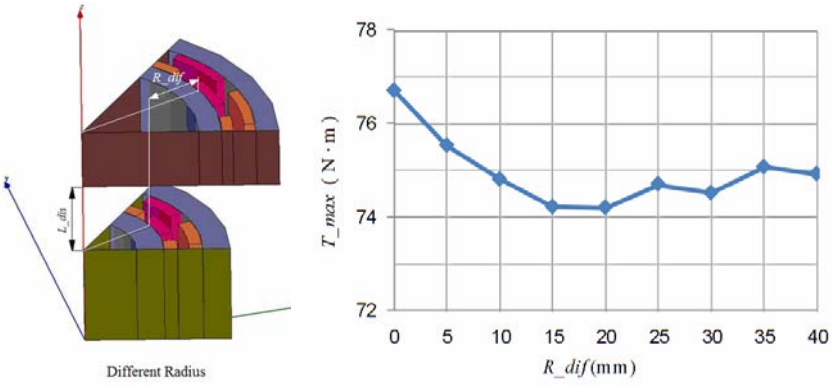


(a) (b)

Fig. 6. Structure of two sets of PMC

The cross coupling effects can be minimized by increasing the spatial distance between two sets of PMC. However, which will increase the dimension of the vacuum chamber and decrease the contamination capability. The spatial distance should be lessened in mechanical design. The closer the two couplings are, the more serious the effects of cross coupling will be. Methods for minimizing cross coupling include the use of magnetic shields to cover the magnets and the rearrange of magnetic pole orientations. This paper proposes another method to reduce effects of cross coupling by increasing the radial difference $R_{dif} = R_{PMC1} - R_{PMC2}$ between the two sets of PMC.

The 3D FEA module of Ansoft's Maxwell 11 is used to simulate and minimize the cross coupling effects. Fig.7 (a) shows the 3D model of PMC groups. In the simulation the lower PMC1 are kept unchanged and the axial distance L_{dif} is set to be 5mm. The radial difference R_{dif} is increased from 0 to 40mm by step length of 5mm. The maximum torque of upper PMC2 is shown in Fig.7 (b). It illustrates that when R_{dif} is less than 15mm, the cross coupling will augment the maximum torque of the upper PMC2 obviously, and the effects of cross coupling decrease as the radial difference increases. When R_{dif} is larger than 25mm, the cross coupling will have negligible influences on each other.



(a) 3D model (b) Simulation result

Fig. 7. 3D FEA simulation of cross coupling

An experiment is carried out to verify the proposed method. The radial difference is 25mm. The angular displacements in Fig.8 show that there are few effects of cross coupling on each other. When PMC1 are driven by the DD motor, PMC2 shake less than 1.5' accordingly then return back. It is due to the distribution of the two PMC magnetic field does not intervene each other strongly.

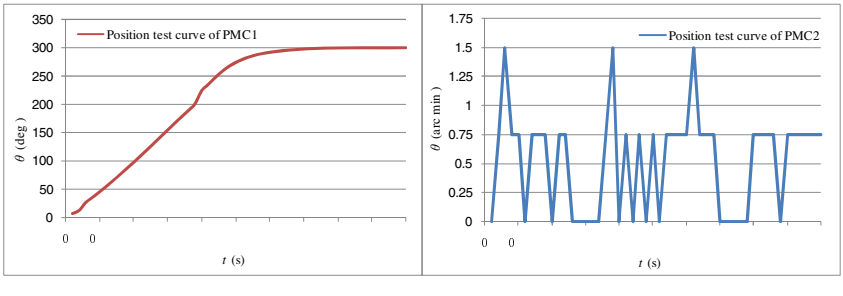


Fig. 8. Experimental result of cross coupling effects

4 Conclusion

This paper describes the dynamic characteristics of synchronous coaxial PMC, presenting an experimental investigation of the tracking performance when well aligned and with radial offset or axial shift misalignments. Experimental results show that PMC shaft performs excellent synchronously when well centered; the tracking error is closely related to the velocity profile of the input; the response time is less than 0.01s; and the tracking error increases barely even with misalignment increases up to 40%. Since in practical applications only limited assembly tolerance may occur, the tracking performance is reliable for vacuum robot to work precisely. The effects of cross coupling are simulated by 3D FEA method, which can be minimized by increasing the radial difference. The synchronous coaxial PMC are verified to be feasible and reliable to be applied in the vacuum robot.

Acknowledgments. This work was supported in part by the Science & Technology Commission of Shanghai Municipality under Grant 07PJ14051 and 071111008. Also it is supported in part by the National Natural Science Foundation of China under Grant 50805095.

References

1. Tsuzuku, C.: The trend of robot technology in semi-conductor and LCD industry. Industrial Robot 28(5), 406–413 (2001)
2. Elies, P., Lemarquand, G.: Analytical optimization of the torque of a permanent-magnet coaxialsynchronous coupling. IEEE Transactions on Magnetics 34(4 Part 2), 2267–2273 (1998)
3. He, J.L., et al.: Analysis of induction-type coilgun performance based on cylindricalcurrent sheet model. IEEE Transactions on Magnetics 27(1), 579–584 (1991)
4. Wallace, A., Wohlgemuth, C., Lamb, K.: High efficiency, alignment and vibration tolerant, coupler using high energy-product permanent magnets. In: IEE Conference Publication (1995)
5. Huang, S.M., et al.: Effects of misalignment on the transmission characteristics of magnetic couplings. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science 215(2), 227–236 (2001)
6. Elles, P., Lemarquand, G.: Analytical study of radial stability of permanent-magnet synchronous couplings. IEEE Transactions on Magnetics 35(4), 2133–2136 (1999)
7. Furlani, E.P., et al.: A three-dimensional model for computing the torque of radialcouplings. IEEE Transactions on Magnetics 31(5), 2522–2526 (1995)
8. Wang, Y.L., et al.: Near-Optimal Design and 3-D Finite Element Analysis of Multiple Sets of Radial Magnetic Couplings. IEEE Transactions on Magnetics 44(12), 4747–4753 (2008)
9. Kim, S.K.: Dynamic System Analysis of Machine Tool Spindles with Magnet Coupling. International Journal of the Korean Society of Precision Engineering 3(4), 87–93 (2002)
10. Yonnet, J.P.: Permanent magnet bearings and couplings. IEEE Transactions on Magnetics 17(1), 1169–1173 (1981)

Complex Surface Machining: Thermo-mechanical Analysis for Error Prediction of Low-Rigidity Workpiece

Yongan Huang[1], Huimin Liu[1,2], Zhouping Yin[1], and YouLun Xiong[1]

[1] State Key Laboratory of Digital Manufacturing Equipment & Technology, Huazhong University of Science and Technology, 430074, Wuhan, China
yahuang@hust.edu.cn
[2] Department of Weaponry Engineering, Navy University of Engineering, 430033, Wuhan, China

Abstract. A thermo-mechanical analysis is presented to predict the cutting force- and temperature-induced deflection in machining low-rigidity workpieces. Firstly, the cutting forces in ball-end milling are discussed. The theoretical flexible force model is considered to model the cutting force due to the coupling effect between force and deflection. Meanwhile, the thermal deformation is studied by including the dynamic temperature load, and it is to be combined into the flexible force model. The cutting force is given by geometric parameter method, the temperature at interface by empirical formula, and the dynamic temperature distribution by physical model. To take into account the deflection-force-temperature-deflection dependency, the workpiece geometry needs to be iteratively updated in computation. Last, the finite element analysis (FEA) is adopted to calculate the deformation.

Keywords: Errors prediction, Finite element analysis, Thermo-mechanical model, Digital manufacturing, Cutting force.

1 Introduction

Low-rigidity parts lie widely in aero-engine, screw propeller, turbo-machinery, which have thin-walled sections with length-thickness ratio greater than 10. In stable milling of low-rigidity workpiece, surface dimensional error due to the workpiece deflection becomes a dominant problem that affects machining precision [1, 2]. Because large low-rigidity workpieces are usually thin-walled, and complex shapes and irregular curvature distribution, it results in many challenges in machining.

The accuracy of the surface profile is one of the most important components of both dimensional and geometric accuracy and usually directly related to the product's functional performance [2]. The direct trial-and-error approach is often expensive and time consuming. Producing the right profile in such parts increasingly depends on specialized packages for defining appropriate cutting strategies and tool paths. However, most are based on idealized geometries and do not take into account the factors such as variable cutting force, thermal load, part/tool deflection, etc [2].

Simulation of machining process is very important to satisfy tight tolerances, which usually involves finite element analysis (FEA), modeling of cutting force,

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 666–677, 2009.

temperature distribution and material removal. The existing methods for predicting the milling forces and deflections can be divided into experimenting and modeling. The force modeling research has been mainly focused on theoretical rigid force models or mechanistic force models that consider the effect of tool/part deflection during machining[3]. Wang et al.[4] studied force-induced errors from the measured force and error, and developed a static/quasi-static error compensation method. Law et al. [5] calculated the force from the measured milling torque, and developed a compensation methodology to integrate both the force and deflection models.

Their applicability to model force in machining of low-rigidity workpiece is limited due to the variability of material properties, cutting force, non-linear dependency on tool immersion angle and chip thickness [3]. The cutting force-induced error will be one of the major error sources in hard machining, low-rigid workpiece machining. Ratchev et al. [6] proposed a flexible force model to study the tool deflection. The model is based on an extended perfect plastic layer model. They considered the end-milling cutter as a cantilever with the force acting at the cutter tip centre position [3]. So far, few thermo-mechanical analysis have been employed to consider the force and temperature-induced deflection. In the cutting process there are many independent influencing factors which including the cutting parameters (feed rate, cutting speed, cutting depth), material properties, properties of the machine-tool-workpiece system, material and tool geometry. The main contribution of our paper is to establish a thermo-mechanical analysis for low-rigidity workpieces where the force- and temperature-induced deflection that is not taken into account by the most existing simulation.

2 Modeling of Cutting Force and Temperature

The surface dimensional error is induced mainly by the deflection of the tool and the workpiece in the case of low-rigid workpiece milling. Most of the existing techniques are based on idealized geometric profile and do not take into account part/tool deflection, which results in a significant deviation between the planned and machined part profiles 3]. Reliable quantitative predictions of the cutting force and temperature components in machining operations are essential for determining geometrical errors of machined workpieces. Force predictions are required for arriving at constrained optimization strategies in computer-aided process planning [7]. Machining process for low-rigidity workpiece consists of various activities defining a digital chain:

1) Definition of a CAD model for the low-rigidity workpiece;
2) Generation of the toolpath by CAM for multi-axis machining;
3) Deflection and force computation by CAE for interface of cutter and workpiece;
4) Optimization of the toolpath for machine tool;
5)Transformation of the data to machining code;
6) Driving and monitoring of the process, etc..

When the workpiece is large and flexible, the force and temperature-induced deformations have to be analyzed in virtue of CAE tool. Steps 3 and 4 are additional process for large low-rigidity workpieces, and do not exist in block ones.

2.1 Cutting Forces in Ball-End Mill

A general model for the determination of cutting forces in ball-end milling operations is presented. The basic concept is the mathematical representation of relations between the milling cutter and the workpiece, the change of chip thickness and the milling cutter rotation angle. The cutting forces are divided into differential cutting forces depending on the number of cutting edges, cutting edge length and milling cutter rotation angle. The following is the main idea how to calculate it [8, 9].

The lag between the tip of the flute at $z=0$ and at axial location z is

$$\psi = \frac{z}{R_0}\tan i_0 \tag{1}$$

A point on the flute j at height z is referenced by its angular position in the global coordinate system,

$$\Psi_j(z) = \theta - \psi + (j-1)\frac{2\pi}{N_f} \tag{2}$$

where N_f is the number of flutes, θ is the tool rotation angle.

The angle position in the direction of z-axis from the center of the hemispherical part to the point on the cutting edge can be written as

$$\begin{cases} f(z) = \sqrt{2R_0(R_0+z)} \\ \psi(z) = z\tan i_0 / R_0 \\ \kappa(z) = \arcsin\left(\sqrt{1-(1-z/R_0)^2}\right) \end{cases} \Rightarrow \begin{cases} x = R(\psi)\sin(\psi) \\ y = R(\psi)\cos(\psi) \\ z = R_0\psi\cot i_0 \end{cases} \tag{3}$$

The tangential, radial and binormal components are calculated as [9, 10]

$$\begin{cases} dF_t(\theta,z) = K_{te}dS + K_{tc}f_{zb}\sin\Psi\sin\kappa db \\ dF_r(\theta,z) = K_{re}dS + K_{rc}f_{zb}\sin\Psi\sin\kappa db \\ dF_a(\theta,z) = K_{ae}dS + K_{ac}f_{zb}\sin\Psi\sin\kappa db \end{cases} \tag{4}$$

where $t_n(\Psi,\theta,\kappa) = f_{zb}\sin\Psi\sin\kappa$ is the uncut chip thickness normal to the cutting edge, and varies with the position of the cutting point, K_{tc}, K_{rc}, K_{ac} (N/mm^2) are the shear specific coefficients, K_{te}, K_{re}, K_{ae} (N/mm) are the edge specific coefficients. dS (mm) is the length of each discrete elements of the cutting edge, $dS = \sqrt{(R'(\psi))^2 + R^2(\psi) + R_0^2\cot^2 i_0}\,d\psi$. f_{zb} is the feeding per tooth, and $\kappa = \arcsin(R(\Psi)/R_b)$. db (mm) is the differential length of cutting edge. It can be noted that it is consistent with the chip width in each cutting edge discrete element. In many mechanistic models for the milling process, the milling force coefficients K_{tc}, K_{rc}, K_{ac}, K_{te}, K_{re}, K_{ae} are established from specially devised milling tests and

mechanistic analysis. Usually there are two methods to predict the parameters which are mechanistic evaluation and prediction from an oblique cutting model [7].

The cutting forces in Eq.(4) are modeled in terms of two fundamental phenomena, an edge force component due to rubbing or ploughing at the cutting edge, and a cutting component due to shearing at the shear zone and friction at the rake face [7]. The cutting force model including explicitly the ploughing component can obtain more precise prediction accuracy.

Once the local tangential $F_t(\theta,z)$, radial $F_r(\theta,z)$, and axial $F_a(\theta,z)$ cutting force on the tooth-workpiece contact point are determined, the resultant forces in Cartesian coordinates are obtained by introducing the transformation matrix T [9]

$$\left\{dF_{x,y,z}\right\}=[T]\left\{dF_{r,t,a}\right\} \tag{5}$$

where, $\left\{dF_{x,y,z}\right\}=\begin{bmatrix}dF_x\\dF_y\\dF_z\end{bmatrix}$, $[T]=\begin{bmatrix}-\sin\kappa\sin\Psi & -\cos\Psi & -\cos\kappa\sin\Psi\\-\sin\kappa\cos\Psi & \sin\Psi & -\cos\kappa\cos\Psi\\\cos\kappa & 0 & -\sin\kappa\end{bmatrix}$, $\left\{dF_{r,t,a}\right\}=\begin{bmatrix}dF_r\\dF_t\\dF_a\end{bmatrix}$.

Then one can get

$$\begin{cases}F_{xj}(\theta)=\int_{z_1}^{z_2}\left(-\sin\kappa_j\sin\Psi_j dF_{rj}-\cos\Psi_j dF_{tj}-\cos\kappa_j\sin\Psi_j dF_{aj}\right)dz\\F_{yj}(\theta)=\int_{z_1}^{z_2}\left(-\sin\kappa_j\cos\Psi_j dF_{rj}+\sin\Psi_j dF_{tj}-\cos\kappa_j\cos\Psi_j dF_{aj}\right)dz\\F_{zj}(\theta)=\int_{z_1}^{z_2}\left(\cos\kappa_j dF_{rj}-\sin\kappa_j dF_{aj}\right)dz\end{cases} \tag{6}$$

Since the cutting force coefficients (K_{tc}, K_{rc}, K_{ac}) may be dependent on the local chip thickness, the integrations given above should be calculated digitally by evaluating the contribution of each discrete cutting edge element at dz intervals [9]. The oblique cutting force components are obtained from orthogonal cutting force using an orthogonal to oblique cutting transformation method [7]. For example, when Cutter radius is 3mm, Feed is 0.02mm/rev, Depth of cutting is 0.02m, the cutting force can be calculated as Fig.1.

2.2 Flexible Force Model for Low-Rigidity Workpiece

Most of the reported papers that has been carried out in the area of cutting force-induced error belongs to those caused by large deformation of thin-walled or low-rigidity workpieces under load [11]. Peripheral milling of flexible components is complicated by periodically varying cutting-forces which statically and dynamically excite the tool and workpieces leading to significant and often unpredictable deflections. Static deflections produce dimensional form errors, and dynamic displacements lead to poor surface finish quality [6, 12].

There are two kinds of force model which are theoretical rigid force model and adaptive theoretical flexible force model [6]. There is a high complexity associated with modeling of cutting forces in machining of low-rigidity workpieces due to the

variable part/tool deflection and changing tool immersion angle. To resolve this complex dependency an interactive approach integrating an extended perfect plastic layer force model needs to be applied which links force prediction with part deflection modeling [6]. The predicted profile of the workpiece is adopted to identify the "real" material volume that is removed during machining instead of the "ideal" one defined by the currently used "static" NC simulation packages [3].

Ratchev et al. [3, 6] proposed a flexible force model for machining dimensional form error prediction of low-rigidity components. Fig. 2 is a transient processing profile of low-rigidity workpiece during machining. The flexible force model has been discussed in detail in Ref. [3]. In this paper, the authors will introduce the thermal deflection into the the flexible force model to consider the temperature's effects.

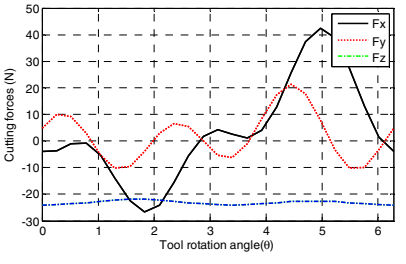


Fig. 1. Predicted cutting forces for slot cutting tests

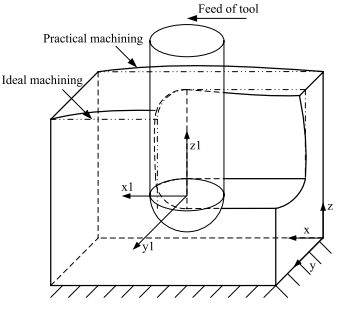


Fig. 2. Flexible milling geometry and coordinate system [3]

3 Thermo-mechanical Analysis for Low-Rigidity Workpiece

The errors are usually caused by excessive deformation at the interface of tool and workpiece due to the cutting force and temperature, so they have to be considered at the same time. The methodology is established based on modeling of cutting forces and temperature, prediction of deflection of the workpiece during machining.

3.1 Time-Dependent Heat Transfer in Machining

Temperature is one of the largest contributors to the dimensional errors of a workpiece in precision machining. Before calculating the temperature distribution, cutting temperature at the interface need to be predicted. The thermal source problem in machining low-rigidity workpiece with speed-, and feed-dependent boundary conditions is very difficult to be solved analytically. Therefore, the temperature prediction at the interface is usually achieved by non-linear empirical modeling approaches. The average interface temperature, as measured by the tool-work thermocouple, is[13]

$$T_{avg,interface}(^{\circ}C) = 1700V^{0.5}d^{0.2}f^{0.4} \tag{7}$$

where V is cutting speed(m/s), d is depth of cut(mm), and f is feed (mm/rev).

The above empirical model of the interface temperature is developed for turning of 4140 Steel alloy with tungsten carbide tools. The relation of cutting temperature and feed is plotted in Fig. 3 where the temperature at the interface is very high.

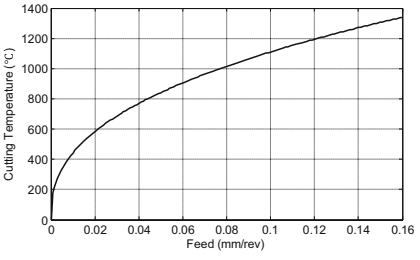


Fig. 3. Cutting temperature versus cutting feed, depth of cut 0.763mm, cutting speed 3m/s

The temperature variations are related to the heat source movement, heat source intensity, and thermal resistance coefficient. The fundamental generalized problem to be solved analytically is the heat conduction in a thin infinite plate with a convective and radial boundary condition on the face. The time-dependent heat transfer process is governed by the following differential equation [14],

$$\frac{\partial^2 T}{\partial r^2}+\frac{1}{r}\frac{\partial T}{\partial r}-\eta T+\frac{g(r,t)}{k}=\frac{1}{\alpha}\frac{\partial T}{\partial t} \tag{8}$$

where, $\eta = h/kw$, h is the convection coefficient of heat transfer, k is the thermal conductivity of the material, w is the plate thickness, α is the thermal diffusivity of the material, and $g(r,t)$ is the internal heat generation rate per unit volume, the variable r is the radial distance from the heat source, and the temperature at the interface can be determined by Eq.(7).

The dynamic non-uniform temperature distribution roots in the non-linear and time/position-dependent thermal deformations which is very difficult to be solved analytically. In general, it is calculated based on the assumption of either perfect thermal contact or perfect thermal insulation. Eq.(8) is to be solved by finite element method here. Ref. [14] proposed another computational methods.

There are two basic thermal error modes, namely, thermal expansion and thermal bending. Fig. 4 is the schematic diagram of machining results due to temperature effects. Fig. 4a is the top view and Fig.4b is the side view. It can be observed from Fig. 4 that the machining topography will be very complex when the temperature effects are considered.

3.2 Interactive Algorithm for Thermo-mechanical Analysis

The cutting forces in machining of low-rigidity workpieces depend on the chip thickness which is a function of the tool immersion angle. In machining low-rigidity parts the tool immersion angle is a function of the part deflection which itself depends on

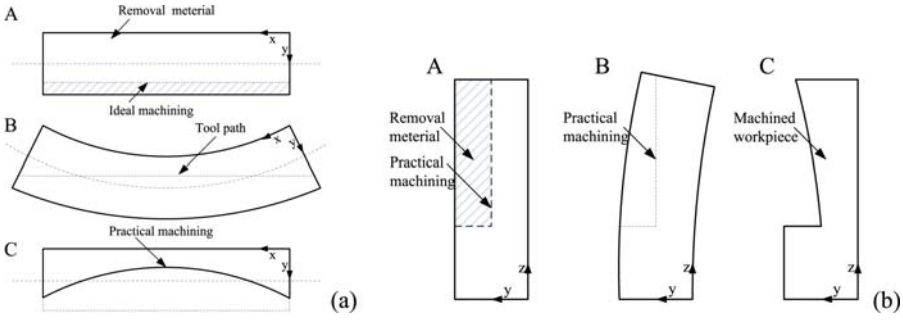


Fig. 4. Temperature-effected machining: (a) Top view; (b) Side view

the cutting forces [13]. The deflection is a function of the cutting force and temperature. The cutting forces depend on the chip thickness which is a function of the tool immersion angle, and the machining temperature is a function of the cutting speed, depth of cut and the feed. An iterative procedure is used to determine the milling error. The predicted deflected part profile is used to identify the practical material volume that is removed during machining instead of the ideal one defined by the current NC simulation packages. The thermo-mechanical analysis is outlined in Fig. 5. While milling a low-rigidity workpiece, as soon as the cutter is engaged, the workpiece deflects to a new position and the cutting temperature changes. In more detail, the tool-workpiece intersection line can be used as an example to explain the impact on the cutting force and deflection [6].

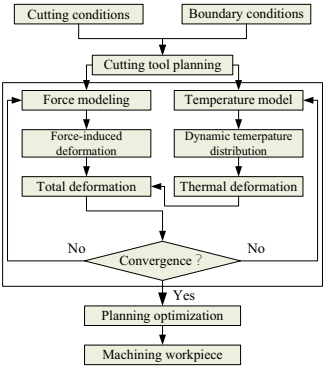


Fig. 5. Iterative "thermo-mechanical analysis"

3.3 Error Definition and Thermo-mechanical Deflection

The surface dimensional error is the normal deviation of the actual machined surface from the desired machined surface. For example, at point P in Fig.6, the corresponding surface dimensional error is e_P

$$e_P = \delta_{t,P} + \delta_{f,P} \tag{9}$$

where $\delta_{t,P}$ and $\delta_{f,P}$ are the normal projections of the temperature- and force-induced deflection corresponding to Point P, respectively, and they may have different signs according to the direction. For the convenience of study, the distance between the initial surface to be machined and the desired machined surface is named as the nominal radial depth of cut symbolized by R_N. In actual machining, to ensure that the surface dimensional error does not violate the tolerance, R_A is often specified to be different from R_N ($R_A \neq R_N$) [1]. In this case, Eq. (9) has to be adjusted to the calculation of surface dimensional error

$$e_P = \delta_{t,P} + \delta_{f,P} + R_N - R_A \tag{10}$$

Note that R_N and R_A are the nominal and specified radial depth of cut, respectively. For a certain surface generation line, the steps adopted to calculate the error distributions can be found in Ref. [1]. Fig.7 shows the section view that is perpendicular to the desired (nominal) tool path at the cutting position.

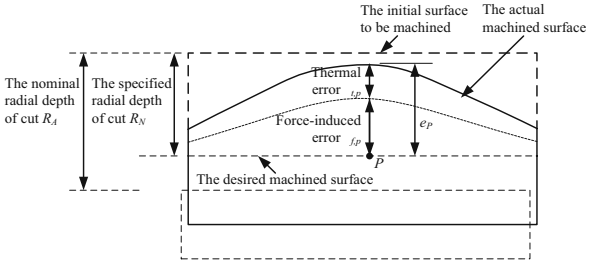


Fig. 6. Definition of the surface dimensional error

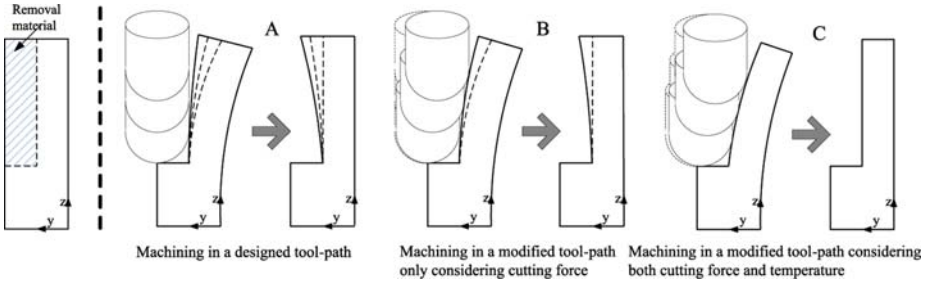


Fig. 7. Tool path optimization in end milling—view along the workpiece length direction

4 Examples and Case Studies

For simplification the part is assumed to be a thin-walled rectangular workpiece. The required machined profile is a flat surface parallel to the plane OXY. During milling,

the workpiece deflection in its thickness direction has a significant impact on forming the surface profile error. The contributions of the workpiece deflection in the feed direction and the tool axial direction can be ignored. Therefore, the investigation is focused only on error prediction in y-axis direction. The simulations were based on clamped-free-free-free cantilever plates with dimension $150\times120\times5mm^3$ and *Aluminium alloy 6063 T83*. The quasi-static cutting force is treated as a moving-distributed load acting on the workpiece-tool contact zone in the milling process.

FEA is employed to estimate the force and temperature-induced deflection. The input to finite element model is the predicted cutting force and temperature and a set of parameters describing material properties, boundary conditions and other constraints. In order to simplify the particularly complex simulation, we employ an assumption that the instantaneous stiff variation due material removal will be not taken into account. To compute the workpiece response, the continuous machining process was simulated by multi-step cutting processes.

During machining, the tool moves along the machining surface. Here, we take a different approach that uses a moving coordinate system fixed at the tool axis. After making the coordinate transformation, the heat transfer problem becomes a stationary convection-conduction problem that is straightforward to model (COMSOL's tutorial). Due to the cutting forces and temperature exerted by the cutting process, the workpiece becomes deformed. With the workpiece undergoing deformation, the cutting force is also changing considerably. These changes is taken into account by computing the cutter on a moving mesh attached to the workpiece.

4.1 Results of Error Prediction

4.1.1 Machining Based on a Designed Tool Path

In this section, the results based on the rigid force model are discussed. Supposed that the force is $0.23N/mm^{-2}$, the temperature is 1200 , and the two loads act on the interface workpiece and cutter at neighborhood of $x=75mm$ of the workpiece.

It can be noted that the max deflection in Fig.8 is 0.348mm and in Fig.9 are 0.384mm in the positive direction and 1.5mm in the passive direction, respectively.

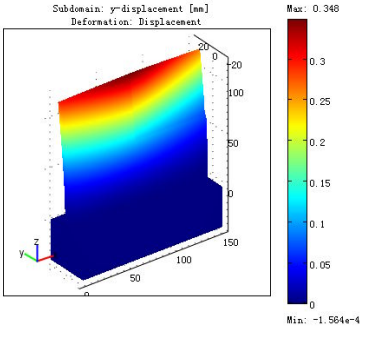


Fig. 8. Deflection results based on the rigid force model

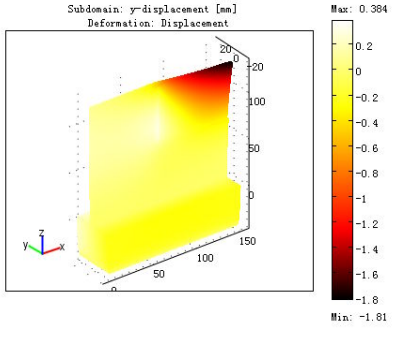


Fig. 9. Thermal deflection results

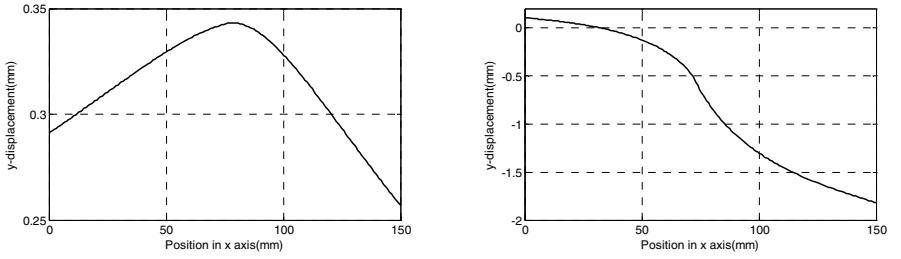


Fig. 10. Deflection of the top edge of the workpiece due to force and temperature

Fig.10 is the force-induced and temperature-induced deflection of the top edge of the workpiece. One can observe an interesting phenomenon that the deflections at $x = 75mm$ in the two figures have different direction. The force-induced deflection is along the positive direction, however the temperature-induced deflection along the passive direction. In reported studies, the force-induced and the temperature-induced deflection are investigated dividually, which will leads large error in machining.

4.1.2 Machining Based on Flexible Force Model

Fig.11 is the results based on the flexible force model. The max deflection is 0.278mm. Compared with Fig.8, one can observe that the max stress and the max deflection are all smaller. The simulation shows the deflection of workpiece affect the cutting force, and the cutting force also affect the deflection in return. The temperature, however, is not considered in the flexible force model. In the above subsection, one can know how important the temperature is in machining deflection.

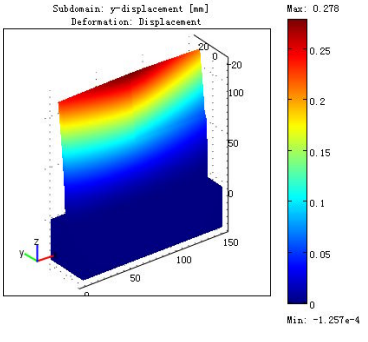


Fig. 11. Stress and deflection results based on the flexible force model

4.1.3 Machining Based on Thermo-mechanical Analysis

The results plotted in Figs.12 and 13 are based on thermo-mechanical analysis. The max deflection in Fig.12 is 0.527mm, which is smaller than the sum of 0.278mm of the flexible force model and 0.384mm of the temperature-induced deflection. So the

thermo-mechanical model is not a simple combination of the flexible force model and the thermal deflection. In practical error compensation, there will overcut or undercut if the force and the temperature studied separately.

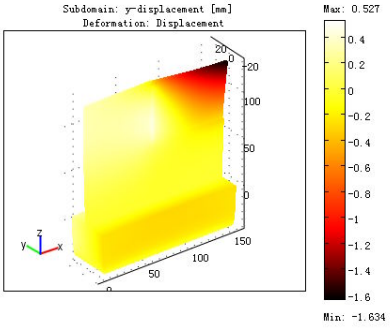


Fig. 12. Deflection results based on the thermo-mechanical analysis

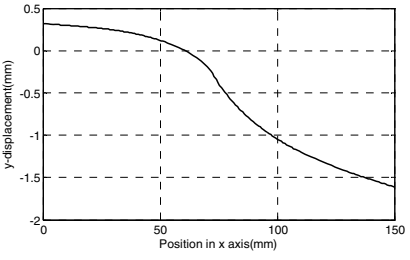


Fig. 13. Deflection of the top edge of the workpiece due to thermo-mechanical action

5 Discuss

In machining low-rigidity workpieces, the temperature and force-induced deflection contribute significantly to the surface error. The proposed methodology is based on coupling effects between cutting forces and temperature and their induced deflection during machining. There is still a knowledge gap in identifying the impact of deflection on the process of metal removal, and hence there are not systematic approaches to modeling, prediction of the component errors due to thermo-mechanical deflection in low-rigidity structures. The reported work is part of an ongoing research on machining of complex low-rigidity workpiece.

Acknowledgement. The authors are grateful for supports from the National Natural Science Foundation of China (50705035, 50835004 and 50625516).

References

[1] Wan, M., Zhang, W.H., Qin, G.H., et al.: Strategies for error prediction and error control in peripheral milling of thin-walled workpiece. International Journal of Machine Tools & Manufacture 48, 1366–1374 (2008)

[2] Ratchev, S., Liu, S.L., Huang, W., et al.: Machining simulation and system integration combining FE analysis and cutting mechanics modelling. International Journal of Advanced Manufacturing Technology 35, 55–65 (2007)

[3] Ratchev, S., Liu, S., Becker, A.A.: Error compensation strategy in milling flexible thin-wall parts. Journal of Materials Processing Technology 162, 673–681 (2005)

[4] Wang, S.M., Liu, Y.L., Kang, Y.A.: An efficient error compensation system for CNC multi-axis machines. International Journal of Machine Tools & Manufacture 42, 1235–1245 (2002)
[5] Law, K.M.Y., Geddam, A.: Error compensation in the end milling of pockets: a methodology. Journal of Materials Processing Technology 139, 21–27 (2003)
[6] Ratchev, S., Liu, S., Huang, W., et al.: A flexible force model for end milling of low-rigidity parts. Journal of Materials Processing Technology 153–154, 134–138 (2004)
[7] Budak, E., Altintas, Y., Armarego, E.J.A.: Prediction of Milling Force Coefficients From Orthogonal Cutting Data. Journal of Manufacturing Science and Engineering-Transactions of the ASME 118, 216–224 (1996)
[8] Milfelner, M., Cus, F.: Simulation of cutting forces in ball-end milling. Robotics and Computer-Integrated Manufacturing 19, 99–106 (2003)
[9] Lee, P., Altintas, Y.: Prediction of ball-end milling forces from orthogonal cutting data. International Journal of Machine Tools and Manufacture 36, 1059–1072 (1996)
[10] Lamikiz, A., de Lacalle, L.N.L., Sanchez, J.A., et al.: Cutting force estimation in sculptured surface milling. International Journal of Machine Tools & Manufacture 44, 1511–1526 (2004)
[11] Ramesh, R., Mannan, M.A., Poo, A.N.: Error compensation in machine tools - a review Part I: geometric, cutting-force induced and fixture-dependent errors. International Journal of Machine Tools & Manufacture 40, 1235–1256 (2000)
[12] Ratchev, S., Govender, E., Nikov, S., et al.: Force and deflection modelling in milling of low-rigidity complex parts. Journal of Materials Processing Technology 143–144, 796–801 (2003)
[13] Leshock, C.E., Shin, Y.C.: Investigation on cutting temperature in turning by a tool-work thermocouple technique. Journal of Manufacturing Science and Engineering-Transactions of the ASME 119, 502–508 (1997)
[14] Fraser, S., Attia, M.H., Osman, M.O.M.: Modelling, identification and control of thermal deformation of machine tool structures, part 1: Concept of generalized modelling. Journal of Manufacturing Science and Engineering-Transactions of the ASME 120, 623–631 (1998)

Modeling and Optimization of Contact Forces for Heavy Duty Robot Grippers

Qunming Li, Dan Gao, and Hua Deng

School of Mechanical and Electrical Engineering
Central South University, Changsha, 410083, China
liqm@mail.csu.edu.cn

Abstract. Different from dexterous robotic hands, the contact status of a two-finger heavy duty gripper between the two tongs and the object are much complex during forging operation, and the contact forces are difficult to be controlled in real-time, because the tong usually is designed to rotate freely around the arm to some extent, and the contact area is usually a surface or a line. Based on the force-closure condition to meet the force and the torque equilibrium equations, this paper presents a real-time calculation model considering the gripping contact areas as equivalent friction points for N robot fingers including four contact points for the heavy gripper gripping a cylinder object. Then the contact force optimization method for multi-fingered hand researches can be used for the gripping forces' calculation between gripper tongs and the forged object, and the task is formulated as an optimization problem on the smooth manifold of linearly constrained positive definite matrices, which has globally exponentially convergent solutions via gradient flows. This is a new approach to optimize the gripping forces in real-time for the gripper's design and control of heavy forging manipulators. Simulation and experimental results are analyzed.

Keywords: Heavy forging robot, gripper, contact force, stable gripping.

1 Introduction

Heavy duty grippers play an important role in a forging robot. They are the interface between the forged work piece and the whole forging manipulator. Now the gripping capacity of large-scale heavy duty forging robots is increasing greatly, the optimization of gripping forces and driving forces of the gripping mechanism becomes more and more important.

Most of the heavy gripping systems installed in industrial automations are designed to be mechanical two symmetrical-finger grippers, which are considered as the simplest efficient grasping configuration. The calculation of the gripping force of the hydraulic cylinder is evaluated statically at some special positions and orientations or by experiences of the designers. And there are little researches about the real-time optimization calculation and control of contact forces and gripping forces for large scale forging grippers. However, the work on manipulation by a multi-fingered robot

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 678–686, 2009.

hand has shown numerous results of theoretical, design and experimental nature [1-5]. The earlier linear and nonlinear programming techniques are off-line force optimization approaches, which cannot be implemented in real-time.

Buss, Hashimoto and Moore (BHM) [2] present the linearly constrained positive definite programming methods for online grasping optimization task for dexterous hands. They find that the nonlinear friction cone constraints are equivalent to the semi-definiteness of certain symmetric matrices, and formulate the grasping force optimization problem as an optimization problem on the Riemannian manifold of linearly constrained symmetric positive definite matrices which there are globally exponentially convergent solutions via gradient flows. Han et al [4] cast the friction constraints into linear matrix inequalities and formulate the optimization problem as a set of convex optimization problem. Helmke [7] proposes a quadratically convergent Newton algorithm for dexterous hand grasping.

Different from multi-fingered hands, the contact status of a two-finger heavy duty gripper between the two 'V'-shape tongs and the object are more complex during operation and the contact forces are difficult to be controlled in real-time, because the tong usually is designed to rotate freely around the arm an angle, and the contact area is usually a surface or a line, where the contact points may change indefinitely when forged. This paper considers the contact forces between the tongs and the work piece as active forces transformed from the hydraulic cylinder to the 'V'-shape contact areas during the forging manipulation, and presents an equivalent friction point gripping model including four contact points when gripping a round object for the heavy forging gripper. Then the optimization model of gripping forces is formulated based on BHM, and the optimization method of constrained gradient flows is used to determine the position of contact points and optimize the contact forces. Based on the optimal contact forces, the optimal gripping forces can also be obtained which can be used for the real-time control of driving forces of the hydraulic cylinder.

2 Configuration of Forging Gripper

Fig.1 is a simplified configuration of the typical forging gripper mechanisms. It can be seen that it is an under-constrained mechanism whose 'V'-shape tongs are free in a little wiggling ranges, and generally the contact areas to the grasped round object are four lines or four surfaces with discrete contacts when gripping a round object, which is different from dexterous robotic hands whose contact models are assumed to have at most one friction contact point for each finger. The calculation of contact forces of forging grippers is more complex than multi-fingered hands: for heavy duty grippers, the distribution of contact points is changed indefinitely during the operation. But to maintain the grasping stability and reliability either for an object grasped by a dexterous robotic hand or for a forged object gripped by a heavy gripper, the force closure condition is the same: the force and the torque balancing equilibrium must be satisfied. So the research results of dexterous robot hands can be used for reference to model the contact forces of heavy forging grippers.

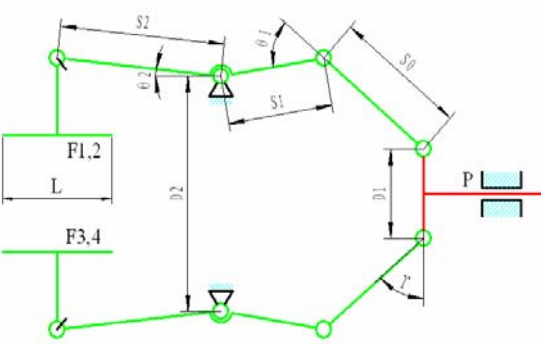


Fig. 1. The configuration of the heavy forging gripper mechanism

3 Modeling of Contact Forces between Tongs and the Forged Workpiece

Fig. 2 shows the simplified model of the large-scale heavy grippers. The coordinate system and the locations and orientations of the contact forces are shown in the figure. Assume that the system is rigid and continuous contact with friction on the contact surface, and the numbers of contact points between the object and the two 'V'-shape tongs are simplified as the four resultant forces with friction point contacts. The gripping model is equivalent to the grasp model as with four robot fingers. However the positions of the four contact forces may be changed within the contact area with external disturbance. The sufficient condition of stable gripping to balance the gravity of the forged object and external forces is force-closure gripping. The contact forces between the gripper and gripped object have the relation to the external wrench, which are determined by equilibrium equations as follows [5].

$$\mathbf{GF} = \mathbf{W} \tag{1}$$

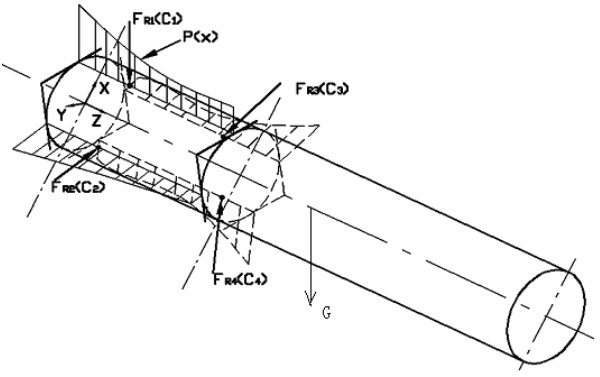
where $G \in R^{6\times 3m}$ is the grasp matrix, $\mathbf{F} = [\mathbf{F}_1^T, \mathbf{F}_2^T, \cdots, \mathbf{F}_m^T]^T$ is the contact force wrench of the grasp, m=4, $\mathbf{F}_i = (f_{i1}, f_{i2}, f_{i3})^T = \int_l \mathbf{P}_i(x, y, z)dz$ is the independent wrench intensity vector of contact point i, $\mathbf{P}_i \in R^3$ is the contact stress vector, $\boldsymbol{W} \in R^6$ is a 6D external wrench on the object.

The friction force constraint for each contact point is assumed to conform to Coulomb's law:

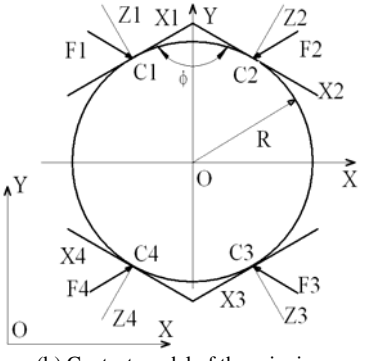
$$\sqrt{f_{i1}^2 + f_{i2}^2} \le \mu_i f_{i3} \tag{2}$$

where f_{i1} and f_{i2} are the tangential friction forces and f_{i3} is the normal force at the ith contact point, μ_i is the Coulomb friction coefficient.

According to [2], the friction cone constraints (2) are equivalent to the positive definiteness of P=Blockdiag(P_1,…,P_i,…,P_m), where P_i for contact i has the following forms:



(a) The distribution of contact stress of gripper



(b) Contact model of the gripping

Fig. 2. The equivalent contact model for heavy gripper

$$P_i = \begin{bmatrix} \mu f_{i3} & 0 & f_{i1} \\ 0 & \mu_i f_{i3} & f_{i2} \\ f_{i1} & f_{i2} & \mu_i f_{i3} \end{bmatrix} \tag{3}$$

P_i has identical diagonal elements and 0 for the (1,2) and (2,1) elements and can be rewritten in the general affine constraint [2]:

$$Avec(P) = \mathbf{q} \tag{4}$$

where $A \in R^{12\times144}$, $\mathbf{q} \in R^{144}$, $P \in R^{12\times12}$ (m=4), and vec($\bullet$) denotes the vec-operation. And the linear constraint (1) which is to balance the external force W is also easily rewritten in the general form of linear constraint (4). Now define cost function:

$$\Phi(P) = tr(W_p P + W_i P^{-1}) \tag{5}$$

where tr() denotes the trace of a matrix, W_p and W_i are two weighting matrices. According to the theorem of linearly constrained gradient flow, the constrained gradient flow –grad((P))=vec($\dot{P}$) is

$$vec(\dot{P}) = Qvec(P^{-1}W_iP^{-1} - W_p) \tag{6}$$

where $Q := I - A^+A = I - A^T(AA^T)^{-1}A$ is the linear projection operator onto the tangent space and + denotes the pseudo-inverse. The unconstrained elements of P converge exponentially to their unique equilibrium. The recursive algorithm of the gradient flow of P is implemented as

$$vec(P_{k+1}) = vec(P_k) + \alpha_k Qvec(P_k^{-1}W_iP_k^{-1} - W_p) \tag{7}$$

with a suitable step-size α_k and an initial P_0 satisfying all constraints.

4 Calculation of Initial Contact Forces

Given the desired external force $\boldsymbol{W}$ (including the weight of the forged object and the external disturbances) for the gripping, a suitable initial condition for optimization has to be found. In BHM [2], their solution is to increase the internal forces until the initial P_0 is inside the friction cones. Wang et al [6] presents a general method to obtain the initial grasp forces automatically by adjusting the Lagrange multipliers of the normal forces of each contact point.

Define a resultant force function:

$$J = \frac{1}{2}F^TF - \beta^TF \tag{8}$$

where $\boldsymbol{\beta} = \left[\boldsymbol{\beta}_1^T, \cdots, \boldsymbol{\beta}_i^T, \cdots, \boldsymbol{\beta}_k^T\right]^T$ is a weighting factor vector for the normal forces, $\boldsymbol{\beta}_i = [0,0,\gamma_i]^T$, γ_i is a constant. Further the contact forces need to balance the external force $\boldsymbol{W}$ as shown in equation (1), and the objective function can be formulated as

$$\Phi(F,\boldsymbol{\lambda}) = \frac{1}{2}F^TF - \boldsymbol{\beta}^TF + \boldsymbol{\lambda}(GF - W) \tag{9}$$

where $\boldsymbol{\lambda}$ is Lagrange multipliers. Let the gradients of Φ on F and $\boldsymbol{\lambda}$ equal to zero, we obtain

$$\begin{cases} F - \boldsymbol{\beta} + G^T\boldsymbol{\lambda} = 0 \\ GF - W = 0 \end{cases} \tag{10}$$

It can be rewritten as a matrix:

$$\begin{bmatrix} I & G^T \\ G & 0 \end{bmatrix}\begin{bmatrix} F \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta} \\ W \end{bmatrix} \tag{11}$$

Then the initial can be calculated as

$$F = G^+W + (I + G^+G)\boldsymbol{\beta} \tag{12}$$

where $G^+ = G^T(GG^T)^{-1}$ is the generalized inverse of G. Given an initial weighting vector $\boldsymbol{\beta}_0$ and a constant δ_0, an iterative form $\boldsymbol{\beta}_i^{k+1} = \delta_i \boldsymbol{\beta}_i^k$ with $\delta_i > 1$ to increase the internal forces can be used to calculate F_{k+1} until they satisfy the friction cone constraints.

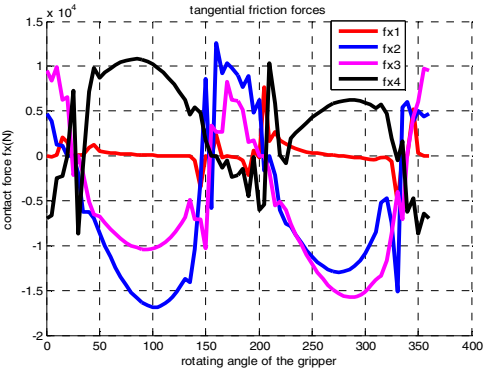
5 Simulation Results

Now consider a forging manipulator gripping a heavy work piece. The experiment is to grip a steel cylinder with 1.915T. The cylinder's outer diameter is 0.5m and its length is 3.957m. Refer to fig.1 and fig.2 (b), the relative design parameters of the gripper are listed in table 1. For simulation calculation, the 'V'-shape tong's span angle is changed from 90 , 100º, 110º, 120º, 130º to 134º respectively. The angle of the experimental tongs is designed as 134º. The pushing force P is produced by a hydraulic cylinder. Contact forces are formulated when the gripper mechanism grasps the 1.915T object by a given gripping force P.

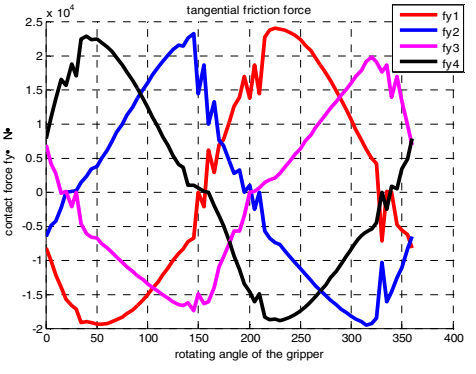
Now let the forging manipulator with 134º tongs grips the cylinder and keep it in horizontal direction and prevent the cylinder from slipping, we will calculate the optimal contact forces and gripping force when the gripper rotates a turn, and the cylinder won't slip to the floor in any angle. The initial position (0º) of the joints of the two tongs is vertical as show in fig.2 (b). Because the resultant contact point of the four contact forces may be changed within L ranges at different angles, first the distribution of the four resultant contact points must be found, then matrix $\boldsymbol{A}$ of equation (4) is calculated, and the initial contact forces $\boldsymbol{F}$ should be represented by form vec(P_0), and then by iterative algorithm of the gradient flow of P as shown in equation (7), the optimal contact forces can be found. According to the gripper mechanism and the structure of the 'V'-shape tongs as show in fig.1 and fig.2 respectively, the optimal resultant force of the gripping cylinder can be derived.

Table 1. Design data of the forging gripper

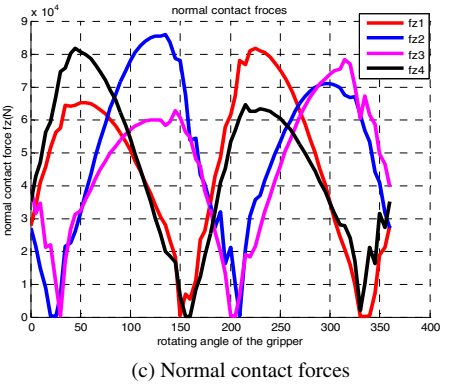
Symbol	Quantity	Unit
R	0.5	m
L	0.46	m
Φ	90 100 110 120 130 134	º
D_1	0.24	m
D_2	0.82	m
S_0	0.964	m
S_1	0.42	m
S_2	0.675	m
Θ_1	61	º
Θ_2	2	º
γ	49	º



(a) Tangential friction forces along x-axis



(b) Tangential friction forces along y-axis



(c) Normal contact forces

Fig. 3. Simulation results of contact forces during an operation

Fig.3 is the simulation results of one of the four contact forces when the gripper rotates to different angles. Fig.3 (a)~(c) show the three divisions of the resultant four contact forces, and fig.4 is the resultant contact forces onto the tongs and the driving force the hydraulic cylinder needs to provide. Where fx_i and fy_i (i=1~4) are two

tangential friction forces of contact points C_i, fz_i is the normal forces, PPjcs and PPjcx are the reverse contact forces of the upper and lower tongs from the tong to the object, PPjc is the composition of upper and lower resultant contact forces, PPjc= |PPjcs|+|PPjcx|, and PPower is the driving force exerted by servo hydraulic cylinder. From fig.4, we know that when the open direction of the gripper is horizontal (90º), the maximum gripping force of the hydraulic cylinder is 3.6×10^5N, and in vertical direction, the needed minimal driving force is 2.15×10^5N. We have measured the two driving forces at these special positions in an actual 3T forging manipulator which grips a rigid steel cylinder, keep the object with a small dropping angle, and the results are 4.2×10^5N and 2.6×10^5N respectively. Simulation and experimental results are very approximate. It demonstrates that the optimal algorithm is quite effective.

We are designing and developing a totally numerical controlled forging robot, which a force sensor is installed to measure and control the gripping force.

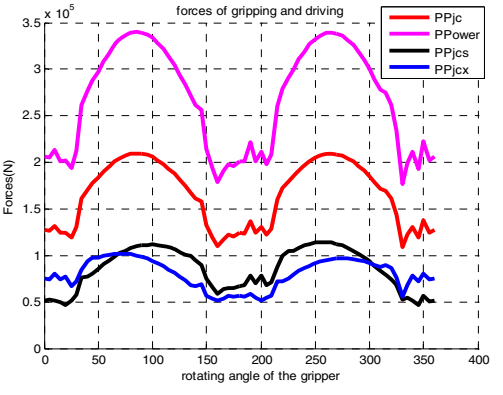


Fig. 4. Gripping and driving forces

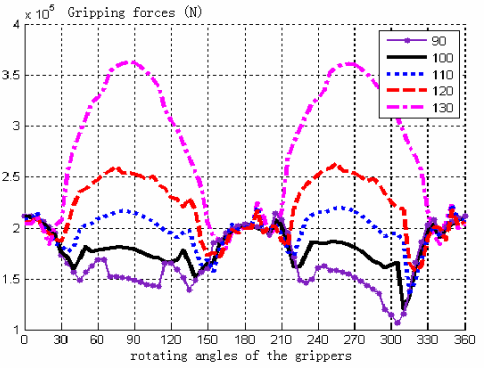


Fig. 5. Gripping and driving forces with different opening angles of the tong

6 Conclusions

This paper presents a gripping model which considers the resultant force of each contact surface as a contact force with an equivalent friction point. In order to obtain the closed-form solutions of contact forces, a three-dimension gripping-contact model is formulated based on the assumption of four contact points with friction between the V-shape tongs and the forged object. Then the heavy gripping mechanism can be simplified as a grasp with N robot fingers including four contact points. A real-time optimal algorithm of gripping forces based on BHM is used to calculate the contact forces. Contact forces and gripping force at different positions are demonstrated by simulation and experimental results. This is a new approach to optimize the gripping force in real-time, which is helpful for the optimal design and online control of the grippers for large-scale heavy forging manipulators.

Acknowledgments. This work was supported by National Basic Research Program of China (Grant No. 2006CB705400).

References

1. Wantanabe, T., Yoshikawa, T.: Grasping optimization using a required external force set. IEEE Trans. Robotics and Automation 44(1), 52–66 (2007)
2. Buss, M., Hashimoto, H., Moore, J.B.: Dextrous hand grasping force optimization. IEEE Trans. Robotics and Automation 12(3), 406–417 (1996)
3. Bicchi, A.: On the closure properties of robotics grasping. Int. J. Robotics Research 4(4), 319–334 (1995)
4. Han, L., Trinkle, J.C., Li, Z.X.: Grasp analysis as linear matrix inequality problems. IEEE Trans. Robotics and Automation 16(6), 663–674 (2000)
5. Murray, R.M., Li, Z.X., Sastry, S.S.: A mathematical Introduction to robotic manipulation, ch. 5. CRC Press, New York (1994)
6. Wang, B., Li, J., Liu, H.: Optimal grasping force computation for multi-fingered robot hand. J. Jilin University 38(1), 178–182 (2008)
7. Han, U., Huper, K., Moore, J.B.: Quadratically convergent algorithms for optimal dextrous hand grasping. IEEE Trans. Robotics and Automation 18(2), 138–145 (2002)
8. Li, Q., Gao, D., Deng, H.: Influence of Contact Forces on Stable Gripping for Large-Scale heavy Manipulator. In: Xiong, C.-H., Liu, H., Huang, Y., Xiong, Y.L. (eds.) ICIRA 2008. LNCS (LNAI), vol. 5315, pp. 839–847. Springer, Heidelberg (2008)

Performance of Inertia Force Analysis for Spherical Bearing Test Stand Mechanism

Yulin Yang[1,2], Xiong Du[1,2], Shaoshuai Fan[1,2], and Xijuan Guo[1,2]

[1] National College of Defence Technology, YanShan University
[2] School of Mechanical Engineering, YanShan University,
066004 Qinhuangdao, P.R. China
y.yang@ysu.edu.cn, dulaysu2006@hotmail.com

Abstract. According to inertia matrix, inertia force performance index of mechanism is proposed, which used to analyze the effect of size and mass on inertia force of mechanism. As an example, 2-RUUS mechanism is analyzed. Without considering those secondary components that have little influence on inertia force, inertia moment of main components are got, the relationship between performance of inertia force, size and mass of mechanism is analyzed and atlas is got. From the atlas, the reasonable range of size was found out.

Keywords: 2-RUUS mechanism, influence coefficient matrices, inertia tensor.

1 Introduction

In the past few decades, robotic technique has gained more and more attention. Control precision of robot is still an objective to improve further. So the study of inertial force performance of robot becomes a typical problem. Lots of performance indices are based on condition number of first-order influence coefficient matrix [1-7]. On the basis of the translation and rotation Jacobian matrices, two global performance indices are proposed in [8]. In literature [9], second influence coefficient matrix had much more important influence on acceleration performance of $4-{}^{x}R\,{}^{x}R\,{}^{x}U^{z}\,{}^{z}R$ mechanism. Because inertial force is related to acceleration, it is reasonable to analyze dynamic performance based on first-order and second-order influence coefficient matrices simultaneously. In recent years, there is lots of development in dynamics analysis of robot, but mostly analyzing acceleration according to above presented indices, then using Newton-Euler Law or Lagrange Law or other methods to analyze dynamics performances. However, the influence of size and mass on the dynamics performances are not taken into consideration completely.

In this paper, inequality of matrix norm between inertia force, size and mass of mechanism is given and inertia force global performance index of mechanism is put forward. As an example, dynamic performance of 2-RUUS mechanism, which is designed to simulate the kinematics of spherical bearing in helicopter swashplate mechanism, is analyzed. Then inertia force performance is analyzed and atlases of performance indices are got. The results will provide theoretical foundation of size optimization and be helpful to analysis and application of parallel mechanisms.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 687–695, 2009.

2 Global Performance Index of Inertial Force

From reference [10], acceleration of moving platform can be expressed as

$$\boldsymbol{a} = \dot{\boldsymbol{q}}^T \boldsymbol{H}_q^H \dot{\boldsymbol{q}} + \boldsymbol{G}_q^H \ddot{\boldsymbol{q}} \ , \tag{1}$$

where $\dot{\boldsymbol{q}}$ and $\ddot{\boldsymbol{q}}$ are first derivative and second derivative of generalized coordinate $\boldsymbol{q}(\in \boldsymbol{R}^6)$ with respect to time respectively. $\boldsymbol{G}_q^H$ and $\boldsymbol{H}_q^H$ are influence coefficient matrices of velocity and acceleration respectively.

Inertial force vector of moving platform can be expressed as [10]

$$\boldsymbol{F}_H^i = -\boldsymbol{I}_{hs}^0 \boldsymbol{a} - \begin{pmatrix} [\boldsymbol{\omega}_h]^T \boldsymbol{I}_{hc}^0 \boldsymbol{\omega}_h \\ \cdots \\ 0 \end{pmatrix} , \tag{2}$$

where $\boldsymbol{I}_{hc}^0$ is the inertial moment tensor matrix that its rank is 3. $\boldsymbol{I}_{hs}^0$ is the inertial force tensor matrix that its rank is 6. The elements of $[\boldsymbol{\omega}_h]$ is

$$[\boldsymbol{\omega}_h] = \omega_x \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} + \omega_y \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix} + \omega_z \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} . \tag{3}$$

And $\boldsymbol{\omega}_h = \begin{pmatrix} \omega_x & \omega_y & \omega_z \end{pmatrix}^T = G_q^h \dot{\boldsymbol{q}}$, it is

$$\boldsymbol{\omega}_h = G_1 \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \dot{\boldsymbol{q}} + G_2 \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix} \dot{\boldsymbol{q}} + G_3 \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \dot{\boldsymbol{q}} \ . \tag{4}$$

In equation (4), $\boldsymbol{G}_i (i = 1,2,3)$ is *i*th row of revolving influence coefficient matrix.

If set

$$[\ \ ] = \begin{bmatrix} [\ _h] & \\ & {}_{3\times3} \end{bmatrix} \in R^{6\times6} , \ I_h = \begin{bmatrix} I_{hc}^0 & \\ & \mathbf{0}_{3\times3} \end{bmatrix} \in R^{6\times6} , \ \boldsymbol{\omega} = \begin{pmatrix} \boldsymbol{\omega}_h \\ \mathbf{0} \end{pmatrix} \in \mathrm{R}^6 ,$$

then

$$\boldsymbol{F}_H^i = -\boldsymbol{I}_{hs}^0 \dot{\boldsymbol{q}}^{\mathrm{T}} \boldsymbol{H}_q^H \dot{\boldsymbol{q}} - [\boldsymbol{\omega}] \boldsymbol{I}_h \boldsymbol{\omega} - \boldsymbol{I}_{hs}^0 \boldsymbol{G}_q^H \ddot{\boldsymbol{q}} \ . \tag{5}$$

Set

$$\begin{cases} \boldsymbol{F}_{H1}^{i} = -\boldsymbol{I}_{hs}^{0}\dot{\boldsymbol{q}}^{\mathrm{T}}\boldsymbol{H}_{q}^{H}\dot{\boldsymbol{q}} \\ \boldsymbol{F}_{H2}^{i} = -\boldsymbol{I}_{hs}^{0}\boldsymbol{G}_{q}^{H}\ddot{\boldsymbol{q}} \\ \boldsymbol{F}_{H3}^{i} = -[\boldsymbol{\omega}]\boldsymbol{I}_{h}\boldsymbol{G}_{q}^{h}\dot{\boldsymbol{q}} \end{cases}, \tag{6}$$

then

$$\begin{cases} \left\|\boldsymbol{F}_{H1}^{i}\right\| = \left\|-\boldsymbol{I}_{hs}^{0}\dot{\boldsymbol{q}}^{\mathrm{T}}\boldsymbol{H}_{q}^{H}\dot{\boldsymbol{q}}\right\| \le \left\|-\boldsymbol{I}_{hs}^{0}\right\|\left\|\boldsymbol{H}_{q}^{H}\right\|\left\|\dot{\boldsymbol{q}}^{\mathrm{T}}\right\|\|\dot{\boldsymbol{q}}\| \\ \left\|\boldsymbol{F}_{H2}^{i}\right\| = \left\|-\boldsymbol{I}_{hs}^{0}\boldsymbol{G}_{q}^{H}\ddot{\boldsymbol{q}}\right\| \le \left\|-\boldsymbol{I}_{hs}^{0}\right\|\left\|\boldsymbol{G}_{q}^{H}\right\|\|\ddot{\boldsymbol{q}}\| \\ \left\|\boldsymbol{F}_{H3}^{i}\right\| = \left\|-[\boldsymbol{\omega}]\boldsymbol{I}_{h}\boldsymbol{\omega}_{h}\right\| \le \sqrt{2}(\|\boldsymbol{G}_{1}\|+\|\boldsymbol{G}_{2}\|+\|\boldsymbol{G}_{3}\|)\|\boldsymbol{I}_{h}\|\left\|\boldsymbol{G}_{q}^{H}\right\|\|\dot{\boldsymbol{q}}\|^{2} \end{cases}, \tag{7}$$

and

$$\begin{aligned} \left\|\boldsymbol{F}_{H}^{i}\right\| &\le \left\|-\boldsymbol{I}_{hs}^{0}\right\|(\left\|\boldsymbol{G}_{q}^{H}\right\|\|\ddot{\boldsymbol{q}}\|+\left\|\boldsymbol{H}_{q}^{H}\right\|\|\dot{\boldsymbol{q}}\|\left\|\dot{\boldsymbol{q}}^{T}\right\|)+\|\boldsymbol{\omega}\|\|\boldsymbol{I}_{h}\|\|\boldsymbol{\omega}_{h}\| \\ &\le \left\|\boldsymbol{I}_{hs}^{0}\right\|(\left\|\boldsymbol{G}_{q}^{H}\right\|\|\ddot{\boldsymbol{q}}\|+\left\|\boldsymbol{H}_{q}^{H}\right\|\|\dot{\boldsymbol{q}}\|^{2})+\sqrt{2}\sum_{j=1}^{3}\left\|\boldsymbol{G}_{j}\right\|\|\boldsymbol{I}_{h}\|\left\|\boldsymbol{G}_{q}^{H}\right\|\|\dot{\boldsymbol{q}}\|^{2} \end{aligned}. \tag{8}$$

Global performance index of inertial force $\boldsymbol{\tau}_{I}$ used to analyze the effect of size and mass on dynamic performance of parallel mechanism can be proposed as

$$\boldsymbol{\tau}_{I} = \frac{\int_{W}\frac{1}{r_{I}}\mathrm{d}W}{\int_{W}\mathrm{d}W}. \tag{9}$$

Where $\boldsymbol{r}_{I} = \left\|-[\boldsymbol{I}_{hs}^{0}]\right\|(\left\|\boldsymbol{G}_{q}^{H}\right\|+\left\|\boldsymbol{H}_{H}^{q}\right\|+\sqrt{2}(\|\boldsymbol{G}_{1}\|+\|\boldsymbol{G}_{2}\|+\|\boldsymbol{G}_{3}\|)\|\boldsymbol{I}_{h}\|\left\|\boldsymbol{G}_{q}^{H}\right\|$, $\boldsymbol{I}_{h} = \begin{pmatrix}\boldsymbol{I}_{hc}^{0} & 0\end{pmatrix}^{T}$, $\boldsymbol{I}_{hc}^{0}$ is inertia tensor, $\boldsymbol{I}_{hs}^{0} = \begin{pmatrix}\boldsymbol{I}_{hc}^{0} & 0 \\ 0 & [\boldsymbol{M}]\end{pmatrix}$, $\boldsymbol{I}_{hs}^{0}$ is general mass matrix.

3 Introduction of 2-RUUS Mechanism

As shown in Fig. 1, platforms 1 and 2 are fixed on the base, and a fixed frame O-XYZ denoted as {B} is created on platform 1. Moving platform III is a complex octagon, which can rotate about axes X and Y respectively, center of point P and diameter of HG and DK. DK is perpendicular to HG, denoted as DK⊥HG. Besides, there are $PA_1 \perp PA_2$, $OC_1 \perp OC_2$, and PO is perpendicular to platform 1. There are four universal joints connecting with platform 1 at points D, G, K and H, which is used to balance weight of mechanism and two horizontal moving joints at points E and F used to constraint moving platform rotating about axis Z. The kinematic joints A_1, A_2, B_1 and

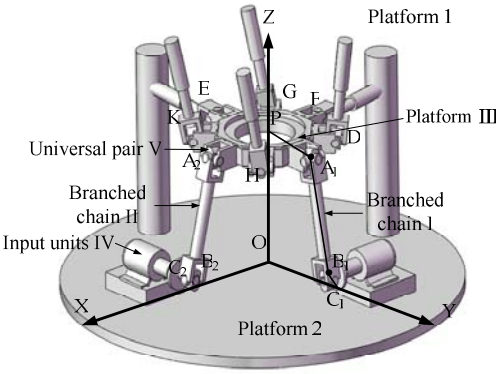


Fig. 1. 2-RUUS mechanism

B_2 are universal joints, bar A_1B_1 is connected with electromotor through eccentric tray. Electromotor is the driver, depicted as revolute joint R. Considering the spherical bearing in the middle of platform III (it is not sketched in Fig. 1), leg I can be expressed as RUUS branched-chain. Leg II is similar to leg I, thus 2-RUUS mechanism is got.

Inverse solution is used to analyze the connection between input and output. Revolute joints C_1 and C_2 are two inputs, and point P is the output point. Two electromotor are two inputs chosen to analyze independently. The coordinate of point P is (0,0,H), H is the height between points P and O, α and β are angles that platform III rotate about axes X and Y respectively. At initiate configuration, $PA_1 \perp PO$, point B_1 is on line OC_1. Lengths of PA_1, A_1B_1 and B_1C_1 are R, L and r respectively. r is the radius of eccentric tray. From Fig. 1, coordinate of points in leg I are $C_1 = [0, R+\sqrt{L^2-H^2}+r, 0]^T$, $A_1 = [0, R\cos\alpha, R\sin\alpha + H]^T$ and $B_1 = [0, B_{1y}, B_{1z}]^T$, where B_{1y} and B_{1z} meet the conditions of $|A_1B_1|=L$ and $|B_1C_1|=r$. If set $s_{11} = (L^2-H^2)^{1/2} + r + R - A_{1y}$ and $s_{21} = r^2 + s_{11}^2 + A_{1z}^2 - L^2$,

$$B_{1z} = \frac{s_{21}A_{1z} - \sqrt{s_{21}^2 A_{1z}^2 - 4(A_{1z}^2 + s_{11}^2)(s_{21}^2 - 4s_{11}^2 r^2)}}{4A_{1z}^2 + 4s_{11}^2}, \quad B_{1y} = \sqrt{L^2 - H^2} + R + r - \sqrt{r^2 - B_{1z}^2}$$

are got. The coordinates of every joint in the leg II can get in the same way.

4 Inertia Tensor Analysis of Mechanism

As shown in Fig. 2 and Fig. 3, 2-RUUS mechanism is simplified into five major components, including branch I, branch II, Platform III, Input units IV and Universal joint V.

Branch I can be divided into three parts, two are composed of 1,2,3 and 4, another composed of 5, shown in Fig. 2 (only show half of the branch). A frame $O_IX_IY_IZ_I$ is

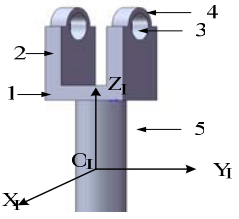


Fig. 2. Sketch of sub-component I

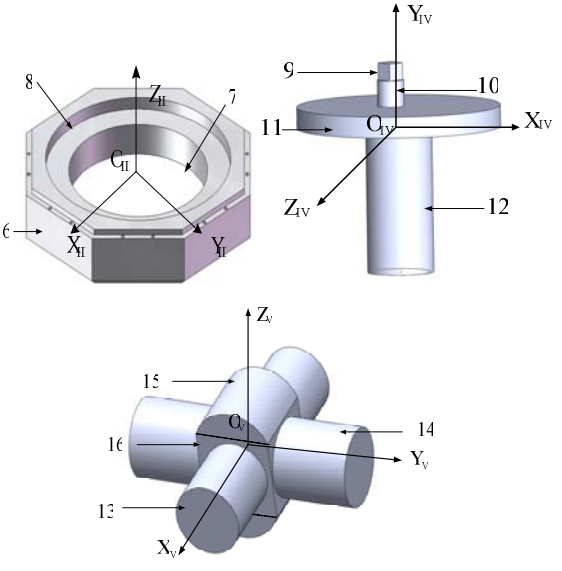


Fig. 3. Sketch of Platform III, Input units IV and Universal joint V. Platform III is on the left, Input units IV is in the middle, Universal joint V is on the right.

created on it, its origin is mass center of component 5. 1 is a cuboid, length of a_1, width of b_1, height of c_1, mass of m_1. 2 is a cuboid, length of a_2, width of b_2, height of c_2, mass of m_2. 3 is a hollow cylinder, radius of R_3, mass of m_3, and the distance between the center of circle section and the bottom surface of 1 is h_3. 4 is a semicircular column, radius of R_4, mass of m_4, and the distance between the center of circle section and the bottom surface of 1 is h_4. 5 is a cylinder, radius of R_5, height of h_5, mass of m_5.

Based on engineering mechanics, moments of inertia of branch I are got as follows.

$$I_{x_I x_I} = 2\left\{ m_1 \frac{b_1^2 + c_1^2}{12} + m_1 \frac{b_1^2}{4} + m_2 \frac{b_2^2 + c_2^2}{12} + m_2 (\frac{c_2}{2} + b_1)^2 + m_4 [\frac{R_4^2}{2}(1 - \frac{64}{9\pi^2}) + \frac{b_2^2}{12}] + m_4 h_4^2 - \right.$$
$$\left. m_3 [\frac{3R_3^2 + b_2^2}{12} + h_3^2 + (\frac{c_1 - b_2}{2})^2] + m_2 (\frac{c_1 - b_2}{2})^2 \frac{(m_1 + m_2 + m_3 + m_4) L^2}{4} \right\} + m_5 \frac{3R_5^2 + h_5^2}{12}$$

$$I_{y_I y_I} = 2\left\{ m_1 \frac{b_1^2 + a_1^2}{12} + m_1 \frac{b_1^2}{4} + m_2 \frac{b_2^2 + a_2^2}{12} + m_2 (\frac{c_2}{2} + b_1)^2 + \frac{m_4 R_4^2}{2} (1 - \frac{32}{9\pi^2}) + m_4 h_4^2 + \right.$$
$$\left. \frac{(m_1 + m_2 + m_3 + m_4) L^2}{4} - m_3 (\frac{R_3^2}{2} + h_3^2) \right\} + m_5 \frac{3R_5^2 + h_5^2}{12}$$

$$I_{z_I z_I} = 2\left\{ m_1 \frac{a_1^2 + c_1^2}{12} + m_2 \frac{a_2^2 + c_2^2}{12} + m_2 (\frac{c_1 - b_2}{2})^2 - m_3 [\frac{3R_3^2 + b_2^2}{12} + (\frac{c_1 - b_2}{2})^2] + m_4 \frac{3R_4^2 + b_2^2}{12} + \right.$$
$$\left. m_4 (\frac{c_1 - b_2}{2})^2 \right\} + \frac{m_5 R_5^2}{2}$$

and $I_{x_I y_I} = I_{x_I z_I} = I_{y_I z_I} = 0$. So, inertia tensor of branch I at any configuration is $\boldsymbol{I}_I' = \boldsymbol{T}_{1r}^T \boldsymbol{I}_I \boldsymbol{T}_{1r}$, where $\boldsymbol{T}_{1r}$ is transformation matrix between $O_I X_I Y_I Z_I$ and {B} and $I_I = diag(I_{xxI}, I_{yyI}, I_{zzI})$. Similarly, inertia tensor of branch II is $\boldsymbol{I}_{II}' = \boldsymbol{T}_{2r}^T \boldsymbol{I}_{II} \boldsymbol{T}_{2r}$.

Inertia tensor of branch II, Platform III, Input unit IV and Universal joint V are, $\boldsymbol{I}_{III}' = \boldsymbol{T}_{3r}^T \boldsymbol{I}_{III} \boldsymbol{T}_{3r}$, $\boldsymbol{I}_{IV_1}' = \boldsymbol{T}_{4r}^T \boldsymbol{I}_{IV} \boldsymbol{T}_{4r}$, and $\boldsymbol{I}_{V_1}' = \boldsymbol{T}_{5r}^T \boldsymbol{I}_V \boldsymbol{T}_{5r}$ respectively.

In sum up, inertia tensor matrix and generalized mass matrix are

$$[\boldsymbol{I}_{hc}^0] = \boldsymbol{I}_I' + \boldsymbol{I}_{II}' + \boldsymbol{I}_{III}' + \boldsymbol{I}_{IV1}' + \boldsymbol{I}_{IV2}' + \sum_{i=1}^{4} \boldsymbol{I}_{Vi}' \tag{10}$$

$$\boldsymbol{I}_{hs}^0 = \begin{pmatrix} [\boldsymbol{I}_{hc}^0] & 0 \\ 0 & [\boldsymbol{M}] \end{pmatrix}. \tag{11}$$

Where $[\boldsymbol{M}]$ is diagonal matrix of M. M is mass of the mechanism.

5 Inertia Force Performance Analysis of Mechanism

According to global conditioning index defined in reference [9], acceleration global performance indices is analyzed. Set *R*=10~80 cm (step is 5cm), *L*=95~210cm (step is 8cm), *r*=5~11cm. acceleration global performance indices are got and shown in Fig. 4.

From Fig. 4, acceleration performance of mechanism is much higher when the value of *L* is bigger. Acceleration performance of mechanism is much lower when the value of *r* is bigger and R is smaller. Acceleration performance index increase gradually until r=9cm from R=10cm to 30cm. From R=10cm to R=20 cm, acceleration performance index become much bigger too when r becoming bigger. That is to say, r=9 is an inflexion of acceleration performance index of mechanism.

According to equation (9), inertia force performance is analyzed. Set *R*=10 ~80 cm, *L*=95~210 cm and *r*=9 cm, atlas of inertia force performance is go and showed in

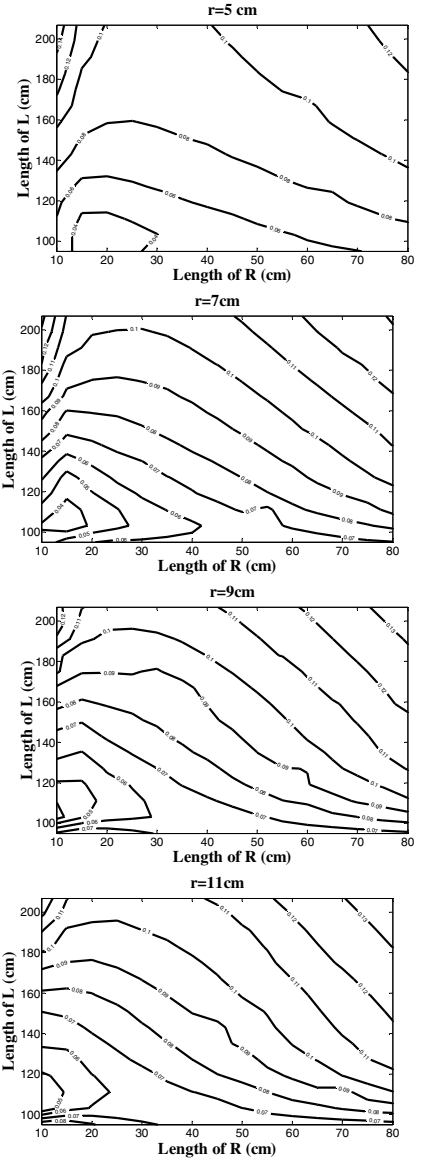


Fig. 4. Acceleration performance atlas of mechanism

Fig. 5. From Fig. 5, inertia force performance of mechanism is higher when L=185~207 cm and R=65~80 cm. That is to say, inertia force performance is higher when the value of L and R are bigger. The results coincide with the results of acceleration performance analysis. So, it can be draw a conclusion that it will improve dynamic performance of 2-RUUS mechanism by increasing the value of L and R.

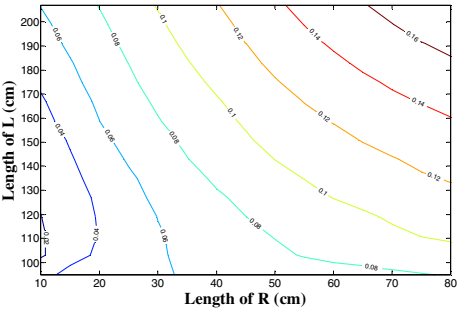


Fig. 5. Atlas of inertia force performance

Based on the above analysis, the range of reasonable size is got. In order to validate the defined performance index, a simulation model is created. By analyzing inertia force performance of different parameters, we can choose better sizes. At first, we fix on the range of reasonable size among the size that inertia force performance is better. Then, we analyze the relationship between inertia force and the radius of leg A_iB_i (30~36cm) when the thickness of the platform is 10cm and 10.8cm in the condition of L=207 cm and R=80 cm, the result is shown in Fig. 6.

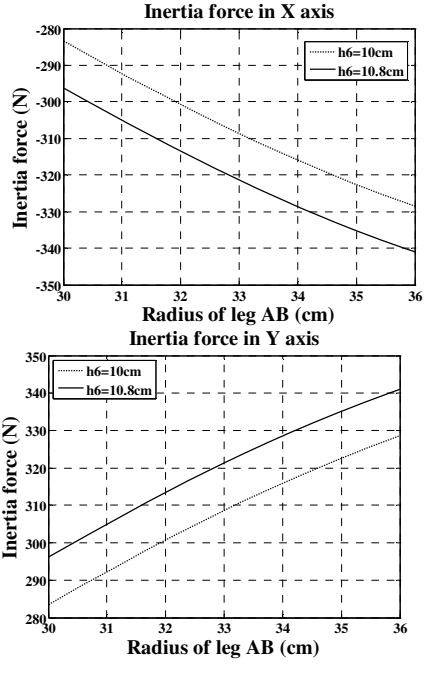


Fig. 6. Contrast of inertia force in axes X and Y

From Fig. 6, we can know

(1) The inertia force is increasing with the radius of leg A_iB_i (i=1, 2), which coincides with analysis presented above.

(2) From two former figures, inertia force at 10.8cm is bigger than at 10cm in the same radius of leg A_iB_i. So inertia force performance of the platform is better when the thickness of platform is 10.8cm.

6 Conclusions

(1) Inequality relationship between inertia force, size and mass of mechanism is given. From the inequality, global inertia force performance index of mechanism is defined. The index can be used to analyze the effect of size and mass on dynamic performance for any parallel mechanism.

(2) As an example, inertia force performance of 2-RUUS mechanism is analyzed, and atlases of these indices are also given. The better sizes of mechanism can be found from the atlases.

(3) Inertia force of mechanism is analyzed by simulation model. The results are consistent with the results of theoretical analysis. The correctness of theoretical analysis and the feasibility of the defined index are verified, which lays a theoretical foundation to optimization design of parallel mechanism.

References

1. Salisbury, J.K., Craig, J.J.: Articulated Hands Force Control and Kinematics Issues. Int. J. of Robot. Res. 1(1), 4–17 (1982)
2. Asada, H.: A geometrical Representation of Manipulator Dynamics and its Application to Arm Design. Trans. of ASME J. of Dynamic Systems, Measurement and Control 105(3), 131–142 (1983)
3. Yoshikawa, T.: Dynamic Manipulability of Robot Manipulators. J. Robot. Syst. 2(1), 113–124 (1985)
4. Yoshikawa, T.: Manipulability of Robotic Mechanisms. Int. J. of Robot. Res. 4(2), 3–9 (1985)
5. Angeles, J., López-Cajún, C.: The Dexterity Index of Serial-type Robotic Manipulator. ASME Trends and Development in Mechanism, Machines and Robotics, 79–84 (1988)
6. Gosselin, C., Angeles, J.: The Optimum Kinematics Design of a Spherical Three-degree-of-freedom Parallel Manipulator. Trans. ASME J. Mechanism Transmissions Automation Des. 111(2), 202–207 (1989)
7. Gosselin, C., Angeles, J.: A Global Performance Index for the Kinematics Optimization of Robotic Manipulator. Transactions of the ASME 113(1), 220–226 (1991)
8. Liu, X.J.: The Relationships between Performance Criteria and Link Lengths of Parallel Manipulators and Their Design Theory (in Chinese). PhD thesis, Yanshan University, China (1999)
9. Guo, X.J., Liu, S.: Analysis for Dynamics Performance Indices of 4-RR(RR)R Parallel Mechanism. Int. J. of Innovative Computing Information and Control 2(4), 849–862 (2006)
10. Huang, Z., Zhao, Y.S., Zhao, T.S.: Advanced spatial mechanism (in Chinese). Higher Education Press, Beijing (2006)

Reordering and Partitioning Jacobian Matrices Using Graph-Spectral Method

Xuelin Wang and Yujin Hu

School of Mechanical Science and Engineering,
Huazhong University of Science and Technology, Wuhan 430074, P.R. China
wangxl@mail.hust.edu.cn, yjhu@public.wh.hb.cn

Abstract. An efficient spectral method is developed for reducing the Jacobian matrix to its block-triangular form in order to solve the inverse kinematics problem. Based on the kinematic structure matrix, the problem of reducing the Jacobian matrix to block-triangular form is transformed into reducing the bandwidth of the matrix. The second Laplacian eigenvector, associated with the bigraph of the structure matrix of the inverse Jacobian, is used to renumber the rows and columns of the Jacobian. The rearranged Jacobian can be divided into subsystems immediately according to the entry value of the Fiedler vector. This algorithm is applied in detail to kinematic analysis for a PUMA robot and T^3 robot. Because of the algebraic nature of spectral algorithm, the algorithm can be implemented in a fairly straightforward manner.

Keywords: Inverse kinematics, Graph theory, Laplacian matrix, Robot analysis.

1 Introduction

The Jacobian matrix defines the transformation between the robot's hand velocity and the joint velocity. The relationship that the Jacobian creates can be expressed as:

$$\dot{\mathbf{x}} = \mathbf{J}(\mathbf{q})\dot{\mathbf{q}} \tag{1}$$

where vector $\mathbf{q}$ comprises the joint variables. Vector $\dot{\mathbf{q}}$ of joint velocities, the time derivative of vector $\mathbf{q}$, is related to vector $\dot{\mathbf{x}}$ of end-effector velocities by Jacobian matrix $\mathbf{J}(\mathbf{q})$.

The Jacobian is a function of the configuration $\mathbf{q}$. The inverse of the Jacobian, $\mathbf{J}^{-1}$, may be used to solve the inverse kinematic problem even when only positions and orientations are specified. Many methods of solution for this problem have been discussed over the past 30 years [1-7]. As computational speeds have improved, there has been greater expectation that this calculation will be performed online. However, inverting the Jacobian is not always an easy task, particularly when this inversion has to be done at an online basis [1,3].

Determining if a robot may be in a singular configuration during its motion is a problem that is of high practical interest. Based on the fact that the robot kinematic singularities are not equally important to the behaviour of the manipulator, Martins

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 696–705, 2009.

and Guenther [9] developed a method using digraph techniques, their method is based entirely on the structure of the Jacobian and on the reduction of the Jacobian matrix to its finest block-triangular form. This form, called hierarchical canonical form, improves the detection of the singularities. However, the block-triangular form is derived from the strong components of the associated digraph, this approach require some expertise of directed graph. Otherwise, an important point is if such block-triangular form is unique when different strategies are employed.

In this paper a spectral algorithm is presented for solving the ordering and partitioning problem of the Jacobian. The given sparse Jacobian matrix is associated to a bi-graph graph, a particular eigenvector of the Laplacian matrix of the graph is computed and its components is used to renumber and partition the vertices of the graph, and applying the order of vertices to rearrange the system (1), the hierarchically rearranged linear system can be obtained.

2 Spectral Methods

A graph G =(V,E) is defined as a set of vertices or nodes V and a set of edges or members E together with a relation of incidence which associates two vertices with each edge.

Let G=(V,E) be a graph with the vertex set V, containing N vertices, and the edge set E. The adjacency matrix $\mathbf{A}=[a_{ij}]_{N\times N}$ of the labeled graph G is defined as

$$a_{ij}=\begin{cases}+1 & \text{if } n_i \text{ and } n_j \text{ are adjacent}\\ 0 & \text{otherwise}\end{cases}$$

The degree matrix $\mathbf{D}=[d_{ij}]_{N\times N}$ is a diagonal matrix, where d_{ii} is equal to the valency of the *i*th vertex.

The Laplacian matrix of a graph is defined as $\mathbf{L}=[l_{ij}]_{N\times N}=\mathbf{D}-\mathbf{A}$. Therefore, the entries of **L** are as

$$l_{ij}=\begin{cases}-1 & \text{if } n_i \text{ and } n_j \text{ are adjacent}\\ \deg(n_i) & \text{if } i=j\\ 0 & \text{otherwise}\end{cases}$$

Consider the following eigensolution problem:

$$\mathbf{L}\boldsymbol{v}=\lambda\boldsymbol{v} \tag{2}$$

where λ n $\boldsymbol{v}$ are the eigenvalue and eigenvector, respectively. **L** is a semi-positive definite matrix, let the eigenvalues of **L** be ordered $0=\lambda_1\le\lambda_2\cdots\cdots\le\lambda_N$.The multiplicity of the zero eigenvalue is equal to the number of connected components of the graph. If G is connected, then the second smallest eigenvalue $\lambda_2>0$. We call an eigenvector $\boldsymbol{v}_2$ corresponding to λ_2 a second eigenvector.

The second eigenvalue λ_2, and the corresponding eigenvector $\boldsymbol{v}_2$ have attractive properties. Fiedler[10] has investigated various properties of λ_2. This eigenvalue is

known as algebraic connectivity of a graph, ν_2 is also known as the Fiedler vector. Several methods exist to find a partition of a graph which minimizes the number of edges cut by the Fielder vector [8]. The Fiedler vector also was used for nodal numbering to reduce the bandwidth of the matrix in finite element analysis [13,14].

3 Jacobian Ordering

In order to obtain the block-triangular form of a generally sparse Jacobian matrix more efficiently, a kinematic structure matrix $\mathbf{Q}$ was defined as [9]:

$$\mathbf{Q} = \begin{bmatrix} \mathbf{0} & \mathbf{J} \\ \mathbf{J}^{-1} & \mathbf{0} \end{bmatrix}$$

Considering the Eq.(1), the below relation can be got [9]:

$$\begin{bmatrix} \dot{\mathbf{x}} \\ \dot{\mathbf{q}} \end{bmatrix} = \mathbf{Q} \begin{bmatrix} \dot{\mathbf{x}} \\ \dot{\mathbf{q}} \end{bmatrix} \tag{3}$$

Eq.(3) can also associate with a bipartite graph, the bigraph consists of two distinct sets X and Q, and undirected edges which join the vertices in X to those in Q. The set X is associated with the rows of the matrix, and the set Q with column. The adjacency matrix $\mathbf{A}_Q$ and Laplacian matrix L_Q of the bipartite graph are respectively□

$$\mathbf{A}_Q = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \overline{\mathbf{J}} \\ \mathbf{0} & \mathbf{0} & \overline{\mathbf{J}}_{-1} & \mathbf{0} \\ \mathbf{0} & \overline{\mathbf{J}}_{-1}^{\mathrm{T}} & \mathbf{0} & \mathbf{0} \\ \overline{\mathbf{J}}^{\mathrm{T}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \overline{\mathbf{Q}} \\ \overline{\mathbf{Q}}^{\mathrm{T}} & \mathbf{0} \end{bmatrix}$$

$$\mathbf{L}_Q = \begin{bmatrix} \mathbf{D}_1 & \mathbf{0} & \mathbf{0} & -\overline{\mathbf{J}} \\ \mathbf{0} & \mathbf{D}_3 & -\overline{\mathbf{J}}_{-1} & \mathbf{0} \\ \mathbf{0} & -\overline{\mathbf{J}}_{-1}^{\mathrm{T}} & \mathbf{D}_4 & \mathbf{0} \\ -\overline{\mathbf{J}}^{\mathrm{T}} & \mathbf{0} & \mathbf{0} & \mathbf{D}_2 \end{bmatrix}$$

where $\overline{\mathbf{J}}$ and $\overline{\mathbf{J}}_{-1}$ are structure matrices of the Jacobian and inverse Jacobian matrices, respectively. $\overline{\mathbf{Q}}$ is the structure matrix of $\mathbf{Q}$. The structure matrix is the relative position in the matrix of the null and non-null elements. The structure matrix can be represented by a Boolean matrix simply by changing the non-null elements by the number 1 and keeping the null elements as 0.

In matrix $\mathbf{L}_Q$, the $\mathbf{D}_1$, $\mathbf{D}_2$, $\mathbf{D}_3$ and $\mathbf{D}_4$ are diagonal matrix, the diagonal element of $\mathbf{D}_1$ and $\mathbf{D}_2$ is the numbers of non-null elements in each row and column of $\overline{\mathbf{J}}$ respectively, the diagonal element of $\mathbf{D}_3$ and $\mathbf{D}_4$ is the numbers of non-null elements in each row and column of $\overline{\mathbf{J}}_{-1}$ respectively.

While $\mathbf{L}_Q$ requires the *prior* knowledge of the structure matrix of the inverse Jacobian $\overline{\mathbf{J}}_{-1}$, the knowledge of the exact values of $\mathbf{J}^{-1}$ are not necessary, therefore, the structure matrix $\overline{\mathbf{J}}_{-1}$ can be obtained without the symbolic inversion of $\mathbf{J}$. There are several methods to obtain $\overline{\mathbf{J}}_{-1}$ either using the graphical interpretation of the determinant or, numerically, by inverting a series of matrices whose incidency matrix is $\overline{\mathbf{J}}$ [9].

Permuting row 1 and row 3, column 1 with 3 in $\mathbf{L}_Q$, we can get

$$\tilde{\boldsymbol{L}}_Q = \begin{bmatrix} \mathbf{L}'_{J_{-1}} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_J \end{bmatrix}$$

where $\mathbf{L}_J = \begin{bmatrix} \mathbf{D}_1 & -\overline{\mathbf{J}} \\ -\overline{\mathbf{J}}^{\mathrm{T}} & \mathbf{D}_2 \end{bmatrix}$ and $\mathbf{L}'_{J_{-1}} = \begin{bmatrix} \mathbf{D}_4 & -\overline{\mathbf{J}}^{\mathrm{T}}_{-1} \\ -\overline{\mathbf{J}}_{-1} & \mathbf{D}_3 \end{bmatrix}$.

$\mathbf{L}_J$ is a Laplacian matrix associated with matrix $\mathbf{J}$, $\mathbf{L}'_{J_{-1}}$ is a Laplacian matrix related to $\mathbf{J}^{-1}$ which permuted rows or columns. According to the property of eigenproblems, the eigenvalues of $\mathbf{L}_Q$ are identical with $\tilde{\mathbf{L}}_Q$. On the other hand, because $\tilde{\mathbf{L}}_Q$ is diagonal block matrix, the spectra of $\tilde{\mathbf{L}}_Q$ can be computed from the spectrum of the diagonal blocks $\mathbf{L}_J$ and $\mathbf{L}'_{J_{-1}}$.

If the adjacency matrix $\mathbf{A}_Q$ or Laplacian matrix $\mathbf{L}_Q$ has the minimum bandwidth, the matrix $\overline{\mathbf{Q}}$ should have lower block-triangular form. Therefore, the problem of reducing the Jacobian matrix to block-triangular form is transformed into reorder the vertices of a bipartite graph to obtain the minimum bandwidth of its adjacency matrix. The algorithm to reduce the bandwidth of matrix can be employed to obtain the block-triangular form of matrix $\boldsymbol{Q}$.

Theory exists to show that spectral partition methods work well on the reduction of matrix bandwidth [8][12]. Substituting for solving the Fiedler vector of matrix $\mathbf{L}_Q$, we can compute the Fiedler vector of $\tilde{\mathbf{L}}_Q$ more easily. In this paper, a simplified scheme is adopted to rearrange the Jacobian. We use only the entries of the second eigenvectors of Laplacian matrix $\mathbf{L}'_{J_{-1}}$, which is one of diagonal blocks of $\tilde{\mathbf{L}}_Q$.

Let $\mathbf{v}_2$ denote the second eigenvector of Laplacian matrix $\mathbf{L}'_{J_{-1}}$, and assumed the largest components of $\boldsymbol{v}_2$ is positive.

The components of $\boldsymbol{v}_2$ are used to renumber and partition the vertices of the graph associated with Jacobian. The following algorithm describes how to obtain the block-triangular form of the Jacobian matrix.

(1) Find $\boldsymbol{v}_2 = [v_2^{(1)}, v_2^{(2)}, \cdots v_2^{(i)}, \cdots v_2^{(N)}]^T$, the second eigenvectors of Laplacian matrix $\mathbf{L}'_{J_{-1}}$.

(2) Order the entries $\boldsymbol{v}_2$ in a descending order.

(3) Reorder the vertices according to their occurrence in $\boldsymbol{v}_2$ for set X and set Q.

(4) Mapping the vertex ordering of the graph into row and column ordering of the matrix, and the set X is associated with the rows of the matrix, and the set Q with columns.

(5) If the entries with the same value in $\boldsymbol{v}_2$, i.e. $v_2^{(i)} = v_2^{(i+1)} = \cdots v_2^{(i+n)}$, the vertices related to the entries are arranged as a sub-block. Then, the vertices in the same sub-block are reordered according to the finest block-triangular form.

(6) Divide Jacobian matrix **J** into subsystems according to the sub-block in set X.

The Jacobian matrix is sparse and its sparsity is crucial to all methods which represent matrices by graphs[8]. Once the row and column are reordered, the canonical form of the Jacobian matrix can be obtained.

4 Computational Results

4.1 PUMA Robot

PUMA robots are one of the most studied configurations found in research papers on robotics[5]. Fig. 1 shows the PUMA robot, the robot is a serial manipulator with six degrees of freedom, all joints are rotative kinematic pairs. Position angles θ_i at the rotative joints are shown in Fig.1. Screws $\$i$ are aligned with the joint axes and drawn as conical arrows. The chosen coordinate system is represented by the triad x, y, z.

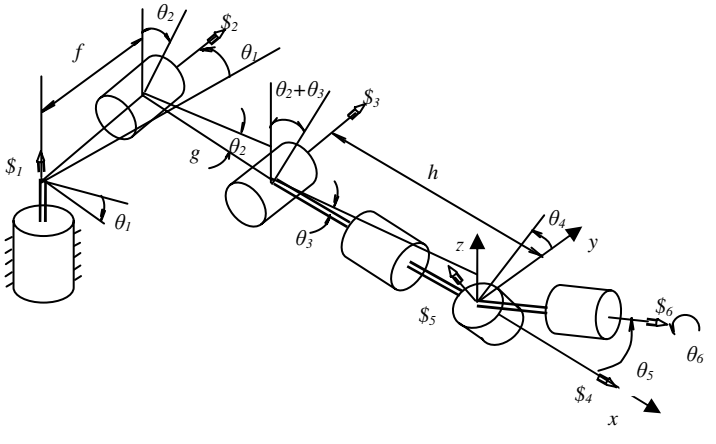


Fig. 1. A PUMA-type arm in a general configuration

The Jacobian of a PUMA robot has been obtained in Ref.[5] using screw theory. For an example of application of the algorithm to PUMA robot, with the coordinate system sketched in Fig.1, the Jacobian matrix is written as Eq.(4), and its corresponding bipartite graph is shown in Fig.2.

$$\mathbf{J}=\begin{bmatrix} -s_{23} & 0 & 0 & 1 & 0 & c_5 \\ 0 & 1 & 1 & 0 & -s_4 & c_4 s_5 \\ c_{23} & 0 & 0 & 0 & c_4 & s_4 s_5 \\ -fc_{23} & gs_3 & 0 & 0 & 0 & 0 \\ x_{14} & 0 & 0 & 0 & 0 & 0 \\ -fs_{23} & x'_{41} & -h & 0 & 0 & 0 \end{bmatrix} \tag{4}$$

In which $s_i=\sin\theta_i$; $s_{ij}=\sin(\theta_i+\theta_j)$; $c_i=\cos\theta_i$; $c_{ij}=\cos(\theta_i+\theta_j)$; and $x_{14}=gc_2+hc_{23}$; $x'_{41}=-(gc_3+h)$, *f, g, h* are the distances shown in Fig.1.

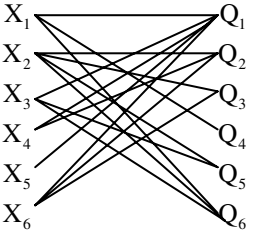


Fig. 2. The bipartite graph corresponding to PUMA

The structure matrix of the Jacobian and the structure of the inverse Jacobian are respectively:

$$\bar{\mathbf{J}}=\begin{bmatrix} 1&0&0&1&0&1 \\ 0&1&1&0&1&1 \\ 1&0&0&0&1&1 \\ 1&1&0&0&0&0 \\ 1&0&0&0&0&0 \\ 1&1&1&0&0&0 \end{bmatrix} \qquad \bar{\mathbf{J}}_{-1}=\begin{bmatrix} 0&0&0&0&1&0 \\ 0&0&0&1&1&0 \\ 0&0&0&1&1&1 \\ 1&1&1&1&1&1 \\ 0&0&0&1&1&1 \\ 0&1&1&1&1&0 \end{bmatrix}$$

We use the proposed algorithm to rearrange the system, where the end-effector screw $\dot{\mathbf{x}}=\$_6=[x_1, x_2, x_3, x_4, x_5, x_6]^{\mathrm{T}}$ is represented in the coordinate system shown in Fig. 1, see details in Ref.[5].

Using the present algorithm, we have:

(1) Calculate the Fiedler vector $\boldsymbol{v}_2$ of Laplacian matrix $\mathbf{L}'_{J_{-1}}$, the eigenvector is obtained as $\boldsymbol{v}_2$ = [0.6592, 0.1552, 0.1552, -0.0292, -0.1497, -0.0082, -0.6592, -0.1458, -0.0840, 0.1497, -0.0840, 0.0408]$^{\mathrm{T}}$.

(2) Arranging the entries of $\boldsymbol{v}_2$ in a descending order and renumbering the vertices according to their occurrence in $\boldsymbol{v}_2$. The rearranged order of vertices can be obtained

according to $\boldsymbol{v}_2$, for set X, the rearranged order is: X_1, X_2, X_3, X_6, X_4, X_5 ,and for set Q, the rearranged order is: Q_4, Q_6, Q_3, Q_5, Q_2, Q_1.

(3) Because $v_2^{(2)} = v_2^{(3)}$ =0.1552, the order of X_2 and X_3 should be determined further by the block triangular form. Repeating the above process to Q_5 and Q_3□Q_5 should be in front of Q_3. As a result, the rearranged order corresponding to $\dot{\mathbf{x}}$ and $\dot{\mathbf{q}}$ is: $\dot{x}_1, \dot{x}_3, \dot{x}_2, \dot{x}_6, \dot{x}_4, \dot{x}_5$ and $\dot{q}_4, \dot{q}_6, \dot{q}_5, \dot{q}_3, \dot{q}_2, \dot{q}_1$.

The rearranged from of the Jacobian of Eq.(4) is:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_3 \\ \dot{x}_2 \\ \dot{x}_6 \\ \dot{x}_4 \\ \dot{x}_5 \end{bmatrix} = \begin{bmatrix} 1 & c_5 & 0 & 0 & 0 & -s_{23} \\ 0 & s_4 s_5 & c_4 & 0 & 0 & c_{23} \\ 0 & c_4 s_5 & -s_4 & 1 & 1 & 0 \\ 0 & 0 & 0 & -h & x'_{41} & -fs_{23} \\ 0 & 0 & 0 & 0 & gs_3 & -fc_{23} \\ 0 & 0 & 0 & 0 & 0 & x_{14} \end{bmatrix} \begin{bmatrix} \dot{q}_4 \\ \dot{q}_6 \\ \dot{q}_5 \\ \dot{q}_3 \\ \dot{q}_2 \\ \dot{q}_1 \end{bmatrix} = \hat{\mathbf{J}} \begin{bmatrix} \dot{q}_4 \\ \dot{q}_6 \\ \dot{q}_5 \\ \dot{q}_3 \\ \dot{q}_2 \\ \dot{q}_1 \end{bmatrix} \tag{5}$$

Matrix $\hat{\mathbf{J}}$ is the hierarchical canonical form of the Jacobian given by Eq.(4). Comparing with Martins and Guenther's method, the difference between Martins' result and the above result is in the sequence of the row 2 and row3.

As mentioned above, the second eigenvector of Laplacian matrix can be used to induce a block partitioning of the canonical from of the Jacobian matrix. Following the procedures described in Section 3, the matrix $\hat{\mathbf{J}}$ can be divided into subsystems immediately according to the value of entries of $\boldsymbol{v}_2$ in set X, if the entries of $\boldsymbol{v}_2$ with the same value in set X, the vertices are arranged in the same sub-block. Thus the block partitioned matrix of $\hat{\mathbf{J}}$ is:

$$\hat{\mathbf{J}} = \left[\begin{array}{c|cc|c|c|c} 1 & c_5 & 0 & 0 & 0 & -s_{23} \\ \hline 0 & s_4 s_5 & c_4 & 0 & 0 & c_{23} \\ 0 & c_4 s_5 & -s_4 & 1 & 1 & 0 \\ \hline 0 & 0 & 0 & -h & x'_{41} & -fs_{23} \\ \hline 0 & 0 & 0 & 0 & gs_3 & -fc_{23} \\ \hline 0 & 0 & 0 & 0 & 0 & x_{14} \end{array}\right] \tag{6}$$

The matrix $\hat{\mathbf{J}}$ has five diagonal blocks whose determinants are x_{14} , gs_3, $-h$, s_5, 1. The terms $-h$ and 1 are invariant. Therefore if x_{14} =0, gs_3=0, s_5=0, these terms can lead to a singularity.

Once the system is subdivided and the subsystems are ordered, the solution of the subsystem can be easily obtained. For example, the subsystem comprised by the rows 2 and 3 of Eq. (5), we have:

$$\begin{bmatrix} \dot{x}_3 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} s_4 s_5 & c_4 \\ c_4 s_5 & -s_4 \end{bmatrix} \begin{bmatrix} \dot{q}_6 \\ \dot{q}_5 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \dot{q}_3 + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \dot{q}_2 + \begin{bmatrix} c_{23} \\ 0 \end{bmatrix} \dot{q}_1 \tag{7}$$

The variables $\dot{q}_5$ and $\dot{q}_6$ can be easily expressed as the inversion of the diagonal block:

$$\begin{bmatrix} \dot{q}_6 \\ \dot{q}_5 \end{bmatrix} = \begin{bmatrix} s_4 s_5 & c_4 \\ c_4 s_5 & -s_4 \end{bmatrix}^{-1} \left\{ \begin{bmatrix} \dot{x}_3 \\ \dot{x}_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \dot{q}_3 + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \dot{q}_2 + \begin{bmatrix} c_{23} \\ 0 \end{bmatrix} \dot{q}_1 \right\}$$

In the above process, the canonical block form describes how the variables affect one another.

4.2 T^3 Manipulator

The T^3 robot can be found in Refs.[5] and [9]. The Jacobian of this manipulator is extracted from Ref.[5]:

$$\mathbf{J} = \begin{bmatrix} -s_{234} & 0 & 0 & 0 & 0 & c_5 \\ 0 & 1 & 1 & 1 & 0 & s_5 \\ c_{234} & 0 & 0 & 0 & 1 & 0 \\ 0 & z_{24} & g s_4 & 0 & 0 & 0 \\ x_{23} & 0 & 0 & 0 & -h & 0 \\ 0 & -x_{34} & -g c_4 & 0 & 0 & h s_5 \end{bmatrix} \tag{8}$$

where $x_{23} = f c_3 + g c_{23}$; $x_{34} = f c_{34} + g c_4$; $z_{24} = f s_{34} + g s_4$.

In this example, the solving process is the same as the PUMA Jacobian. For T^3 manipulator we have:

(1) The Fiedler vector $\boldsymbol{v}_2$ of Laplacian matrix $\mathbf{L}'_{J_{-1}}$ is obtained as $\boldsymbol{v}_2 = \{$-0.0576, 0.9176, -0.1148, -0.0212, -0.1148, -0.0212, -0.2040, -0.0799, -0.0799, 0.1148, -0.2040, -0.1351$\}^{\mathrm{T}}$.

(2) Arranging the entries of $\boldsymbol{v}_2$ in a descending order and renumbering the vertices according to their occurrence in $\boldsymbol{v}_2$, the rearranged order of the vertices is obtained initially as X_2, X_4, X_6, X_1, X_3, X_5 and Q_4, Q_2, Q_3, Q_6, Q_1, Q_5.

(3) Because there is the entries with the same value in $\boldsymbol{v}_2$, reorder these groups with the identical entries, such as X_4 and X_6 , Q_1 and Q_5. The final order corresponding to $\dot{\mathbf{x}}$ and $\dot{\mathbf{q}}$ is obtained as: $\dot{x}_2, \dot{x}_6, \dot{x}_4, \dot{x}_1, \dot{x}_3\ \dot{x}_5$ and $\dot{q}_4, \dot{q}_2, \dot{q}_3, \dot{q}_6, \dot{q}_5, \dot{q}_1$.

The hierarchical canonical from of the Jacobian of Eq.(8) in a linear system form, becomes:

$$\begin{bmatrix} \dot{x}_2 \\ \dot{x}_6 \\ \dot{x}_4 \\ \dot{x}_1 \\ \dot{x}_3 \\ \dot{x}_5 \end{bmatrix} = \left[\begin{array}{c|cc|c|cc} 1 & 1 & 1 & s_5 & 0 & 0 \\ \hline 0 & -x_{34} & -g c_4 & h s_5 & 0 & 0 \\ 0 & z_{24} & g s_4 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & c_5 & 0 & -s_{234} \\ \hline 0 & 0 & 0 & 0 & 1 & c_{234} \\ 0 & 0 & 0 & 0 & -h & x_{23} \end{array}\right] \begin{bmatrix} \dot{q}_4 \\ \dot{q}_2 \\ \dot{q}_3 \\ \dot{q}_6 \\ \dot{q}_5 \\ \dot{q}_1 \end{bmatrix} \tag{9}$$

Compared with Martins' method [9], the above equation is different in the sequence of $\dot{x}_6$ and $\dot{x}_4$, $\dot{q}_5$ and $\dot{q}_1$.

Since the Jacobian matrix rearranged is a block upper triangular matrix, its determinant is easier to obtain:

$$\det\mathbf{J}=-fgs_3c_5(x_{23}+hc_{234})=-\ -fgs_3c_5(fc_2+gc_{23}+hc_{234})$$

The singularities of the Jacobian are: $x_{23}+hc_{234}=0$, $c_5=0$, $s_3=0$. The results indicate that the method described will provide a useful addition to existing techniques for the singularity analyses of robot.

5 Concluding Remarks

The problem of reducing the Jacobian matrix to block-triangular form is transformed into reorder the vertices of a bipartite graph. The Fiedler vector of Laplacian matrices corresponding to inverse Jacobian matrix are employed in reordering of the Jacobian matrix. The Jacobian matrix can be divided into subsystems immediately according to the entry value of the Fiedler vector. The proposed method is tested with practical examples, the results indicate that there are other alternative of sequence of variables compared with Martins' results [9].

The dominant computation in the present method is an eigenvector computation, because of its algebraic nature, the algorithm can be implemented in a fairly straight forward manner.

Acknowledgement. This work is supported by the NSF of China (Grant No. 50675077).

References

1. Zomaya, A.Y., Smith, H., Olariu, S.: Computing robot Jacobian on meshes with multiple buses. Microprocessors and Microsystems 23, 309–324 (1999)
2. Sciavicco, L., Siciliano, B.: Modeling and Control of Robot Manipulators. In: Electrical and Computer Engineering. McGraw-Hill, New York (1996)
3. Lucas, S.R., Tischler, C.R., Samuel, A.E.: Real-time solution of the inverse kinematic-rate problem. International Journal of Robotics Research 19, 1236–1244 (2000)
4. Nakamura, Y., Hanafusa, H.: Inverse kinematic solutions with singularity robustness for robot manipulator control. Journal of Dynamic Systems, Measurement and Control 108, 163–171 (1986)
5. Hunt, K.H.: Robot kinematics–a compact analytic inverse solution for velocities, Transactions of ASME. Journal of Mechanisms, Transmissions and Automation in Design 109, 42–49 (1987)
6. Wang, S.L., Waldron, K.J.: A study of the singular configurations of serial manipulators. Journal of Mechanisms, Transmissions, and Automation in Design 109, 14–20 (1987)
7. Waldron, K.J., Wang, S.L., Bolin, S.J.: A study of the Jacobian matrix of serial manipulators. Transactions of ASME, Journal of Mechanisms, Transmissions and Automation in Design 107, 230–238 (1985)

8. Holzrichter, M., Olieira, A.: A Graph based method for generating the Fiedler vector of Irregular problems. LNCS, pp. 978–985. Springer, London (1999)
9. Martins, D., Guenther, R.: Hierarchical kinematic analysis of robots. Mechanism and Machine Theory 38, 497–518 (2003)
10. Fiedler, M.: Algebraic connectivity of graphs. Czechoslovac Math. J. 23, 298–305 (1973)
11. Mohar, B.: The Laplacian spectrum of graphs. In: Alavi, Y., et al. (eds.) Graph Theory, Combinatorics and Applications, vol. 2, pp. 871–898. Wiley, New York (1991)
12. Pothen, A., Simon, H., Liou, K.P.: Partitioning sparse matrices with eigenvectors of graphs. SIAM J. Matrix Anal. Appl. 11, 430–452 (1990)
13. Paulino, G.H., Menezes, I.F.M., Gattass, M., Mukherjee, S.: Node and element resequencing using Laplacian of a finite element graph, Part I –general concepts and algorithms. Int. J. Numer. Meth. Eng. 37, 1511–1530 (1994)
14. Kaveh, A., Bondarabady, H.A.R.: A multi-level finite element nodal ordering using algebraic graph theory. Finite Elements in Analysis and Design 38, 245–261 (2002)

Workspace Analysis and Parameter Optimization of a Six DOF 6-3-3 Parallel Link Machine Tool

Minghui Zhang and Baohai Zhuo

College of Mechanical and Electronic Engineering,
Shandong University of Science and Technology, China

Abstract. Taking the 6-3-3 parallel mechanism as the researching target and the inverse kinematics equation of the mechanism is built. Based on this, the relationships between the displacement of the slides and the full reachable workspace are deduced in the mathematic functions. Then boundary curves of the full reachable workspace are obtained by means of the algorithm of mathematical analysis under the traversed range of the slides, spherical joint fit corner and parallel bar interference constraints. And then the projection and section curve graphs of work space boundary surfaces are drawn. Finally, the relations between the full reachable workspace and the structure parameters are simulated, and the Genetic Algorithm is used to optimize the structural parameters in order to obtain the maximal workspace volume.

1 Introduction

Since Stewart invented the six freedom parallel machine, many researchers pay close attention to the parallel robot. The parallel robot takes advantage of strong carrying capacity, minor error, high precision, small turndown radio, excellent dynamic performance and control easily. A parallel link manipulator is considered to be an important construction feature of machine tools in the future. However, in comparison with serial robot, the workspace of parallel robot is small to be affected by the constraint of kinematic pair, the links interference, singularity, the pose error and so on. The workspace is important performance index to evaluate the robot. It is meaningful to study the relation between the structure parameter and the workspace, in order to obtain the bigger workspace volume by optimal design. The research work about the workspace mainly include the workspace analysis of six freedom parallel mechanism[2,3,4] and parameter optimization based on workspace analysis[5,6]. The full reachable workspace of a six DOF 6-3-3 parallel link machine tool can hardly solve with traditional search algorithm[7], so the workspace analysis mainly focused on the reachable workspace with fixed pose of the moving platform in the past references. However, the volume of the full reachable workspace is usually greater than of the fixed pose workspace, even singular configuration of the machine is avoided by adjusting the moving platform pose[8]. Thus taking the 6-3-3 parallel mechanism for an example, an mathematical analysis algorithm is raised in order to obtain the boundary of the full reachable workspace and research the relationship between the structure parameters and workspace. And the structure parameters is optimized with targets of the maximum workspace.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 706–712, 2009.

2 Computation of the Full Reachable Workspace

2.1 The Inverse Kinematics

In this paper, the construction of a six DOF 6-3-3 parallel link machine is considered and is shown in Fig. 1.It is constituted of a fixed base plate and a mobile plate connected by 6 variable-length links. One of the extremities of each link is articulated with the base plate through a prismatic joint and the other extremities of each link articulated with the mobile plate through a spherical joint. Six moving axes driven by six servo motors are included in this system. A fixed coordinate system *OXYZ* for this system is located on the upper base frame. Set a_1, a_2, a_3 and a_4 are the distances between two moving axes, $a = [-a_1, 0, a_2, -a_3, 0, a_4]$. b_1, b_2, b_3, b_4, b_5 and b_6 are the displacements of the sliders $B_i (i = 1, \cdots 6)$.

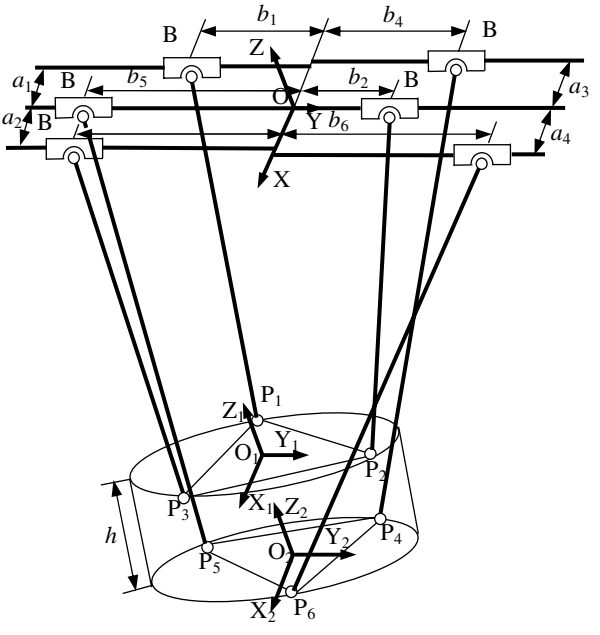


Fig. 1. The geometrical relationships of the six DOF 6-3-3 parallel link machine tool

The intersection points of the links and the slides relative to the origin of the fixed coordinate system *OXYZ* are described as follows:

$$B_i = [a \quad b_i \quad 0] \qquad i = 1, \cdots 6 \tag{1}$$

Similarly, The expression of intersection points of the links and the moving platform relative to the moving coordinate system $O_1X_1Y_1Z_1$ are obtained as follows:

$$P = \begin{bmatrix} -d/2 & -\sqrt{3}/6d & 0 \\ 0 & \sqrt{3}/3d & 0 \\ d/2 & -\sqrt{3}/6d & 0 \\ -d/2 & \sqrt{3}/6d & 0 \\ 0 & -\sqrt{3}/3d & 0 \\ d/2 & \sqrt{3}/6d & 0 \end{bmatrix} \tag{2}$$

R represent the orientation of the moving platform. α, β, γ are the Euler angles. From the geometry relationship, the following results are obtained:

$$R = \begin{bmatrix} \cos\beta\cos\gamma & -\cos\beta\sin\gamma & \sin\beta \\ \sin\alpha\sin\beta\cos\gamma+\cos\alpha\sin\gamma & -\sin\alpha\sin\beta\sin\gamma+\cos\alpha\cos\gamma & -\sin\alpha\cos\beta \\ -\cos\alpha\sin\beta\cos\gamma+\sin\alpha\sin\gamma & \cos\alpha\sin\beta\sin\gamma+\sin\alpha\cos\gamma & \cos\alpha\cos\beta \end{bmatrix} \tag{3}$$

The length of the link is L_i, the location of the moving platform is express:

$$O_1 = L_i - P_i R + B_i \qquad (i = 1, \cdots 3) \tag{4}$$

$$O_2 = L_i - P_i R + B_i \qquad (i = 4, \cdots 6) \tag{5}$$

The relationship between the distance of slides and the space coordinates and the pose of the moving platform is solved from the equation of (4) and (5):

$$\begin{aligned} &(x_j + P_{ix}R_{11} + P_{iy}R_{12} + a_i)^2 + (y_j + P_{ix}R_{21} + P_{iy}R_{22} + b_i)^2 + \\ &(z_j + P_{ix}R_{31} + P_{iy}R_{32} + a_i)^2 = L_i^2 \qquad (j = 1, i = 1,2,3; j = 2, i = 4,5,6) \end{aligned} \tag{6}$$

2.2 Structure Constraints

(1) The slides move within a restricted range because of the limits of the real structure and driving motors. The maximum distance and the minimum distance of slides are $b_{\max}$, $b_{\min}$, then the displacement of sliders should satisfy with the expression as follows:

$$b_{\max} - b_i \geq 0 \qquad b_i - b_{\min} \geq 0 \tag{7}$$

(2) θ_i is the conically angle of sphere pair of the links and the fixed platform, $\theta_i = \arccos(P_{iz}/l_i)$, θ_i' is the conically angle of sphere pair of the links and the moving platform then $\theta_i' = \arccos(\vec{n} \cdot \vec{l}_i / l_i)$ $\vec{n}$ is the unit normal vector of the moving platform**.** The rotation angle of the sphere pair is described as:

$$\theta_{\max} - \theta_i \geq 0 \qquad \theta_{\max} - \theta_i' \geq 0 \tag{8}$$

(3) Interference detection among the links**.** Because each link have some physical size in the mechanism, the links could produce an interference when the mechanism moves. The direction vector of adjacent links is B_iP_i $B_{i+1}P_{i+1}$,their common normal length is d_i , the vertical position vector between the common normal line and the two links is c_i, c_{i+1} , then

$$d_i = \left| \frac{(B_iP_i \times B_{i+1}P_{i+1}) \cdot (c_{i+1} - c_i)}{\left| B_iP_i \times B_{i+1}P_{i+1} \right|} \right|$$

The cross section diameter of the link is the conditions of no interference as follows:

$$d_i \geq D, \qquad i = 1, \cdots 6 \tag{9}$$

2.3 The Relationship between the Slide Displacement and the Full Reachable Workspace

The definition of the full reachable workspace is a set which containing every position that the moving platform can arrive. It is closely related to the moving platform shape, volume and loads. Firstly, the full reachable workspace of the reference point O_1 is researched in this paper. The movement of the platform is decomposed with the three resolving motions and three translation, the equation about the slider displacement and the motion parameters is obtained basis of the formula(6). According to this, the limit moving range of the platform can be found. The relation of the slide displacement and *Z* direction of the moving platform is set up, by now,*X*1=0, *Y*1=0, *i*=1, 2, 3; *X*2=0, *Y*2=0, *i*=4, 5, 6, the rotation angle is 0.Acording to the formula(6),z_i is described as follows:

$$z_i = sqrt(l_i^2 - b_i^2 - 2 * P_{iy} * b_i - P_{iy}{}^2 - (P_{ix} + a_i)^2) \qquad (i = 1,2, \cdots 6) \tag{10}$$

Let $z_a = [z_{\min}, z_{\max}]$ the search space is divided into lots along *Z* axis, a polar coordinate system is set up on the parallel *Xa Ya* plane**,** the angle between a polar diameter and the positive half of the x-axis is a polar angle ϕ . Thus the relations of a polar diameter ρ and the slide displacement can be obtained as follows:

$$\rho^2 + 2\rho\left(b_i \sin\phi + P_{iy} \sin\phi + P_{ix}\cos\phi - a_i \cos\phi\right) + b_i^2 + 2P_{iy}b_i + P_{iy}^2 + \left(P_{ix} - a_i\right)^2 + z_a^2 - L_i^2 = 0 \tag{11}$$

When the moving platform revolves around *Xa* axis, the relations of the slide displacement and the rotation angle can be obtained as follows:

$$b_i^2 - 2P_{iy}Z_a \sin\alpha + 2b_iP_{iy}\cos\alpha + P_{iy}^2 + Z_a^2 + \left(P_{ix} + a_i\right)^2 - L_i = 0 \tag{12}$$

When the moving platform revolves around axis, the relations of the slide displacement and the rotation angle can be obtained as follows:

$$b_i^2 - 2P_{ix}Z_a \sin\beta + 2a_i P_{ix}\cos\beta + P_{ix}^2 + Z_a^2 + 2b_i P_{iy} + a_i^2 - L_i = 0 \qquad (13)$$

When the moving platform revolves around Z axis, the relations of the slide displacement and the rotation angle can be obtained as follows:

$$b_i^2 - 2\left(b_i P_{ix} - a_i P_{iy}\right)\sin\gamma + 2\left(b_i P_{iy} + a_i P_{ix}\right)\cos\gamma + P_{iy}^2 + Z_a^2 + P_{ix}^2 - L_i = 0 \qquad (14)$$

3 Mathematical Analysis Algorithm and Example

3.1 Mathematical Analysis Algorithm

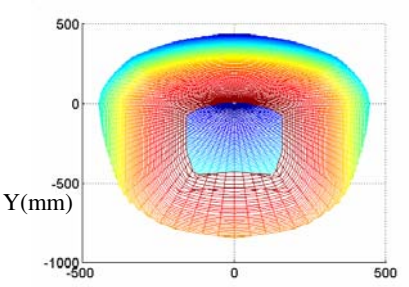
Firstly, the partial derivatives of the formula (11) is calculated**,** let its partial derivatives $\frac{\partial\rho}{\partial b_i} = 0$,the extreme points of the polar diameter is obtained $[\rho_{i\min}, \rho_{i\max}]$. Let $\rho_{\min} = \max[\rho_{i\min}], \rho_{\max} = \min[\rho_{i\max}]$, the workspace boundary content with restrains is obtained. Because the points on the workspace boundary must satisfy the maximum constraints, the formula （8 （9 will turn equality, the length of the polar diameter is solved. Comparing the former polar diameter, if it is in the range of $[\rho_{i\min}, \rho_{i\max}]$, the workspace boundary is obtained,if it is out the range of $[\rho_{i\min}, \rho_{i\max}]$, the pose angle which is greatest influence factor to the slider displacement is found by means of the formula(12), (13) and (14), based on the angle value, the polar diameter is computed according to the formula (6).

3.2 Example Analysis

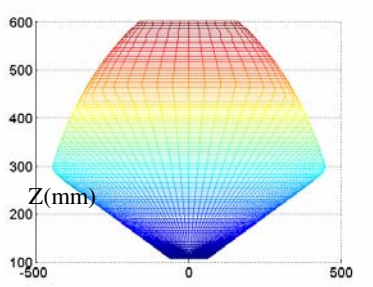
The structure parameters of the 6-3-3 parallel mechanism is given in the table 1. The range of the slide displacement is $b = [100,700]$. The maximum conically angle of sphere pair is $\theta_{\max} = 65^\circ$. The range of rotation angle around x, y, z axis is $-10^\circ \le \alpha \le 10^\circ \quad -10^\circ \le \beta \le 10^\circ \quad -10^\circ \le \gamma \le 10^\circ$

Table 1. The structure parameters of the 6-3-3 parallel mechanism unit：mm

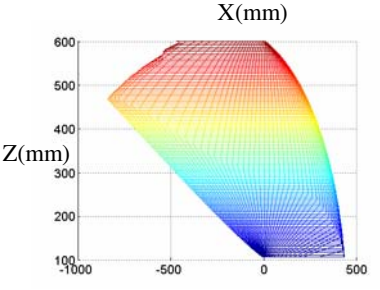
a_1=250	a_2=250	a_3=250	a_4=250	d=160	h=20
L_1=684.18	L_2=630.76	L_3=684.18	L_4=714.40	L_5=657.57	L_6=714.40



(a) Projection in the ***XY*** plane



(b) Projection in the ***XZ*** plane



(c) Projection in the ***YZ*** plane

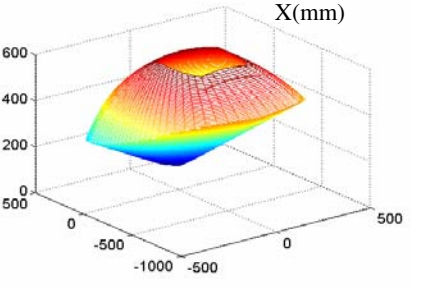


Fig. 2. The full reachable workspace of the platform whenthe upper platform's attitudes change random

4 Conclusion

1. An mathematical analysis algorithm is raised in order to obtain the boundary of the full reachable workspace based on the theoretical analysis and formula derivation. This method takes advantage of a high calculation precision and fast in speed of operation. At the same time, the difficult problem which the traditional search algorithm can hardly give the full reachable workspace of a six DOF 6-3-3 parallel link machine tool is solved.

2. Because the volume of the full reachable workspace is usually greater than of the fixed pose workspace, the performance of the parallel robot is proved. The method in this paper can be used to other parallel mechanisms

References

[1] Stewart, D.: A platform with six degrees of freedom. In: Proc. Instn. Mech Engrs (part I), vol. 180(15), pp. 371–385 (1965)

[2] Liu, Y.W.: Study on theWorkspace and Parameter Optimum of Parallel Mechanisms. Mechanical Science and Technology 25(2), 184–188 (2006)

[3] Cao, Y.G.: Workspace Analysis and Parameter Optimization of 6-RSS Parallel Mechanism. Chinese Journal of Mechanical Engineering 44(1), 19–24 (2008)
[4] MajidM, Z.A., et al.: Workspace analysis of a six-degrees of freedom, three-ppr parallel manipulator. Advanced Manufacturing Technology 16, 41–44 (2000)
[5] Zhang, L.J.: Mechanism Design of Spherical 5R Parallel Manipulator Based on Workspace. Chinese Journal of Mechanical Engineering 43(2), 55–59 (2008)
[6] Snyman, J.A., du Plessis, L.J., Duffy, J.: An optimization approach to the determination of the boundaries of manipulator workspaces. Journal of Mechanical Design, Transactions of the ASME 122, 447–456 (2000)
[7] Cheng, J.: Research on Full Reachable Workspace of the Parallel 62DOF Platform. Journal of Astronautics 28(2), 389–393 (2007)
[8] Chen, S.L., You, I.T.: Kinematic and singularity analyses of a six DOF 6-3-3 parallel link machine tool. Advanced Manufacturing Technology (16), 835–842 (2000)

Keeping a Stable Position of Walking Robot with Vibration

Grigory Panovko[1], Evgeniya Myalo[1], and Teodor Akinfiev[2]

[1] Mechanical Engineering Research Institute (IMASH RAS),
4, Maly Kharitonievsky per, 101990, Moscow, Russia
gpanovko@yandex.ru
[2] Instituto de Automática Industrial (IAI CSIC), Carretera de Campo Real,
km 0.200. La Poveda. Arganda del Rey. Madrid. E-28500 Spain
teodor@iai.csic.es

Abstract. A walking robot is considered. For leg lightness it is suggested in some cases to miniaturize its cross section. At that buckling of the leg rectilinear form is possible under dead weight. To keep the rectilinear form it is proposed to place a vibroexciter for vertical oscillations in the bottom of the leg. It results in parametrical oscillations which make onset for effect of vibrational stabilization of the rectilinear form. This effect can be treated as an increase of flexural stiffness of the flexible rod. Scheme of the leg's model in the form of flexible rod with distributed mass and concentrated mass at its free end is developed. Taking into account finite rotations of its sections we derive the equation of oscillations for the rod. It is shown that the effect of rectilinear form stabilization occurs under certain magnitudes of vibration frequency and amplitude. Stability conditions for obtained solutions are analyzed.

Keywords: Walking robot, robot's leg, elastic rod, vibrational stabilization, vibroexciter, parametric oscillations, mathematical modeling, stability domains.

1 Introduction

It is well known that walking robots are characterized by low velocity and high energy expenses [1]. This is largely due to the fact that drives of such robots have to work in two considerably different regimes: 1) displacement of a heavy body of the robot; 2) displacement of the robot's legs.

Naturally, setting up of a drive is made for the most difficult of the regimes - the regime of movement of the robot's body. As a result, the movement of the legs of the robot is far from the optimum mode resulting in underutilization of the capacity of the motor. The developed earlier in IAI CSIC (Spain) Dual SMART Drive [2, 3] allows to solve in part the problem of increasing the speed of walking machines by allowing independent optimal configuration at using motions with two different masses (Fig. 1). This allows to increase several times the speed of the robot's legs movement. Further increase in speed of robot's legs is not limited by capacity of the drive, but is limited by the dynamic properties of the robot itself.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 713–722, 2009.

Thus, to obtain high-speed movement of the robot it is necessary to have high accelerations (positive at acceleration and negative when braking). This brings about the considerable forces of inertia, which can cause a rollover of legged walking robot. As the forces of inertia are proportional to the mass to be displaced of the leg of the robot, the only way to increase further the speed of movement of the legs of the robot is to reduce the mass to be displaced.

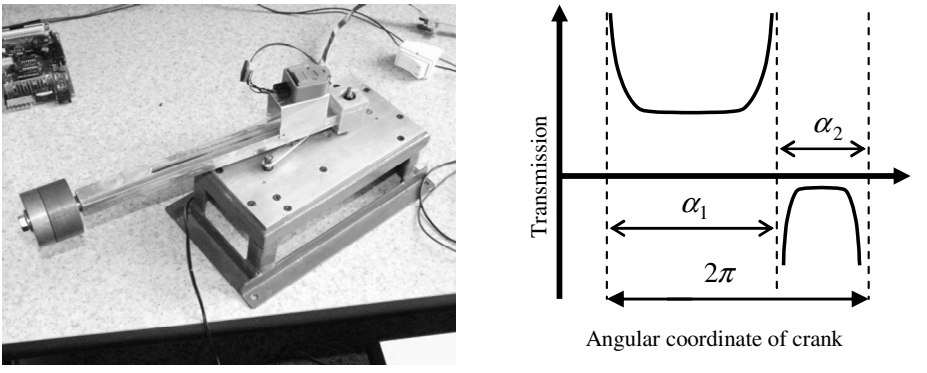


Fig. 1. Prototype of DUAL SMART drive

Reducing the mass of the legs of the robot leads to a reduction in their stiffness and, consequently, to appearance of deformation under the influence of robot's body weight as well as an interruption of the normal mode of operation.

The same problem - the need to reduce the mass of the legs of the robot – appears in case of jumping robots. Thus, in self-resonance jumping robots case with the tension or compression springs (Fig. 2, 3), the forces of inertia are compensated by passive spring, and the motor serves only to compensate for the losses of energy [4-9]. Reducing these losses results in significant reduces of the motor power or increase of the height of the robot's jump.

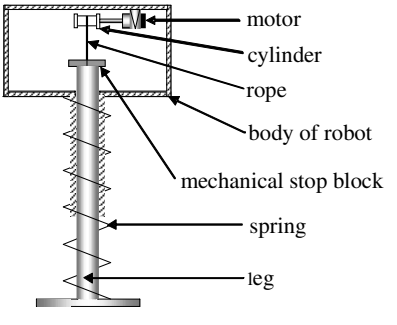


Fig. 2. Hopping robot with a compression spring

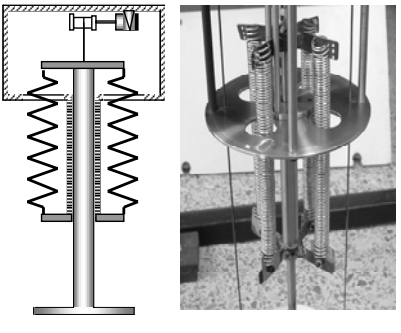


Fig. 3. Hopping robot with extension spring

Significant energy losses occur at the time of inelastic impact of the robot leg upon the supporting surface and at the moment of inelastic impact of mechanical stop block of movement of the leg and body of the robot. These losses can be substantially reduced by reducing the mass of the leg. However, in this case the rigidity of the legs of the robot is reduced, leading to its deformation and disturbance of the normal mode of operation of the robot.

In static conditions (especially at the point of contact with surface) the vertical axis of the leg may bend under gravity (and force of moving space). To keep the rectilinear form of leg's axis it is supposed to place vibroexciter for vertical oscillations at the bottom part of the leg. Such kind of excitation leads to parametric oscillations of the leg. Due to them the effect of vibrational stabilization of the rectilinear form of a flexible rod can be achieved under certain conditions [10 - 12]. On the other hand the effect can be treated as increasing of bending stiffness of the rod.

In this paper, we study a possibility of the use of additional drive, which generates high-frequency parametric oscillation of the leg with small amplitude, to prevent deformation of the flexible leg of the robot with distributed and concentrated masses.

2 Model Description

Scheme example of a jumping robot with the leg *2* realized as flexible elastic rod is represented in the Fig. 4. Robot's elements which provide its vertical displacement (jump) are modeled by perfectly rigid body *1*. Of course, for more complete model description it is desirable to take into account the elastic suspension of bumping drive elements. In the paper [13] it was shown that this accounting gives slightly more precise results but does not lead to additional effects qualitatively. Therefore, for simplification we neglect the influence of additional elastic suspended masses. In the base of the robot's leg a vibrator *3* is mounted. It excites high frequency oscillations of the robot's leg. It is possible to make this drive like motor with disbalance, piezocrystal or electromagnet.

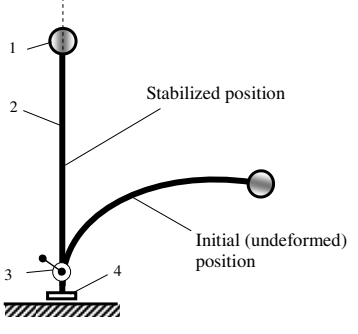


Fig. 4. Scheme of robot's leg with a vibrator

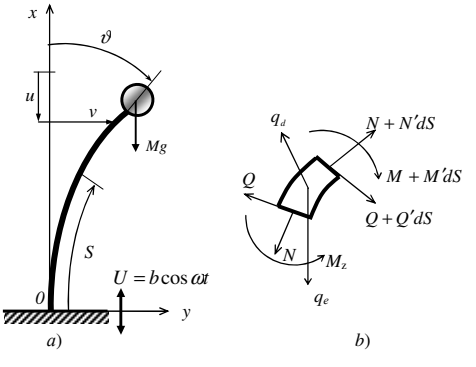


Fig. 5. Scheme of the model of the rod (*a*) and scheme forces applied to the rod's element (*b*)

For simulation and estimation of excitation parameters, under which the effect of the robot's leg straightening (the effect of flexural hardening) is reached, let's consider the cantilevered rectilinear uniform elastic rod (Bernoulli-Euler's beam) with constant cross section. The perfectly rigid body with mass M is fixed at another end of the rod. The rod's axis is nonstretchable. For estimation of amplitudes of lateral oscillations finite rotations of rod's section are taken into account. Material internal damping is described by Fought model; external friction is proportional to absolute velocity.

At the unstrained position the rod's axis coincides with the vertical axis x of the rectangular coordinate system xOy with origin placed in the rigid support of the rod (Fig. 5, *a*). In the absence of external forces the flexible rectilinear rod buckles under gravity force of the concentrated mass Mg and uniformly distributed gravity force mgL, and its axis bends (Fig. 5, *a*) (m - mass of rod length unit, L - length of the rod, g - gravity acceleration). This rod is called as supercritical one [10, 12].

In general case in dependence of vibrational excitation a periodical force $P = P_0 \cos \omega t$ (P_0 - force amplitude, ω - frequency), that is force excitation, or vertical movement $U(t) = b \cos \omega t$ (b - amplitude of oscillations), that is kinematic excitation, can be applied to the base of the rod.

3 Differential Equation of Parametric Oscillations of the Rod

Motion of the arbitrary cross section of the rod is described in the moving coordinate system $x0y$ with origin placed in the rigid support of the rod (Fig.5, *a*). Actual coordinate S of the rod's section is count off from the base. Forces and moments acting on the rod's element are shown in the Fig. 5, *b*. M_z is bending moment, Q and N are transversal and longitudinal forces correspondingly, q is total distributed external axial load produced by gravity and forces of moving space, q_d is distributed force of external damping proportionate to the absolute velocity of the rod's element.

For the Bernoulli-Euler rod the internal bending moment with internal damping is described by the expression [14]:

$$M_z(t,S) = EI\left[\kappa(t,S) + d_1 \frac{\partial \kappa(t,S)}{\partial t}\right], \quad \kappa(t,S) = \frac{\partial \vartheta(t,S)}{\partial S} = \vartheta'(t,S), \tag{1}$$

E - elastic modulus, I - moment of inertia of a section, $\kappa(t,S)$ - rod's curvature, ϑ - rod's axis angle of rotation relative to the vertical, d_1 - the internal damping coefficient.

Positive direction of the vertical displacement u of rod's sections is deflection against axis x. Positive direction of the bending v coincides with the direction of y axis. From the definition of the rod's axis angle of rotation it is follows: $v' = \sin \vartheta$, $u' = 1 - \cos \vartheta$

In the issue the expression for rod's axis curvature κ in dependence of the bending v:

$$\kappa = \vartheta' = v''\left(1 - v'^2\right)^{-1/2} = v''\left(1 + v'^2/2 + 0\left(v'^4\right)\right) \approx v''\left(1 + v'^2/2\right). \tag{2}$$

Axial deflection u can be expressed in terms of lateral deflection v :

$$u = S - \int_0^S \left(1 - v'^2\right)^{1/2} dS \approx \frac{1}{2}\int_0^S v'^2 dS \,. \tag{3}$$

At this point it is considered that $u\big|_{S=0} = 0$, $u' = 1 - \sqrt{1 - \sin^2 \vartheta} \approx \frac{1}{2} v'^2$.

Eq. (1) for bending moment taking into account Eq. (2):

$$M_z = EI\left[v''\left(1 + v'^2/2\right) + d_1 \frac{\partial v''\left(1 + v'^2/2\right)}{\partial t} \right]. \tag{4}$$

Equation for the moments balance is given by:

$$M_z' dS + Q dS = 0 \,. \tag{5}$$

Whence expression for lateral force taking into account Eq. (4):

$$\begin{aligned} Q &= -EI\left[v''\left(1 + v'^2/2\right) + d_1 \frac{\partial v''\left(1 + v'^2/2\right)}{\partial t} \right]' = \\ &= -EI\left(v''' + v'^2 v'''/2 + v' v''^2\right) - EId_1 \frac{\partial\left(v''' + v'^2 v'''/2 + v' v''^2\right)}{\partial t}. \end{aligned} \tag{6}$$

In case of force excitation rod motion is considered in the absolute coordinate system. In the case of kinematic excitation motion of the rod's element is considered in the relative coordinate system and following calculations have general character. Therefore, we consider indeed the case below.

In the relative coordinate system it is necessary to take into account inertia forces from translational frame motion along axis x . Then total external axial distributed load (direction against axis x , Fig. 5, *b*)

$$q_e(t) = \left[m + M\delta(S - S_M)\right]\left(g + \ddot{U}\right) = \left[m + M\delta(S - S_M)\right]\left(g - b\omega^2 \cos \omega t\right), \tag{7}$$

S_M is the position of mass fixation and for the case considered $S_M = L$, $\delta(S - S_M)$ is delta-function.

External friction proportionate to the absolute velocity $\mathbf{V}$ (linear viscous model) results in the additional distributed force,

$$\mathbf{q}_d = -d\mathbf{V}\,, \qquad \mathbf{V} = -\left(b\omega \sin \omega t + \partial u/\partial t\right)\mathbf{i} + \left(\partial v/\partial t\right)\mathbf{j}\,, \tag{8}$$

d is external damping coefficient, $\{\mathbf{i}, \mathbf{j}\}$ are unit vectors of the coordinate system axes $\{x, y\}$. x and y projections of the external friction force vector are given by:

$$q_{dx} = d\left(b\omega \sin \omega t + \partial u/\partial t\right);\; q_{dy} = -d\left(\partial v/\partial t\right). \tag{9}$$

Equilibrium equation for the differential rod's element for x and y projections:

$$\begin{aligned}\left[m+M\delta(S-S_M)\right]\frac{\partial^2 u}{\partial t^2}&=-(N\cos\vartheta)'+(Q\sin\vartheta)'+q_e(t)+q_{dx}(t),\\ \left[m+M\delta(S-S_M)\right]\frac{\partial^2 v}{\partial t^2}&=(N\sin\vartheta)'+(Q\cos\vartheta)'+q_{dy}(t).\end{aligned}\qquad(10)$$

Inextensibility condition imposed on rod axis limits the axial deflections u. Thus, the first equation of the system (10) allows to determine normal force N. From Eq.(10) after required transformation it is possible to obtain the equation which describes third-order infinitesimal (relative to lateral deflections) nonlinear lateral oscillations of the vertical cantilevered rod mounted on the vibrating base with nonstretchable axis under gravity.

$$\begin{aligned}&\left[m+M\delta(S-S_M)\right]\frac{\partial^2 v}{\partial t^2}+d\frac{\partial v}{\partial t}+EIv^{IV}+d_1EI\frac{\partial v^{IY}}{\partial t}+q_\Sigma(t)\left[(L-S)v'\right]'=\\ &=\frac{1}{2}\left[m+M\delta(S-S_M)\right]\left[v'\int_S^L d\tilde{S}\int_0^{\tilde{S}}\frac{\partial^2\left(v'^2\right)}{\partial t^2}d\hat{S}\right]'-EI\left[v'\left(v'v''\right)'\right]'-\\ &-d_1EI\left[\frac{\partial v'\left(v'v''\right)'}{\partial t}\right]'-q_\Sigma(t)\left[(L-S)\frac{v'^3}{2}\right]'+\frac{d}{2}\left[v'\int_S^L d\tilde{S}\int_0^{\tilde{S}}\frac{\partial(v')^2}{\partial t}d\hat{S}\right]',\end{aligned}\qquad(11)$$

$$\left(q_\Sigma(t)=q_e(t)+db\omega\sin\omega t\right).$$

For subsequent analysis let us reduce Eq. (11) to dimensionless form using dimensionless parameters:

$$\tau=\frac{t}{L^2}\sqrt{\frac{EI}{m}}\text{ - time, }\ \zeta=\frac{S}{L}\text{ - angular position, }\ \xi=\frac{v}{h}\text{ - lateral deflection,}\qquad(12)$$

h is characteristic scale, which magnitude is chosen to fulfill the conditions for infinitesimal parameter ε introduced below (for example, h is rod diameter).

With such parameters Eq. (11) becomes dimensionless:

$$\begin{aligned}&\left[1+\mu\delta(\zeta-\zeta_M)\right]\ddot{\xi}+2\psi\dot{\xi}+\xi^{IV}+2\psi_1\dot{\xi}^{IV}+p_\Sigma(\tau)\left[(1-\zeta)\xi'\right]'=\\ &=\varepsilon\left\{\begin{array}{l}\left[1+\mu\delta(\zeta-\zeta_M)\right]\xi'\int_\zeta^1 d\zeta_1\int_0^{\zeta_1}\left(\dot{\xi}'^2+\xi'\ddot{\xi}'\right)d\zeta_2-\xi'\left(\xi'\xi''\right)'-\\ -2\psi_1\left[\xi'\left(\xi'\xi''\right)'\right]^{\bullet}-p_\Sigma(\tau)(1-\zeta)\xi'^3/2+\xi'\int_\zeta^1 d\zeta_1\int_0^{\zeta_1}2\psi\xi'\dot{\xi}'d\zeta_2\end{array}\right\}',\end{aligned}\qquad(13)$$

$$p_\Sigma(\tau)=\left[1+\mu\delta(\zeta-\zeta_M)\right]\left(\gamma-\beta\Omega^2\cos\Omega\tau\right)-2\psi\beta\Omega\sin\Omega\tau.$$

Following are non-dimensional complexes:

$$\Omega = \omega L^2 \sqrt{\frac{m}{EI}}, \beta = \frac{b}{L}, \gamma = \frac{mgL^3}{EI}, \varepsilon = \left(\frac{h}{L}\right)^2, \psi_1 = \frac{d_1}{2L^2}\sqrt{\frac{EI}{m}}, \psi = \frac{dL^2}{2\sqrt{mEI}}, \mu = \frac{M}{m}. \quad (14)$$

4 Solution of Equation for Parametric Rod's Oscillation

Solution $\xi(\zeta,\tau)$ of the equation (13) is given by multimodal approximation:

$$\xi(\zeta,\tau) \approx \sum_i^n q_i(\tau)\varphi_i(\zeta), \quad (15)$$

where $q_i(\tau)$ are amplitude functions, $\varphi_i(\zeta)$ are continuous and four times differentiable coordinate functions satisfying to boundary conditions:

$$\zeta = 0: \ \xi = 0, \xi' = 0; \quad \zeta = 1: \xi'' = 0, \xi''' = 0. \quad (16)$$

Consider linearized Eq. (13) for determination of stability domains of the rod vertical position. After substitution of the Eq. (15) into the linearized Eq. (13) and posterior orthogonalizaiton with coordinate functions $\varphi_i(\zeta)$ in the range $\zeta \in [0,1]$ we deduce system of *n* ordinary differential equations for amplitude functions $q_i(\tau)$.

As coordinate functions, forms of rod transverse oscillations, determining from equation $\varphi^{IV} - \alpha^4\varphi = 0$ and respective boundary conditions (16), are taken. Solution of the eigen modes problem of the cantilever rod is produced with help of Krylov's functions [14].

For determination of stability domains for vertical rectilinear form of the rod's axis the linearized system of equations regarding to amplitude functions $q_i(\tau)$ is considered. Parameters domain under investigation is limited by third frequency of the main parametric resonance (*n*=6). Two-parameter domains, where stabilization of rectilinear form of the rod's axis is possible, were obtained as the result of numeric simulation. Stability analysis is carried out according to Floquet - Lyapunov's theory [15]. As a stability criterion it is taken inequality $\max|\mu_i| < 1, \ i = \overline{1,2n}$, where μ_i are multipliers.

Amplitude – frequency of excitation parameter domain (β,Ω) corresponding to similar behavior of multipliers for flexibility parameter $\gamma = 8$ is shown in Fig. 6, *a* (white regions – stable state). In Fig. 6, *b* amplitude – frequency of excitation parameter domain (β,Ω) is shown for flexibility parameter $\gamma = 10$. Low boundary of stability domain (Fig. 6, *a* – black curve), satisfies the next inequality [10]:

$$\beta_{crit} = \frac{1}{\Omega}\sqrt{\frac{2(\gamma - \gamma_{crit})}{|J|}}, \qquad J = -\int_0^1 (\varphi_1')^2 (1-\zeta) d\zeta, \quad (17)$$

γ_{crit} - value of flexibility parameter in case the rod buckles in static conditions under gravity.

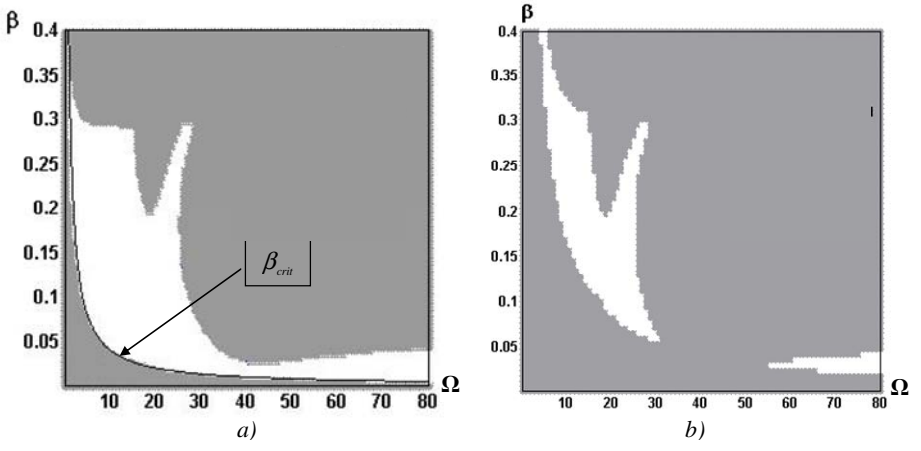


Fig. 6. Instability domains for the supercritical rod ($\psi = 0, \psi_1 = 0{,}01$)**:** *a* - $\gamma = 8$, *b* - $\gamma = 10$

At frequency interval (15, 27) for excitation amplitude $\beta = 0{,}25$ one can see (Fig. 6) unstable values of trivial solution; it corresponds to subresonance at excitation frequency $\Omega_{sub} = 22$. At frequency interval (25, 80) - second main appears parametric resonance at $\Omega = 2\Omega_{sub} = 44$.

Increase of flexibility parameter γ makes instability domains wider. It can be seen from comparison between Fig. 6, *a* and Fig. 6, *b*. Analysis shows that linear damping increases stability domains.

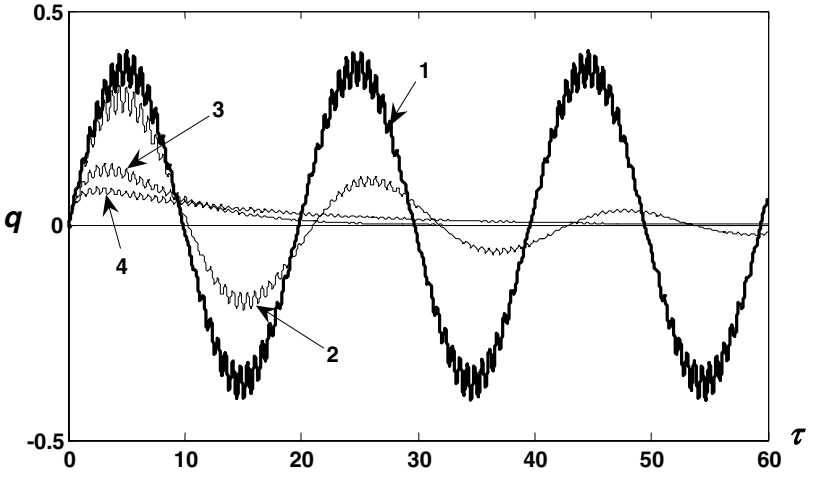


Fig. 7. Oscillating proccess for the rod: 1 – without friction; 2 – with external friction $(\psi = 0{,}05;\ \psi_1 = 0)$; 3 – with external and internal fricrion $(\psi = 0{,}025;\ \psi_1 = 0{,}025)$; 4 – with internal friction $(\psi = 0;\ \psi_1 = 0{,}05)$

For the problem under consideration the most interest represents the process of oscillations steadiness for vibrational stabilization of the rod. In Fig. 7 time depending curves for different values of external and internal friction coefficients are shown.

Analysis of the results shows that for the same equivalent value of external and internal friction coefficients internal friction has more influences on stabilization (processes 2 and 4, Fig. 7).

Simulation shows that for the excitation level which is lower than critical $\beta < \beta_{crit}$ stabilization does not take place and the rod oscillates with small amplitude near deflected equilibrium position. If the excitation level is high enough ($\beta > \beta_{crit}$) then the rod stabilizes near vertical position.

5 Conclusion

Analysis of results shows that for the same coefficients of external and internal friction the last one exerts more stabilization influence (processes 2 and 4 in the Fig.7) Simulation shows that if the excitation level is lower than critical one when stabilization does not occur and the rod oscillates near deflected equilibrium position with small amplitude. If the excitation level is high enough the rod becomes stabile near vertical position. As follows from Eq. (24) the higher excitation frequency the less excitation amplitude is required for stabilization. This conclusion allows to recommend piezoelectric vibroexciters as vibrational stabilizer for the walking robot's flexible leg.

Acknowledgements. The paper was supported by grants: Russian Foundation for Basic Research N 07-08-00253-a, Ministry of Education and Sciences of Russian Federation N 2.1.1/5248 and CRDF REC-018.

T. Akinfiev acknowledges the financial support received from CSIC under the project "New actuators with high efficiency and control algorithms for automation and robotics".

References

1. The walking machine catalogue, `http://www.fzi.de/ids`
2. Akinfiev, T., Armada, M., Fernandez, R.: Actuator for a working element, particularly a walking robot and the method of controlling said actuator. EC patent EP 1514776, USA patent 7341122
3. Akinfiev, T., Fernandez, R., Armada, M.: Dual SMART drive for walking robots. In: 5th International Conference on Climbing and Walking Robots, pp. 595–602. Professional Engineering Publishing limited, London (2002)
4. Raibert, M.H.: Legged robots that balance. MIT Press, Cambridge (1986)
5. Rad, H., Gregorio, P., Buehler, M.: Design, modeling and control of a hopping robot. In: IEEE/RSJ Int. Conf. Intelligent Systems and Robots, Yokohama, Japan, pp. 1778–1785 (1993)
6. Prosser, J., Kam, M.: Control of hopping height for a one-legged hopping machine. Mobile Robots VII 1831, 604–612 (1992)

7. Gregorio, P., Ahmadi, M., Buehler, M.: Design, control and energetic of an electrically actuated legged robot. IEEE Trans. Systems, Man and Cybernetics 27(4), 626–634 (1997)
8. Akinfiev, T., Armada, M., Fernandez, R., Montes, H.: Hopping robot and control method. Robot saltador y su método de control. Spanish patent ES2209617
9. Akinfiev, T., Armada, M., Montes, H.: Vertical Movement of Resonance Hopping Robot with Electric Drive and Simple Control System. In: 11th IEEE Mediterranean Conference on Control and Automation, MED 2003 (2003)
10. Gouskov, A.M., Panovko, G.Y.: Vibrating stabilization of a vertical axis of a flexible rod. J. Problems of mechanical engineering and reliability of machines (Вибрационная стабилизация вертикальной оси гибкого стержня. Проблемы машиностроения и надежности машин) 5, 13–19 (2006)
11. Hamdan, M.N., Al-Qaisia, A.A., Al-Bedoor, B.O.: Comparison of analytical techniques for nonlinear vibrations of a parametrically excited cantilever. International J. of Mechanical Sciences 43, 1521–1542 (2001)
12. Champneys, A.R., Fraser, W.B.: The 'Indian rope trick' for a parametrically excited flexible rod: linearized analysis. Proc. R. Soc., London, A 456, 553–570 (2000)
13. Blekhman, I.I., Sperling, L.: Stabilization of the synphase postcritical regime of rotation of vibroexciters by means of internal degree of freedom. In: 3rd Polyakhov Readings, St.Petersburg, pp. 107–111 (2003)
14. Biderman, V.L.: The applied theory of mechanical oscillations. The manual for high technical schools (Прикладная теория механических колебаний. Учебное пособие для втузов). High school, Moscow (1972)
15. Iooss, G., Joseph, D.D.: Elementary stability and bifurcation theory, p. 296. Springer, Heidelberg (1990)

An Algebraic Approach for Accurate Motion Control of Humanoid Robot Joints

Jorge Villagra and Carlos Balaguer

Universidad Carlos III, Departamento de Ingenieria de Sistemas y Automatica
Av. Universidad, 30, 28911 Leganes (Madrid), Spain
{jvillagr,balaguer}@ing.uc3m.es

Abstract. Humanoid robots are extremely complex systems, where a multi-layer control architecture is necessary to guarantee a stable locomotion. In the lower layer, joint control has to track as finely as possible references provided by higher layers. A new approach to precisely control humanoid robot joints is presented in this paper. It is based on algebraic control techniques and on a model-free control philosophy. An online black-box identification permit to compensate neglected or uncertain dynamics, such as, on the one hand, transmission and compliance nonlinear effects, and on the other hand, network transmission delays.

1 Introduction

1.1 Preliminaries

Each joint of a humanoid robot is subject to complex and varying loads as the robot moves. These loads are mostly due to the effect of gravity on the robots mass, but also include centrifugal and Coriolis forces from the robots complex motion. In addition, joints may become loaded and unloaded as the robot lifts and places its feet. Indeed, the alternation of support and swinging phases in biped robots are characterized by low speeds and strong loads for the former, and high speeds and weak loads for the latter. These different conditions makes the system very sensitive to friction, flexibility and compliance effects.

Furthermore, joints may have some compliance due to structural stiffness, clearances, and backlashes (cf. [14]). The local compliance at the joint can influence the system dynamics of the humanoid robot when it walks or performs any motion.

It is then challenging to design controllers that maintain a high level of tracking and stability performance under the aforehand mentioned range of load conditions. In typical robotic applications the problem may be addressed by calculating the forward model of the torques on the joints and appropriately compensating. However, since humanoid robots are difficult to precisely model, the determination of feedforward dynamics for model based control can be both a complicated and time consuming process.

Moreover, a belt transmission and a Harmonic Drive reduction are between motor shaft and joint axis in our specific design (see [18] for further details). As a result, there may appear control errors due to external torques, internal flexibilities and compliances. For a precise and robust joint trajectory tracking, the joint angular position measurement is needed, instead of the classical motor shaft position information.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 723–732, 2009.

For this reason, all joints in our RH2 humanoid robot [19] have been conceived with an additional absolute encoder in the shaft joint output, which will complement the classical motor incremental encoder (Fig. 1 shows the physical location of these components).

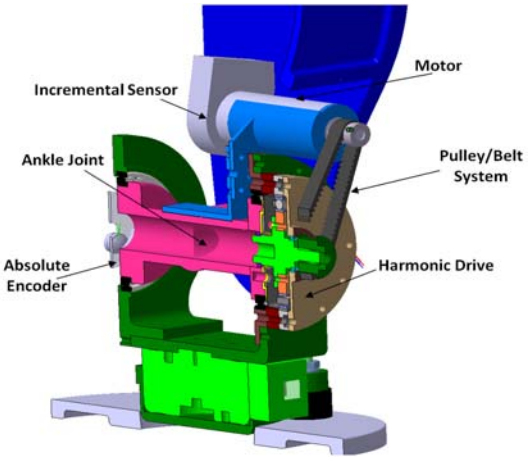


Fig. 1. RH-2 ankle design

Due to the critical importance of some specific joints, an accurate tracking for flexible joints is sought, within a context with a wide payload range and uncertain transmission models.

1.2 Joint Motion Control under Uncertainty

The accurate position control of robot joints has been extensively studied in the last decades. Thus, many different feedback techniques have achieved good position tracking when considering electrically driven rigid robots (cf. e.g. [22]) This task turns out to be critical in humanoid robots, since the stability margins must be respected as precisely as possible in a high range of speeds and under very different payload conditions. However, the transmission systems usually introduce nonlinear dynamics between the motor output and the real joint. Moreover, it is not easy to obtain a precise model of these effects, because of the great number of parts that intervene in the transmission.

Some authors (e.g. [12], [23]) have tried to obtain, through careful modeling, an accurate feedforward control which, in turn, allows to follow a desired trajectory with low-feedback gains. The main problem of this sort of approaches is that they have to deal with highly cross-coupled non-linear models, where the parameters are not always identifiable. Thus, several robust/intelligent control approaches (e.g. [3], [11], [13], [17]) and adaptive control techniques (e.g. [2], [10], [21]) have been proposed to tackle with an appropriate feedback law the transmission uncertainty problem in rigid joint tracking. However, not so many efforts (e.g [4]) have been addressed to control electrically driven flexible-joints under uncertainty. This is because the presence of

joint flexibility greatly increases the complexity of the system dynamics.[1] I In addition, the hardware architecture imposes absolute encoder information to pass through the CAN bus before coming back to the motor driver (see Fig. 2). This message round-trip implies significant delays that make the analysis and control design more complex than with standard control techniques. Thanks to a novel model-free approach, transmission friction, uncertainty related to flexible joints and network delays will be fast and properly compensated. These techniques, initiated in [7], propose an algebraic framework to deal with fast numerical derivatives estimation, and thereafter, nonlinear model free control design.

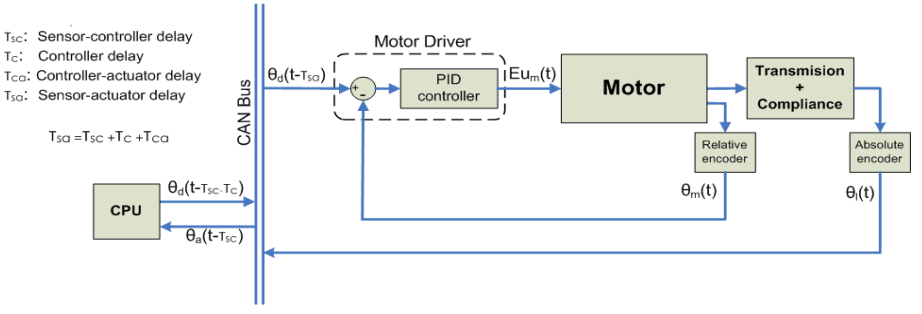


Fig. 2. Global joint control scheme

1.3 Outline

Section 2 will be devoted to detail the system to be controlled. The model used in simulations will be composed of a DC motor, a Harmonic Drive transmission (modeled as a friction torque) and a flexible joint which reproduces the humanoid structure compliance. As depicted in figure 2, the outer loop control will be closed via the CAN bus, whose details in terms of transmission delay will be summarized also in section 2. Since severe nonlinearities as well as complex time-varying phenomena appear in our system, an algebraic model-free controller will be implemented. Its main features will be highlighted in section 3. To show the advantages of the proposed control strategy, it will be compared in simulation with two more standard techniques. Section 4 will be devoted to detail the benchmark and comment the results of this control law comparison. Finally, some concluding remarks and prospective ideas will be given in section 5.

2 System Model

As presented in Fig. 1 and schematized in Fig. 2, the system to be controlled consist of a DC motor and an Harmonic Drive transmission system. Furthermore, flexion torques due to the robot structure compliance, and Coulomb and viscous frictions of the Harmonic Drive have been taken into account.

[1] As it will be detailed in section 2, the consideration of flexibility results in a second-order dynamics, which, if the motor dynamics is also included, implies that a fifth-order dynamics should be considered.

DC Motor. The classical direct current (DC) motor model is considered:

$$J_m\ddot{\theta}_m(t) = K_t i(t) - \tau_{fr}(t) - \tau_{fl}(t)$$
$$L_m\frac{di(t)}{dt} = -R_m i(t) + Eu_m(t) - K_b\dot{\theta}_m(t) \quad (1)$$

where i denotes the armature current, θ_m the angular position of the motor shaft, J_m the rotor inertia, L_m the terminal inductance, R_m the motor terminal resistance, K_b the back electromotive force constant and K_b the torque constant. E represents the maximum available voltage, in absolute value, which excites the machine, while u_m is an input voltage modulation signal, acting as the ultimate control input, with values restricted to the closed real set $[-1,1]$. Friction torque τ_{fr} generated by the Harmonic Drive transmission and flexible torque τ_{fl} due to the structure compliance will be above detailed.

Harmonic Drive Transmission. The model of the reducer can be written as follows (cf. [1] for more details)

$$\dot{\theta}_t = \frac{\dot{\theta}_m}{N},\ \tau_t = \alpha_m N\tau_m \quad (2)$$

where θ_t is the transmission angular position, τ_t is the transmission at the Harmonic Drive output, N is the gearbox ratio and α_m represents the torque transfer coefficient.

Friction. As for any mechanical systems, frictions are difficult to model and many works have been published on the subject either with a wide scope (cf. e.g. [23]) or more focused in Harmonic Drive transmissions (e.g.[1], [5], [12]).

Results from [1] will be used in this work due to the similar biped robot context for which that study was developed. Consider a velocity $\dot{\theta}$ and load τ_l dependent friction force

$$\tau_{fr} = (F_r + \mu|\tau_l|)\left(1 + f_1 e^{-\left(\frac{\dot{\theta}}{\dot{\theta}_0}\right)^2}\right)sign(\dot{\theta}) + f_2\dot{\theta} + f_3\dot{\theta}|\dot{\theta}| \quad (3)$$

where F_r, μ θ_0, f_1, f_2 and f_3 are experimentally obtained constant values.

The static friction is felt at the start but it is not very influent in general working conditions. Indeed, when the direction of the movement is reversed, the speed is canceled without relieving the materials so the Stribeck effect can be neglected ($f_1 = 0$). In addition, the quadratic viscous friction effect appears little in the robot transmissions, therefore the coefficient f_3 can be neglected. As a result, the previous friction torque can be written as $\tau_{fr} = (F_r + \mu|\Gamma|)sign((\dot{\theta}) + f_2\theta$. It can be shown (cf. [1]) that a precise model integrating non symmetrical terms of friction for the two directions of rotation can be obtained from the previous equation. Thus, the motor τ_{fr_m} and joint τ_{fr_t} friction torques (or in other words, the harmonic drive friction model) can be expressed as follows

$$\tau_{fr_t}(\dot{\theta}) = \frac{C_{c_1}}{2}(sign(\dot{\theta}_t + 1) + \frac{C_{c_2}}{2}(sign(\dot{\theta}_t - 1) + f_c\dot{\theta}_t \quad (4)$$

$$\tau_{fr_m}(\dot{\theta}_m) = \frac{C_{m_1}}{2}(sign(\dot{\theta}_m + 1) + \frac{C_{m_2}}{2}(sign(\dot{\theta}_m - 1) + f_m\dot{\theta}_m \quad (5)$$

with C_{c_1} and C_{c_2} terms of Coulomb friction in direct and opposite direction for the connection of the axis of the leg, C_{m_1} and C_{m_2} similar terms for the motor side, and (f_t,

f_m) transmission and motor viscous friction constants. Finally, if the motor-joint torque transmission relationship (2) is introduced in (4) and (5), a single expression can be used to express overall friction effects:

$$\tau_{fr} = f_t \dot{\theta}_m + C_t sign(\dot{\theta}_m) + \Delta C_t \tag{6}$$

with $f_t = f_m + \dfrac{f_c}{\alpha_m N^2}$, $C_t = \dfrac{C_{m_1} + C_{m_2}}{2} + \dfrac{C_{c_1} + C_{c_2}}{2\alpha_m N}$ and $\Delta C_t = \dfrac{C_{m_1} - C_{m_2}}{2} + \dfrac{C_{c_1} - C_{c_2}}{2\alpha_m N}$

Flexible joint. The dynamic effects induced by the structure compliance has been characterized as a flexible link. In the last years an important effort has been devoted to modeling and control flexible joints (cf. i.e. [4], [6], [9], or [16]). However, since a realistic behavior needs complex partial derivative equations, an approximate dynamic model has been chosen in order to make identification and simulation easier.

Consider the dynamics of the flexible system as follows (cf. [9])

$$J_l \ddot{\theta}_l + \tau l - \tau_{fl} = 0,\ \tau_{fl} = c(\theta_m - \theta_l) \tag{7}$$

where J_l the joint inertia, c the joint stiffness constant, τ_l the load torque, and the τ_{fl} the flexible torque.

Remark that the torque τ_{fl}, generated by joint flexible dynamics, is here the same than in equation (1), but with opposite sign.

CAN Bus. As mentioned in the introduction, the absolute encoder located at the output shaft will provide, through a Controller Area Network (CAN) bus, the necessary information to properly correct the joint reference from a global stabilizer.

As any other networked control system, the plant and the controller are spatially separated and the control loop is physically closed through the communication network. This wide spread approach in humanoid robotics results in decreased complexity and cost, easier maintenance and system diagnosis, and higher flexibility. However, the communication between the controller and the plant in such a way introduces a time delay into the closed loop system. It is well-known that time delay in a closed control loop degrades the performance and can lead to instability. Hence, with the aim of considering a model as realistic as possible, both sensor-controller and controller-sensor transmission delays (T_{sc} and T_{cs}, respectively) have been taken into account in simulation. Furthermore, the CPU cycle time has been considered as an additional delay T_{ca} (see Fig. 2).

3 Model-Free Tracking

In order to find out which is the reference angle $v(t)$ for the motor driver to get a better tracking at the output shaft, the following black-box closed-loop control scheme is proposed (cf. [8] for further details).

Take a nonlinear finite-dimensional SISO system

$$\Phi(t, \theta_l, \dot{\theta}_l, \ldots, \theta_l^{(\iota)}, v, \dot{v}, \ldots, v^{(\kappa)}) = 0$$

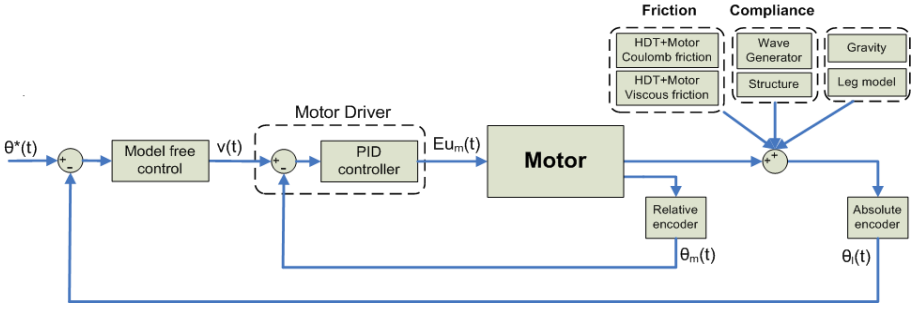


Fig. 3. Model-free joint control scheme in the absence of delays

where Φ is a sufficiently smooth function of its arguments. Assume that for some integer n, $0 < n \leq \iota$, $\frac{\partial \Phi}{\partial \theta_l^{(n)}} \not\equiv 0$. The implicit function theorem yields then locally

$$\theta_l^{(n)} = \Upsilon(t, \theta_l, \dot{\theta}_l, \ldots, \theta_l^{(n-1)}, \theta_l^{(n+1)}, \ldots, \theta_l^{(\iota)}, v, \dot{v}, \ldots, v^{(\kappa)})$$

This equation becomes by setting $\Upsilon = F + \alpha v$:

$$\theta_l^{(n)} = F + \alpha v \tag{8}$$

where

- $\alpha \in \mathbb{R}$ is a constant parameter, which is chosen in such a way that F and αv are of the same magnitude,
- F is determined thanks to the knowledge of u, α, and of the estimate of $\theta_l^{(n)}$. It plays the role of a nonlinear back-box identifier.

Remark 1. This approach obviously necessitates robust derivatives estimation of noisy signals. A very powerful technique developed in [15] will be used.

If the simplest case is considered ($n = 1$), the desired behavior is obtained with the following controller

$$v = \frac{1}{\alpha}\left(\dot{\theta}_d - F\right) + K_P e + K_I \int e dt + K_D \frac{de}{dt} \tag{9}$$

where θ_d is a the reference trajectory coming from humanoid stabilizing control and inverse kinematics, $e = \theta_l - \theta_d$ is the tracking error, and K_P , K_I , $K_D \in \mathbb{R}$ are suitable gains.

4 A Control Law Comparison

In order to show the efficiency of the proposed model-free tracking control, 3 different control algorithms will be compared. The joint output shaft will have to track a sinusoidal signal $\theta_d(t)$, whose frequency (5 Hz) is approximately equivalent to a 0.2

$m{\cdot}s^{-1}$ walking motion. Even if support and swinging phases are completely different in terms of loads and joint speeds, a relatively high frequency (15 Hz) and high amplitude ($25Nm$) sinusoidal load $\tau_l(t)$ is considered all along the simulation. As a result, the control robustness to load variations can be evaluated under different situations.

All three strategies will be simulated on the above presented model (equations (1), (6), (7)), with the parameter values summarized in table 1. These data come from a RE35-150W Maxon DC motor, friction parameters suggested in [1], and inertial and stiffness parameters based on approximate structure mass distribution and compliance around critical joints (cf. [18] or [19]). Besides, since CAN bus has a limited bandwidth ($1\ MB{\cdot}s^{-1}$), an approximate symmetric round trip time delay of 4 ms is considered (i.e $T_{sc} = 2$ ms, $T_{ca} = 2$ ms). Additionally, the computer cycle time to generate joint set-points is fixed at 2 ms. Finally, the absolute encoder is supposed to provide a perturbed signal, where the perturbation is modeled as an additive white Gaussian noise whose probability distribution is $\mathcal{N}(0,\sigma)$, $\sigma = 5 \cdot 10^{-5}$

Table 1. Nominal values of the system

Parameter	Value
Motor terminal resistance (R_m)	4.75 Ω
Terminal inductance (L_m)	$1.29 \cdot 10^{-3}$ H
Back EMF Constant (K_b)	0.0235 Vs/Rad
Torque Constant (K_t)	0.0758 Nm/A
Rotor Inertia (J_m)	$6.52 \cdot 10^{-6}$ Kgm^2
Joint stiffness (c)	200 Nm/rad
Global Coulomb friction (C_t)	0.203 Nm
Difference of Coulomb friction back on driving axis (ΔC_t)	-0.0025 Nm
Joint inertia (J_t)	0.3 Kgm^2
Global viscous friction (f_t)	$6.5 \cdot 10^{-3}$ Nms/rad
Transmission ratio (N)	100

Standard PID control. The simplest control strategy that can be imagined in a robotic system is the one that control each joint axis as a SISO system in a decentralized way. Thus, coupling effects between joints due to varying configurations during motion are treated as disturbance inputs. Therefore, the structure of any controller should achieve an effective rejection of different disturbances appearing on the output shaft. Such constraint suggest the use of, on the one hand, an efficient feedforward control, and on the other hand, an integral action over the tracking error.

In this connection, [20] proposes the following position based controller[2]:

$$u(t) = u_{ff}(t) + K_{P_i} e_{\theta_m}(t) + K_{I_i} \int e_{\theta_m}(t)dt + K_{D_i} \frac{de_{\theta_m}}{dt}, \tag{10}$$

[2] This approach is equivalent to a position-velocity feedback (cf.[20] for further details) which is usually implemented on DC motor drivers.

where $u_{ff}(t) = K_b\dot{\theta}_d(t - T_{ca}) + \dfrac{R_m J_m}{K_t}\ddot{\theta}_d(t - T_{ca})$ is the feedforward control, $e_{\theta_m}(t) = \theta_m(t - T_{ca}) - \theta_d(t - T_{ca})$ the angular position tracking error and $K_{P_i} = R_m$, $K_{I_i} = \omega_n^2$ and $K_{D_i} = 2\eta\omega_n$ suitable PID gains. Note that these control parameters are optimally chosen (cf. [20]) in terms of the inner loop natural frequency ω_n and damping ratio η.

Remark 2. All continuous-time operators $\int dt$ and $\dfrac{d}{dt}$ are implemented on discrete time using its respective z-transforms.

High Level PID (HL-PID) control. An outer PID control loop will generate appropriate set points for inner loop described above with absolute encoder information. Thus, the system governed by equations (1), (6), (7) and (10) can be controlled by modifying the position reference, or in other words, by adding a new control variable $v(t)$. This outer control variable will have in this case the classical PID form

$$v(t) = K_{P_o} e_{\theta_l}(t) + K_{I_o}\int e_{\theta_l}(t)dt + K_{D_o}\frac{de_{\theta_l}(t)}{dt},\ e_{\theta_l}(t) = \theta_l(t - T_{sc} - T_{ca}) - \theta_d(t - T_{ca})$$

Model free control. HL-PID control will be here slightly modified to obtain an i-PID as in (9):

$$v(t_k) = \frac{1}{\alpha}\left(\dot{\theta}_d(t_k) - F(t_k)\right) + K_P e_{\theta_l}(t_k) + K_I\int e_{\theta_l}(t_k)dt + K_D\frac{de_{\theta_l}(t_k)}{dt},$$
$$F(t_k) = \hat{\theta}_l(t_k) - \alpha v(t_{k-1})$$

Simulation results. Control parameters have been chosen with the aim of minimizing the mean tracking error. To achieve such a task, the Matlab function *fmincon* (based on Sequential Quadratic Programming) has been applied to each control scheme.

Figure 4 shows simulation results for the above 3 approaches, from which several conclusions can be highlighted:

- Tracking quality of model-free algebraic control (i-PID) is considerably better than with the two other approaches.
- High level PID (HL-PID) control and no absolute encoder based control provide very similar tracking behavior (in terms of maximum and mean error). This surprising phenomenon (HL-PID should a priori improve tracking performance) is mainly due to the feedback network delays (which do not exist in standard PID control). On the contrary, i-PID satisfactory compensates these delays thanks to the ultra-local black-box identification.
- As a result of the previous statement, control action is much more jerky with HL-PID than with standard PID. In any case, i-PID always provide softer control actions than the 2 other techniques.

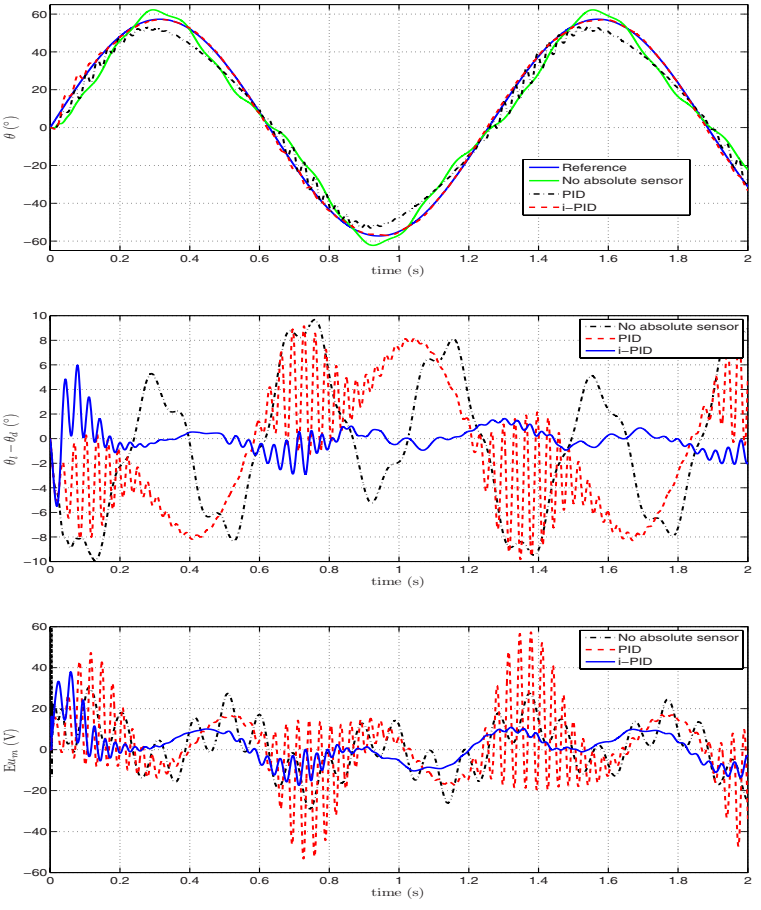


Fig. 4. (a) Position comparison (b) Position error comparison (c) Control action comparison

5 Concluding Remarks

A new approach to precisely control humanoid robot joint has been presented. It is based on model-free algebraic control techniques, which are able to compensate not only transmission and compliance nonlinear effects, but also CAN bus transmission delays. The promising preliminary results shown in this paper will be pursued, on the one hand, to develop a systematic tuning procedure for i-PID and, on the other hand, to deeply analyze robustness to delays and friction/flexion parameters. In that respect, friction losses due the belt-pulley system will have to be considered in this study. Furthermore, since some undesired transient responses still remains around zero-velocity zones, an eventual feedforward friction model will be studied. Finally, an experimental validation with more realistic trajectories and loads will be soon implemented.

References

1. Abba, G., Sardain, P.: Modeling of frictions in the transmission elements of a robot axis for its identification. In: Proc. of the 16th IFAC World Congress, Prague (2005)
2. Chang, Y.C.: Adaptive tracking control for electrically-driven robots without over-parametrization. Int. J. of Adaptive Control and Signal Process 16(2), 123–150 (2002)
3. Chen, B.S., Uang, H.J., Tseng, C.S.: Robust tracking enhancement of robot systems including motor dynamics: a fuzzy-based dynamics game approach. IEEE Trans. on Fuzzy Systems 6(4), 538–552 (1998)
4. Chien, M.-C., Huang, A.-C.: Adaptive Control for Flexible-Joint Electrically Driven Robot With Time-Varying Uncertainties. IEEE Trans. on Ind. Electronics 54(2), 1032–1038 (2007)
5. Dhaouadi, R., Ghorbel, F.H., Gandhi, P.S.: A New Dynamic Model of Hysteresis in Harmonic Drives. IEEE Trans. on Ind. Electronics 50(6), 1165–1171 (2003)
6. Dwidey, K.D., Eberhard, P.: Dynamic analysis of flexible manipulators, a literature review. Mechanism and Machine Theory 41, 749–777 (2006)
7. Fliess, M., Join, C., Sira-Ramirez, H.: Complex continuous nonlinear systems: their black box identification and their control. In: Proc. of 14th IFAC Symp. on System Identification, Newcastle, Australia (2006)
8. Fliess, M., Join, C.: Intelligent PID Controllers. In: Proc of 16th Mediterrean Conf. on Control and Automation, Ajaccio, France (2008)
9. Ghorbel, F., Hung, J.Y., Spong, M.W.: Adaptive control of flexible-joint manipulators. Control Systems Magazine 9, 9–13 (1989)
10. Ishii, C., Shen, T., Qu, Z.: Lyapunov recursive design of robot adaptive tracking control with L2-gain performance for electrically-driven robot manipulators. International Journal of Control 74(8), 811–828 (2001)
11. Kaynov, D., Balaguer, C.: Joint Control of a Humanoid Robot. In: Proc. of 7th IEEE-RAS Int. Conf. on Humanoid Robots, Pittsburgh (2007)
12. Kennedy, C.W., Desai, J.P.: Modeling and Control of the Mitsubishi PA-10 Robot Arm Harmonic Drive System. IEEE/ASME Trans. on Mechatronics 10(3), 263–274 (2005)
13. Kwan, C., Lewis, F.L., Dawson, D.M.: Robust neural-network control of rigid-link electrically driven robots. IEEE Trans. on Neural Networks 9(4), 581–588 (1998)
14. Lee, K.J., Yim, H.J., Jang, S., Kang, Y.S., You, Y.K., Park, T.W.: A Study on Joint Compliance for a Biped Robot. In: Proc. of 5th IEEE-RAS Int. Conf. on Humanoid Robots, Tsukuba (2005)
15. Mboup, M., Join, C., Fliess, M.: Numerical differentiation with annihilators in noisy environment. Numerical Algorithms 50, 439–467 (2009)
16. Ozgoli, S., Taghirad, H.D.: A survey on the control of flexible joint robots. Asian Journal of Control 8(4), 1–15 (2006)
17. Oya, M., Su, C.Y., Kobayashi, T.: State observer-based robust control scheme for electrically driven robot manipulators. IEEE Trans. on Robotics 20(4), 796–804 (2004)
18. Pabon, L., Perez Martinez, C., Villagra, J., Balaguer, C.: Mechatronic design and control of a critical biped robot joint. In: Proc. of IEEE Int. Conf. on Mechatronics, Malaga, Spain (2009)
19. Pérez, C., Pierro, P., Martínez de la Casa, S., Pabón, L., Arbulú, M., Balaguer, C.: RH-2 an Upgraded Full-size Humanoid Platform. In: Proc. of the 12th CLAWAR, Instanbul, Turkey (2009)
20. Siciliano, B., Sciavicco, L., Villani, L., Oriolo, G.: Robotics: modelling, planning and control. Springer, Heidelberg (2009)
21. Su, C.Y., Stepanenko, Y.: Redesign of hybrid adaptive/robust motion control of rigid-link electrically-driven robot manipulators. Trans. on Robotics and Automat. 14(4), 651–655 (1998)
22. Tarn, T.J., Bejczy, A.K., Yun, X., Li, Z.: Effect of motor dynamics on nonlinear feedback robot arm control. IEEE Transaction on Robotics aand Automation 7(1), 114–122 (1991)
23. Tjahjowidodo, T., Al-Bender, F., Van Brussel, H., Symens, W.: Friction characterization and compensation in electromechanical systems. J. of Sound and Vibration 308(3), 632–646 (2007)

Modeling and Impedance Control of a Chewing Robot with a 6RSS Parallel Mechanism

L. Huang, W.L. Xu, J. Torrance, and J.E. Bronlund

School of Engineering and Advanced Technology, Massey University
Private Box 756, Wellington, New Zealand
l.huang@massey.ac.nz

Abstract. Chewing robots are designed to mimic human mastication process . Dynamic force and position control is needed for the robot to produce the chewing force and the trajectory typical for the foods being chewed. The controller design is challenging due to the complexity of the dynamic model of the robot which is normally in a parallel structure. In this paper, a simplified joint space based impedance control scheme is proposed for a 6RSS chewing robot. The special features of the kinematic, force and dynamic models of the robot are explored for the controller design. The effectiveness of the proposed approach is proved.

1 Introduction

To assess a food's property, it has to be chewed before any analysis and conclusion can be made. It is laborious, tedious and time consuming for a person to do that [1][2]. Several chewing robots which are able to mimic the human mastication behavior have been developed. One such a robot named *WJ-3* is presented in [3]. It has nine artificial muscle actuators combined with various position, speed and force sensors. It can achieve three degrees of freedom motion and force control at each actuator. It cannot move in six degrees of freedom as found in a normal chewing process. The system looks clumsy and the controller is complex. Another chewing robot named *WY-5R* is reported in [4]. It has six degrees of freedom with only two of which being actively controlled. The robot is driven by linear actuators. The development of a similar robot named *JSN/2A* is presented in [4]. Though it can generate the chewing movement for simulation purpose, which though is not implemented on a real food. A complete position /force control is not implemented in the above robot systems.

A more dexterous and versatile 6RSS chewing robot has been reported in [5][6]. With a 6 RSS parallel mechanism, the robot has six degrees of freedom and is able to generate chewing trajectories required by a real food chewing process. This paper is intended to address the development of the position and force control scheme for that robot. There are three main methods for the position and force control of a robot, namely *constrained robot control* [7] [8] [9][10][11][12], *hybrid position/force control* [13] [14] [15] and *impedance control* [16] [17]. These methods were mainly developed for serial robotic manipulators, and cannot be

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 733–743, 2009.

directly applied in a chewing robot with parallel structure. In this paper, the special features of the chewing robot concerned make it possible possible to design the controller in the robot's joint space, and thus greatly simplifies the controller design. Considering the contact and non-contact movements of the robot teeth, impedance control scheme is adopted.

The rest of paper is organized as follows. In Section 2, the kinematic, force and dynamic model of the chewing robot concerned is given. In Section 3, an impedance controller is presented. A conclusion of the work is given in Section 4.

2 Kinematic and Force Model

This section deals with the position, velocity and force relations between the driving units and the mandible during the robot's chewing process. The picture of the robot and its schematic are shown in Figures 1 and 2 respectively.

The motion of the robotic jaw is generated by six RSS type linkages, each of which consists of a driver unit (D_i), a crank (C_i) and a coupler link (L_i) ($i = 1, 2, \ldots 6$). The crank is connected to the ground and the coupler through a revolute joint (R) and a ball-socket joint (S) respectively. The coupler is linked to the mandible by another ball-socket joint (S). The fixed part at which the driver unit is installed is called *skull* or *ground*.

A simplified schematic of the chewing robot is redrawn in Figure 3 after taking out physical details of the mechanism. For the ith linkage, its joints are denoted by $g_i(R)$, $s_i(S)$ and $m_i(S)$ respectively. Several Cartesian coordinate frames at the mandible ($\{M\} := O_m X_m Y_m Z_m$), the skull ($\{S\} := O_s X_s Y_s Z_s$) and the joint g_i of driving unit ($\{D_i\} := g_i X_{Di} Y_{Di} Z_{Di}$) are set up respectively for the

Fig. 1. Chewing robot [18]

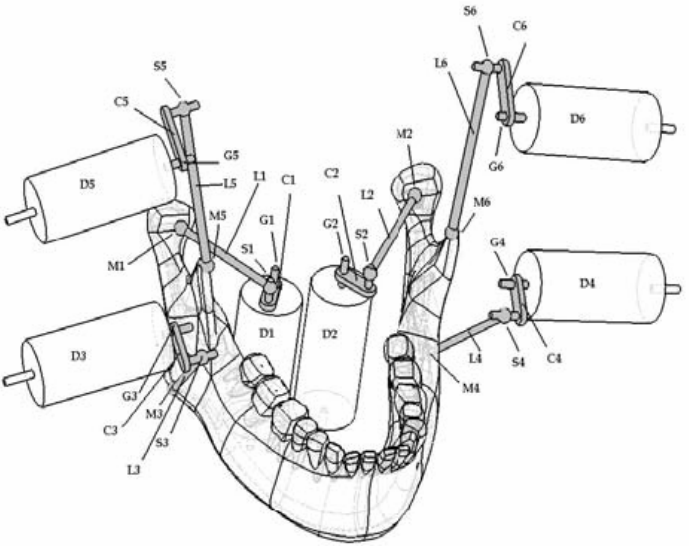


Fig. 2. Schematic of the chewing robot [18]

system description. The origin of $\{M\}$, O_m, is at the center of mass of the mandible and the origin of $\{S\}$, O_s, can be chosen at any suitable point of the skull. For the frame $\{D_i\}$, the axis z_{Di} is along the rotational axis of the driving unit, X_{Di} is along the vector from g_i to s_i and $Y_{Di} = Z_{Di} \times X_{Di}$. The positions of the joints with respect to $\{S\}$ are denoted by r_{gi}, r_{si} and r_{mi} $(i = 1, 2, \ldots 6)$ respectively. The joint variable (rotational displacement) of the ith driving unit is denoted by θ_i. The resulting force caused by chewing is denoted by f_m acting in the frame $O_m X_m Y_m Z_m$, and the accompanied moment is denoted by τ_m. The constraint torque generated at the ith driving unit is denoted by τ_{ci}.

Note in the following, the vector from a point a to a point b is denoted as $\overline{ab}$. The reference frame in which it is described is indicated by a superscript at its top left. For example, ${}^M\overline{O_m m_i}$ means a vector from O_m to m_i described in the frame M. For simplicity and clarification, there is no superscript for a vector described in the fixed frame S.

2.1 Position Relations

The position (linear and angular) of the mandible in reference to S is represented by its rotation matrix ${}^S R_M$, and a positional vector r at its center of mass O_m. The matrix ${}^S R_M$ consists of three unit vectors of the frame $\{M\}$: u_x, u_y and u_z

$$ {}^S R_M = [u_x^T \; u_y^T \; u_z^T]^T, \;\; \|u_x\| = \|u_y\| = \|u_z\| = 1, \;\; u_x^T u_y = u_x^T u_z^T = u_y^T u_z = 0 \quad (1) $$

It is obvious that

$$ r_{mi} = r + {}^S R_M(\theta_i) \; {}^M\overline{O_m m_i}, i = 1, 2, \ldots 6 \quad (2) $$

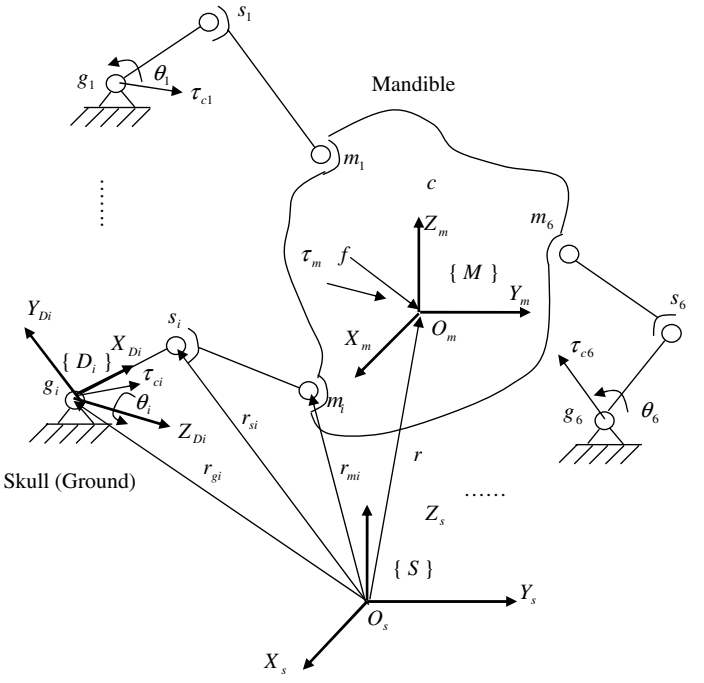


Fig. 3. Schematic for modeling of the chewing robot

$$r_{si} = r_{gi} + {}^{S}R_{Di}\ {}^{D_i}\overline{g_i s_i} \tag{3}$$

$$\overline{s_i m_i} \quad = r_{mi} - r_{si} \tag{4}$$

where ${}^{S}R_{Di}(\theta_i)$ is the orientation matrix of the frame $\{D_i\}$ in reference to $\{S\}$. It is the function of θ_i. The vectors r_{gi}, ${}^{M}\overline{O_m m_i}$ and ${}^{D_i}\overline{g_i s_i}$ are determined by the design parameters of the robot mechanism and are constant in the frames in which they are described.

For simplicity without ambiguities, the dependence of one variable on other variables is not shown in its expression if this is clear in the context of the presentation. For example, ${}^{S}R_M(\theta_i)$ will be simply expressed as ${}^{S}R_M$.

Due to the constraint

$$\|\overline{s_i m_i}\|^2 = constant \tag{5}$$

we have

$$\begin{aligned} &\|r + {}^{S}R_M\ {}^{M}\overline{O_m m_i} - r_{gi} - {}^{S}R_{Di}\ {}^{D_i}\overline{g_i s_i}\|^2 \\ &= (r + {}^{S}R_M\ {}^{M}\overline{O_m m_i} - r_{gi} - {}^{S}R_{Di}\ {}^{D_i}\overline{g_i s_i})^T (r + {}^{S}R_M\ {}^{M}\overline{O_m m_i} - r_{gi} - {}^{S}R_{Di}\ {}^{D_i}\overline{g_i s_i}), i = 1, \ldots 6 \end{aligned} \tag{6}$$

Given the posture of the mandible, r and ${}^{S}R_M$, six joint variable θ_i $(i = 1, \ldots 6)$ are solved directly from the six quadratic equations (6).

It is much tougher to find r and ${}^{S}R_M$ from θ_i $(i = 1, 2, \ldots 6)$. This is a typical direct kinematic problem of a parallel robot which has no unique closed form solution. There are a lot of numerical methods to tackle this problem [19].

2.2 Velocity Relation

To get the velocity relation of the robot, differentiate equation (6) with respect to time t,

$$\overline{s_i m_i}^T(\dot{r} + {}^{S}\dot{R}_M \ {}^{M}\overline{O_m m_i} - C_i \ {}^{D_i}\overline{g_i s_i}\dot{\theta}_i) = 0, i = 1, 2, \ldots 6 \tag{7}$$

where

$$C_i = \frac{\partial {}^{S}R_{D_i}}{\partial \theta_i}$$

As ${}^{S}R_M$ is rotation matrix, its derivative with time t is [20] (pp. 71)

$${}^{S}\dot{R}_M = [\omega\times] \ {}^{S}R_M \tag{8}$$

where $\omega = [w_x \ w_y \ \omega_z]^T$ is the angular velocity of the frame $\{M\}$, and $[\omega\times]$ is a skew symmetric matrix defined as

$$[\omega\times] \triangleq \begin{bmatrix} 0 & -\omega_z & \omega_y \\ w_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix} \tag{9}$$

With definitions in equations (9) and (9) and noting that

$${}^{S}R_M \ {}^{M}\overline{O_m m_i} = \overline{O_m m_i},$$

equation (7) is expanded and simplified in a more compact form,

$$\overline{s_i m_i}^T(\dot{r} - [\overline{O_m m_i}\times]\omega - C_i \ {}^{D_i}\overline{g_i s_i}\dot{\theta}_i) = 0, i = 1, 2, \ldots 6 \tag{10}$$

Defining

$$\begin{aligned} B_i &= [I^{3\times 3} \ -[\overline{O_m m_i}\times]^T]^T \\ \theta &= [\theta_1 \ \theta_2 \ldots \theta_6]^T \\ \dot{x} &= [\dot{r}^T \ \omega^T]^T \end{aligned}$$

equation (10) which includes six equations for all the linkage, is rewritten as

$$J_x \dot{x} = J_\theta \dot{\theta} \tag{11}$$

where

$$J_x = \begin{bmatrix} \overline{s_1 m_1}^T B_1 \\ \vdots \\ \overline{s_i m_i}^T B_i \\ \vdots \\ \overline{s_6 m_6}^T B_6 \end{bmatrix}, \ J_\theta = diag(\overline{s_i m_i}^T C_i \ {}^{D_i}\overline{g_i s_i}), i = 1, 2, \ldots 6 \tag{12}$$

are called *Jacobians* mapping the velocity input $\dot{\theta}$ to the linear and angular velocities $\dot{x}$ of the mandible. In the above, J_θ is expressed in a form $diag(x_i)$ which is a diagonal matrix consisting of the elements x_i.

From equation (11), we have

$$\dot{\theta} = J\dot{x} \tag{13}$$

where

$$J = J_\theta^{-1} J_x = diag(\frac{1}{\overline{s_i m_i}^T C_i \ ^{D_i}\overline{g_i s_i}}) J_x \tag{14}$$

is called an overall Jacobian.

From the definition of the definition of J_x (equation (11)) and B_i (equation(10)), J can be explicitly expressed as the function of the geometric parameters of the system.

$$J = diag(\frac{1}{\overline{s_i m_i}^T C_i \ ^{D_i}\overline{g_i s_i}}) \begin{bmatrix} \overline{s_1 m_1}^T & -(\overline{O_m m_1} \times \overline{s_1 m_1})^T \\ & \vdots \\ \overline{s_i m_i}^T & -(\overline{O_m m_i} \times \overline{s_i m_i})^T \\ & \vdots \\ \overline{s_6 m_6}^T & -(\overline{O_m m_6} \times \overline{s_6 m_6})^T \end{bmatrix} \tag{15}$$

The validness of the velocity mappings in equations (11) and (13) requires that J_θ and J_x are of full rank. This is equivalent to

- $[\overline{s_i m_i}^T \quad -(\overline{O_m m_i} \times \overline{s_i m_i})^T]^T$ are linear independent, and
- $\overline{s_i m_i}^T C_i \ ^{D_i}\overline{g_i s_i} \neq 0$, $i, j = 1, 2, \ldots 6$

or , more explicitly,

- both $\overline{s_i m_i}$ and $\overline{O_m m_i} \times \overline{s_i m_i}$ are linear independent, and
- $\overline{s_i m_i}$ is not perpendicular to $C_i \ ^{D_i}\overline{g_i s_i}$, $i = 1, 2, \ldots 6$

If J_θ and J_x are not of full rank, there are following implications [21],

- when $rank(J_q) < 6$, the robot is in a so called inverse kinematic singular configuration where some directions cannot be reached by the mandible, and thus the degrees of freedom are lost;
- when $rank(J_x) < 6$, the robot is in a so called direct kinematic singular configuration where some directions can be achieved without any inputs of the driving units, and thus the degrees of freedom are gained.

In the following discussion, it is assumed that the full rankness of J_θ and J_x are kept by proper robot structure design and trajectory planning. In this case, given the desired position and velocity of the mandible, all the terms in equation (13) are well defined and the required joint velocity inputs $\dot{\theta}$ can be found.

2.3 Force Relation

Let τ_{ci} is the torque of the driving unit of ith linkage required to contribute to balance the force /moment caused by chewing the foods, and define

$$\tau_c = [\tau_{c1}\ \tau_{c2}\ \ldots \tau_{c6}] \tag{16}$$

Let δx and δq are the virtual displacements corresponding to $\dot{x}$ and $\dot{q}$ respectively, and define

$$f = [f_m^T\ \tau_m^T]^T$$

which represent the chewing force acting at the center of the mass of the mandible.

From the principle of virtual work

$$\tau_c^T \delta q = f^T \delta x$$

As $\delta q = J\delta x$, we have

$$(\tau_c^T J - f^T)\delta x = 0$$

This relation holds for any δ_x, and accordingly

$$f = J^T \tau_c \tag{17}$$

From equation (17), f is obtained indirectly through measuring τ_c, or the input currents of the driving units which are proportional to τ_c. This fact is very useful to overcome the difficulties in the direct measurement of f. For example, it is not easy to install the force sensor in the narrow space between the mandible and the skull, or the readings from the force sensor is very noisy.

Assuming that J is a full rank, we have

$$\tau_c = J^{-T} f \tag{18}$$

2.4 Dynamic Model

The chewing robot is an example of generic robotic manipulators whose complete dynamic model has been well studied in 1980s[22][23][24]. Those models are derived by considering the couplings among all the elements in the system, and are highly nonlinear and coupled. For a robot system with a high degree of freedom, they are too computationally intensive to be used in real time.

Considering that the mass of the linkage and crank are negligible compared to that of the mandible and a high gear ratio gear box is used to link the driving unit to the crank , the dynamics of the driving unit where the control input is applied can be decoupled from that of the rest of the system. As a result, the system dynamic model can be greatly simplified.

Assume the effective moment of inertia of the shaft of the ith driving unit with respect to the axis Z_{Di} is I_i and the zero friction force at the joint, the dynamics of the ith driving unit can be simplified to

$$I_i \ddot{\theta}_i + h_i(\theta, \dot{\theta}) + \tau_{ci} = \tau_i, \quad i = 1, 2 \ldots 6 \tag{19}$$

where I_i is the moment of inertia of the motor, $h_i(\theta, \dot{\theta})$ is the term related to the centrifugal force, Corilos force, gravitational force and the force due to the contact between the linkage and the mandible, τ_i is the driving torque and τ_{ci} is the torque generated by the chewing force f which has been defined in equation (16).

Combining the dynamics models of all the driving units together, we have

$$I_R\ddot{\theta} + h_R(\theta, \dot{\theta}) + J^{-T}f = \tau \tag{20}$$

where $I_R = diag(I_1\ I_2 \ldots I_6)$, $h_R = diag(h_1\ h_2 \ldots h_6)$ and $\tau = [\tau_1\ \tau_2 \ldots \tau_6]^T$.

From equations (13) and (20), we have the dynamic model of the driving unit as a function of the states of the mandible

$$I_R J\ddot{x} + I_R\dot{J}\dot{x} + J^{-T}f = \tau \tag{21}$$

which is described in the Cartesian coordinate frame $S : O_sX_sY_sZ_s$.

The dynamics of the mandible is described by

$$M\ddot{x} + h_m(x, \dot{x}) = f, \quad i = 1, 2 \ldots 6 \tag{22}$$

where M is the inertia of the mandible, $h_m(x, \dot{x})$ is the term related to the centrifugal force, Corilos force, gravitational force and the force acting from the linkage on the mandible. As it is decoupled from the dynamics of the driving unit, it will not be used directly in the controller design

The dynamic equations (20) and (21) show that the driving torque τ not only produces the motion of the robot, but also overcomes the torque caused by the chewing force f. Through Jacobian J^T which is the function of the configuration of the whole robot system, f is mapped to the joint space of each linkage. Obviously, this mapping is not decoupled from one linkage to another, though their dynamic models are decoupled.

3 Controller Design

Before designing the controller, the robot's desired position and force trajectories: θ_d, f_d and $\tau_{cd} = J^{-T}f_d$ are obtained from the measurement from a real chewing process. From the desired position of the robot, its desired velocity and acceleration $\dot{\theta}_d$ and $\ddot{\theta}_d$ are also obtained.

We take the same 2nd order impedance model as that in [25], that is

$$e_{\tau_c} = M\ddot{e}_\theta + D\dot{e}_\theta + Ke_\theta \tag{23}$$

where $e_{\tau_c} = \tau_{cd} - \tau_c = J^{-T}e_f$, $e_f = f_d - f$, $\ddot{e}_\theta = \ddot{\theta}_d - \ddot{\theta}$, $\dot{e}_\theta = \dot{\theta}_d - \dot{\theta}$, $e_\theta = \theta_d - \theta$ and $e_{\tau_c} = \tau_{cd} - \tau_c$. Impedance parameters $M > 0$, $D > 0$ and $K > 0$ are all constant positive definite (p.d.) matrices . Our control objective is to achieve the desired impedance as defined in equation (23).

Consider the following control input

$$\tau = I_R(\ddot{\theta}_d + M^{-1}(D\dot{e}_\theta + Ke_\theta - e_{\tau_c})) + h_R(\theta, \dot{\theta}) + \tau_c \tag{24}$$

Substituting τ into equation (20) gives

$$MI_R(M\ddot{e}_\theta + D\dot{e}_\theta + Ke_\theta - e_{\tau_c}) = 0 \tag{25}$$

As M and I_R are positive definite, it follows that

$$M\ddot{e}_\theta + D\dot{e}_\theta + Ke_\theta = e_{\tau_c} = J^{-T}e_f \tag{26}$$

which is the desired impedance.

Remarks

– The chewing force mapped into the robot's joint space, τ_c, is obtained through measuring the chewing force described in the coordinate frame of the mandible. In practice, the force sensor is normally installed at a location near or at some distance away from the chewing point. A space transformation from the sensor frame to the mandible coordinate frame is needed to get the force f from the sensor reading f_s. If this transformation is represented by the matrix ${}^m_sD$, then f and f_s is related by

$$f = {}^m_sDf_s$$

– If the chewing force cannot be measured directly and the actuators of the robot are DC motors, an alternative approach is to get the chewing force from the readings of the currents of the motors. This method is based on the fact that the motor torque which include the part to overcome the chewing force is proportional to the motor current, that is, $\tau_i = k_t i$, where k_t is the torque constant of the motor and i is the motor current. From equation (19), τ_{ci} can be derived as

$$\tau_{ci} = k_t i - I_i\ddot{\theta}_i - h_i(\theta, \dot{\theta})$$

In addition to the motor current, position and the velocity, the acceleration of the motor is also needed to get τ_{ci}. It can be measured with an accelerometer or is ignored if its magnitude is small.

4 Conclusion

This paper presents the modeling and impedance control of a chewing robot with 6RSS mechanism. The robot's kinematic, force and dynamic models are established by considering its special features. A joint space based impedance control scheme is then developed to achieve the position /force control for the robot in the process of chewing foods or in a free motion. Simulation and experimental study are undergoing to implement the controller.

References

1. Wilinson, C., Dijksterhuis, G.B., Minekus, M.: From food structure to texture. Trends in Food Science & Technology 11, 442–450 (2000)
2. Cardello, A.V.: Sensory-Instrument Research. Cereal Foods Works 39 (1994)

3. Takanobu, H., Yajima, T., Nakazawa, M.: Quantification of Masticatory Efficiency with a Mastication Robot. In: Proceedings of the 1998 IEEE Internal Conference on Robotics & Automation (1998)
4. Takanobu, H., Soyama, R., Takanishi, A., Ohtsuki, K., Ozawa, D., Ohnishi, M.: Remote Treatment with Mouth Opening and Closing Training Robot. IEEE Intelligent Robots and Systems 1, 553–558 (2000)
5. Pap, J.S.: A Chewing Robot Based on Parallel Mechanism- Analyse and Design, Master of Engineering Thesis, Massey University, New Zealand (2006)
6. Xu, W.L., Torrance, J.D., Chen, B.Q., Potgieter, J., Bronlund, J.E.: Kinematics and Experiments of a Life-Sized Masticatory Robot for Characterizing Food Texture. IEEE Transactions on Industrial Electronics 55, 2121–2132 (2008)
7. McClamroch, N.H., Wang, D.W.: Feedback stabilization and tracking of constrained robots. IEEE Trans. on Auto. Contr. 33(5), 419–426 (1988)
8. Mills, J.K., Goldenberg, A.A.: Force and Position Control of Manipulators During Constrained Motion Tasks. IEEE Trans. on Robotics and Auto 5(1), 30–46 (1989)
9. Kankaanranta, R.K., Koivo, H.N.: Dynamics and Simulation of Compliant Motion of a Manipulator. IEEE J. of Robotics and Auto 4(2), 163–173 (1988)
10. Ge, S.S., Huang, L., Lee, T.H.: Model-based and neural-network-based adaptive control of two robotic arms manipulating an object with relative motion. Int. J. of Systems Sci. 32(1), 9–23 (2001)
11. Huang, L., Ge, S.S., Lee, T.H.: Position/force control of uncertain constrained flexible joint robots. Mechatronics (16), 111–120 (2006)
12. Liu, G., Li, Z.: A unified geometric approach to modeling and control of constrained mechanical systems. IEEE Trans. on robotics and Auto 18(4), 574–587 (2002)
13. Yoshikaw, T.: Dynamic hybrid position/force control of robot manipulators - description of hand constraints and calculation of joint driving force. IEEE Journal of Robotics and Auto RA-3(5), 386–392 (1987)
14. Hsu, F.-Y., Fu, L.-C.: Intelligent Robot Deburring Using Adaptive Fuzzy Hybrid Position/Force Control. IEEE Trans. on Robotics and Automation 16(4), 325–335 (2000)
15. Kiguchi, K., Fukuda, T.: Position/Force Control of Robot Manipulators for Geometrically Unknown Objects Using Fuzzy Neural Networks. IEEE Trans. on Industrial Electronics 47(3), 641–649 (2000)
16. Hogan, N.: Impedance control, an approach to manipulation: Part I–III. Journal of Dynamic Systems, Measurement, and Control 107, 1–24 (1985)
17. Seraji, H., Colbaugh, R.: Force tracking in impedance control. The Int. J of Robotics Research 16(1), 97–117 (1997)
18. Torrance, J.D., Pap, J.-S., Xu, W.L., Bronlund, J.E., Foster, K.D.: Motion Control of a Chewing Robot of 6RSS Parallel. In: The 3rd International Conference on Autonomous Robots and Agents, New Zealand (December 2006)
19. Raghavan, M.: The stewart platform of general geometry has 40 configurations. ASME J. MEch. Des. 115, 277–282
20. Sciavicco, L., Siciliano, B.: Modeling and control of robot manipulators. McGraw-Hill, New York (1996)
21. Tsai, L.-W.: Robot Analysis. John Wiley & Sons, Inc., Chichester (1999)
22. Hollerbach, J.M.: A recursive Lagrangian formulation of manipulator dynamics and a comparative study of dynamics formulation complexity. IEEE Trans. Syst. Man Cybern. SMC-10(11), 730–736 (1980)

23. Luh, J.Y.S., Walker, M.W., Paul, R.P.: On-line computational schemen for mechanical manipualtors. ASME J. Dyn. Syst. Meas. Control 120, 69–76 (1980)
24. Lee, C.S.G., Lee, B.H., Nigam, R.: Development of the generalized d'Alembert equations of moption of mechnanical manipulators. In: Proc. 22nd Conference on Decision and Control, San Antonio, TX, pp. 1205–1210 (1983)
25. Canudas de Wit, C., Siciliano, B., Bastin, G. (eds.): Theory of Robot Control. Springer, Heidelberg (1996)

Decentralized Control for Swarm Flocking in 3D Space

Xiang Li [1], M. Fikret Ercan[1], and Yu Fai Fung[2]

[1] Singapore Polytechnic, School of Electrical and Electronic Engineering, Singapore
[2] The Hong Kong Polytechnic University, Department of Electrical Engineering, Hong Kong SAR
xl_huse@126.com, mfercan@sp.edu.sg, eeyffung@polyu.edu.hk

Abstract. This paper presents a decentralized control strategy for a robot swarm where each robot tries to form a regular tetrahedron with its three neighbors. The proposed method is based on virtual spring. Robots can form regular tetrahedron regardless of their initial positions and they require minimum amount of information about their neighbors. The control strategy is made scalable by integrating a neighbor selection procedure so that it can be expanded to large swarms easily. In addition, an obstacle avoidance mechanism, based on artificial physics, is also introduced. By utilizing this control strategy, basic swarm behaviors such as aggregation, flocking and obstacle avoidance are demonstrated through simulations in an unknown three dimensional environment.

Keywords: Swarm robotics, Swarm Intelligence, Distributed control.

1 Introduction

Swarm robotics is a new direction in robotics in which coordination of multiple simple robots by using basic local interactions is studied. It is inspired by the swarming animal behavior from nature [1], [2]. Natural swarms such as ants, termites, and bees are capable of complex tasks such as carrying a prey, foraging and flocking. Design of individual robots that can achieve a desired collective behavior is still an open problem. However, there are numerous swarm behaviors widely studied in literature, such as aggregation, flocking, formation, etc. Nowadays, swarm robotics is gaining even more attention since it is becoming possible to utilize them in real world applications such as surveillance, odor localization and search and rescue [2], [3], [4], [5], [6], [7], [8]. For instance, Correll et al. show inspection of turbines in power plants and jets using a swarm of miniature robots that acts as self-actuated sensors [8]. Applications of swarms in complex three dimensional environments (see for instance, swarms of aerial robots in urban environments [9], [10], or submersible robots for underwater explorations [11]) are also currently being studied. Deploying swarm of robots in challenging environments such as underwater has practical value. As these robots perform collectively, a task can be completed in a shorter period of time in a large working area. Furthermore, it is unavoidable to have a robot to break down in harsh and unpredictable environments. However, when a swarm of robots is employed, failure of a few robots will not interrupt the completion of a given task.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 744–754, 2009.

Flocking is a key swarm behavior and realizing it artificially has been an interest in the robotic research community. It is a collective and coordinated movement of individuals in a swarm so that entire swarm maneuvers as one body. In one of the early studies [12], it has been shown that if individuals in an artificial swarm can sense bearing, range and orientation of their neighbors then they can achieve flocking by using three simple behaviors: collision avoidance, velocity matching and flock centering. Subsequently, there are many methods presented to achieve swarm flocking which are inspired from a wide range of fields such as artificial physics to social animal behaviors. The following references provide a good survey of various techniques developed so far [13],[14], [15], [16], [17].

Above mentioned methods may be broadly summarized into two control strategies: centralized and decentralized [18]. In centralized strategy, a leader is responsible for the swarm behavior, usually members of the swarm follows the leader. In decentralized strategy, individuals rely on local rules and act independently which also determines the group behavior. In this paper, we study a decentralized control strategy for a swarm of mobile robots in three-dimensional space to achieve flocking behavior. That is, robots flock towards a target point by avoiding obstacles in the environment as a whole. This work is extended from our earlier study, in which we have presented a decentralized control strategy, named Triangular Formation Algorithm (TFA) [19]. The TFA is based on triangular geometry and developed for swarm robots moving in two-dimensional space. The principle of TFA, which is basically forming and maintaining isosceles among the three neighboring robots, is extended to the forming f regular tetrahedrons among four neighboring robots in a 3D space. Distributed control strategy employs artificial spring to achieve aggregation and flocking behaviors. Robots negotiate the environment and dynamically adapt to a collective behavior. By integrating a neighbor selection procedure, this control strategy is also extended to a large swarm. Aggregation and unconstrained flocking behavior are also achieved for the large swarm. Our empirical study shows that the proposed control strategy delivers the desired results regardless of the initial distribution of the robots.

2 Robot Model and Problem Description

We consider a swarm consisting of n robots which are denoted by $r_1, r_2, \ldots, r_n$ respectively. It is assumed that in practice these robots are equipped with adequate controllers, actuators and sensors, such as digital compass, range sensors, accelerometer so that they can operate autonomously. For the sake of simplicity, they are modeled as dynamic points in 3D space. In our model, it is assumed that each robot can recognize its distance from other robots situated within its sensing range. Hence, a robot determines the magnitude and orientation of its acceleration in each time step based on the distances from its three neighbors. There is no direct communication among the robots. All individuals follow the same control method and their acts are independent and asynchronous.

For any given robot, r, the distance from robot r_i is denoted as d_i and the desired distance from each other is denoted as d_u. For the convenience of mathematical

description, $dist(r_i, r_j)$ denotes the distance between any two robots, r_i and r_j. $area(r_i, r_j, r_k, r_l)$ denotes the surface area of the tetrahedron configuration formed by the four robots, r_i, r_j, r_k, and r_l as shown in Fig.1. The robots locating at the sensing range of robot r make up a neighbor set denoted as N_r. R_s and A_s represent a robot's sensing radius and the corresponding sensing area respectively. RT represents a regular tetrahedron configuration constructed by four robots where all four robots have the same distance d_u from each other. Basic behavior of each individual robot in a swarm is then to maintain distance d_u with the other three specific neighbors. Consequently, the swarm can form a uniform interval network which is made of multiple RT s in space. Firstly, we are interested in the basic behavior, that is, four neighboring robots always form a RT configuration starting from any arbitrary positions. This behavior is also the basic element of the swarm behavior. Apparently, the behavior of each robot is local. Since each robot determines its actions by using local positions of three other robots, the number of neighbors needed for the controller to operate is three.

Uniform 3D flocking control. This swarm behavior enables n robots to move as a whole towards a stationary or moving target while adapting to an environment populated with obstacles. Robots start from an arbitrary initial distribution in 3D space. It is expected that they maintain a uniform distance from each other while flocking towards a target. This uniform flocking behavior in 3D space has many potential applications such as search and rescue, sampling in sea or air, surveillance and so on.

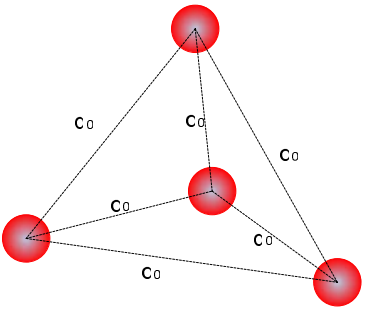


Fig. 1. Regular tetrahedron (RT) configuration desired for four neighboring robots

3 Control Strategy

The basic control principle employed in this study is inspired from virtual physics of spring mesh [20], [21]. However, a major difference and the advantage of our control strategy is that we employ distance information of invariable number of neighbors. This basic control strategy, employed by all the robots in swarm, is formulated in equation 1.

$$\vec{a} = \left[\sum_{i=1,2,3} k_s (d_i - d_u)\vec{n}_i \right] - k_d \vec{v} \quad (1)$$

There are a number of inputs to the controller, such as $d_i (i = 1,2,3)$ represents the distance from three neighboring robots collected by the range sensors; $\vec{v}$ is the velocity of the robot r executing the control strategy; $\vec{n}_i$ is the unit vector from robot r to the corresponding neighboring robot r_i ($i = 1, 2, 3$). On the other hand, acceleration, $\vec{a}$, is the output from controller at the next time instant. The constants employed in the controller are the natural spring length d_u , the spring stiffness k_s , and the damping coefficient k_d . The output of each robot controller relies on distance information from only three specific neighboring robots. The control strategy is local and scalable.

3.1 Basic Behavior and Its Convergence

As mentioned earlier, the basic behavior of the robots is to form a regular tetrahedron starting from arbitrary initial positions as shown in Fig.2. The parameter settings for the simulations were d_u =20, k_s =0.1, and k_d =0.3 and the convergence of the basic behavior is shown in Fig3. Apparently, all the sides were at different length and larger than d_u at the beginning. We can observe that at the beginning of the tuning process, the side lengths of tetrahedron configuration oscillate about several cycles while decreasing in magnitude and finally settling around the equilibrium d_u =20. That is, all the side lengths will eventually converge to the same length of d_u . This result is acceptable as the initial configuration formed by the four robots is a dissipative energy system. This oscillatory pattern takes most of the convergence time. There are some

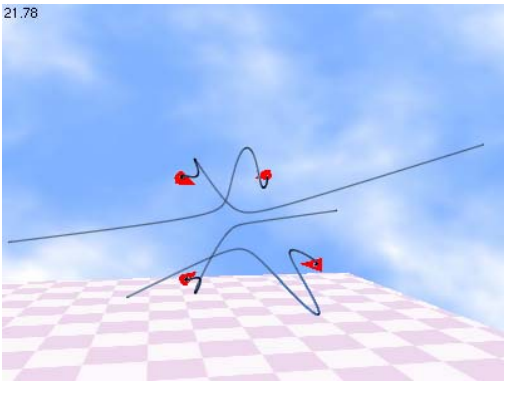


Fig. 2. RT configuration by four

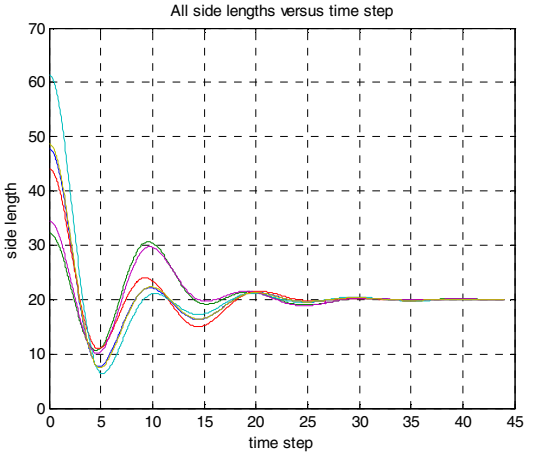


Fig. 3. *RT* convergence curves for all six side lengths (d_u =20)

experiments reported earlier, to observe the effects of k_s and k_d on the convergence time of a seven robot mesh [22]. This report gave us some idea though it does not reflect the case in our study. Test mesh used in [22] is a 2D mesh using acute angle test while ours is a 3D mesh and employing a complete connection topology of four robots. Finding an optimal combination of k_s and k_d for RT formation in 3D space with respect to the initial distribution is going to be the subject of our future work.

4 Control Strategy for Large Swarms

The discrete control strategy presented above basically unites four robots into a RT formation. This strategy is made scalable by including a neighborhood selection procedure. It is expected that robot r will select three neighbors as reference objects even though it may detect more than three neighboring robots within its sensing range. The neighborhood selection for the first, second and third neighbors are given by equations r_1, r_2, and r_3 respectively.

$$r_1 := \arg\left[\min(dist(r, r_i))\right],\ r_i \in N_r \tag{2}$$

$$r_2 := \arg\left[\min\left[dist(r_i, r) + dist(r_i, r_1)\right]\right],\quad r_i \in (N_r - r_1) \tag{3}$$

$$r_3 := \arg\left[\min\left[area(r, r_1, r_2, r_3)\right]\right],\quad r_3 \in (N_r - r_1 - r_2) \tag{4}$$

In the above equations, N_r is the neighborhood set containing all the robots detected by robot r. $N_r - r_1$ and $N_r - r_1 - r_2$ are the neighborhood subsets excluding r_1 and r_1, r_2 respectively. The neighborhood selection mechanism described above is integrated to our control strategy. All the simulation experiments are conducted using BREVE platform which is designed for simulations of decentralized systems and artificial life [23]. Fig 4. shows snapshots of the six robot aggregation behavior. It can be seen from the results that robots form a multiple configuration which has identical distance among individuals.

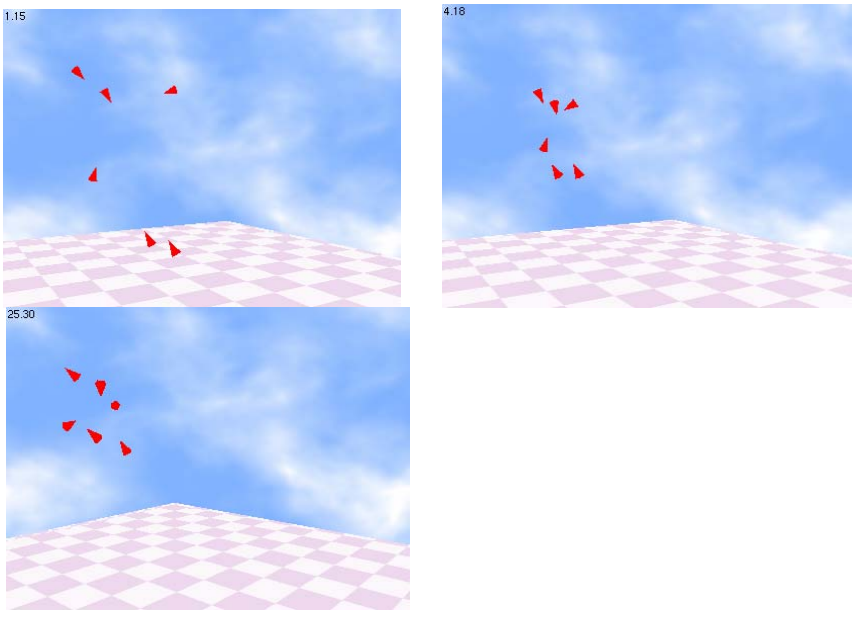


Fig. 4. Aggregation of more than four robots

In our experiments with four, five, six and seven robots, we observed that swarm can always aggregate as a whole. For the swarm consisting of more than seven robots, large d_u and concentrated initial distribution can also enforce swarm to aggregate as a whole. When the initial distribution is relatively scattered, multiple aggregations is most likely to occur. More specifically, the number of sub aggregations that will appear is closely related to the size of the swarm, the initial distribution space and the parameter d_u. However, this might be desired for some applications such as multiple-target search. If d_u and the initial distribution space are fixed, the swarm with larger size tends to produce one aggregation. The swarm will have more chance to form

multiple aggregations for the smaller d_u and the wider distribution space, as an example shown in Fig.6. From the above analysis, we can deduce that the possible number of aggregations will be between 1,…,$\lfloor n/4 \rfloor$, where n is the size of the swarm.

4.1 Obstacle Avoidance

The ability of adaptive flocking in a constrained environment is the basic requirement for a swarm system. In order to avoid obstacles in an environment, we employed a mechanism based on artificial physics to work with the distributed control strategy discussed above. Here, we mimic the interaction between an electron in motion and an atom nucleus as illustrated in Fig 5. As a result, robots maintain RT formation while adapting to an environment with obstacles and flocking towards a target. If a robot approaches an obstacle, a repulsive force f_r will exert on it. A repulsive velocity v_r is the equivalent of the effect of f_r on the robot. The direction of v_r depends on the nearest point of the obstacle and directs from the point to the robot. For simplicity, an obstacle is considered as a closed sphere and has a safety margin of d. The nearest point of the obstacle from the robot is denoted as O. In order to avoid the obstacle, robot r_i will adjust its current velocity by adding a component v_r to the current velocity. If the distance between r_i and O is larger than d, there is no force on r_i. Otherwise, there will be a repulsive force exerting on r_i. In practice, a maximal velocity, $v_{\max}$, can be set for all the robots uniformly. The repulsive velocity is then calculated as:

$$v_r = \frac{p_i - O}{|p_i - O|} r\left(1 - \frac{|p_i - O|}{d}\right) \cdot v_{\max} \tag{5}$$

Where $r(x)$ is defined as a ramp function:

$$r(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases} \tag{6}$$

Equation (5) implies that, in the worst case, the repulsive velocity exerting on a robot linearly increases from 0 to $v_{\max}$ during the course that the robot enters safety margin until it just touches the obstacle. This will cause the robot to stop when it is just touching the obstacle because its total velocity will become zero. In fact, as a member of the swarm, a robot will eventually stay away from the obstacle by following its neighbors even before collision happens. In other words, individuals will possibly avoid the local minima with the help of their neighbors. However, dealing with local minima is a fundamental problem in the potential field methods as it has been studied extensively [16].

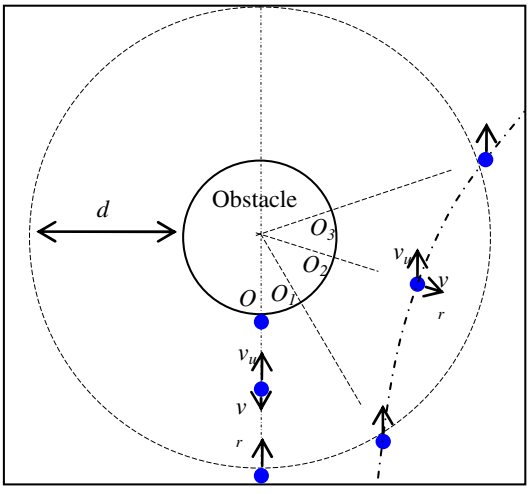


Fig. 5. Obstacle Avoidance mechanism

4.2 Swarm Flocking Behavior

In this study, we try to enable a swarm to move towards a target in a 3D environment. The environment is populated with obstacles as shown in Fig 7. The target is positioned in the middle on the right hand side; however it is not explicit in these pictures due to limitations of graphics. In order to move as a whole, instead of neighborhood set N_r, a modified neighbor set N_{rt} is utilized by each robot. Set N_{rt} consists of the neighbors that are locating at the half space pointing towards the target side. Only the robots, who have detected the target, will have flocking vectors. The magnitude of the vector is a constant and the direction of it points towards the target. Other robots near by, due to their local control strategy to maintain RT with their neighbors, will follow these robots and move together with them.

During the experiments, we found that, if the size of an obstacle is larger than d_u, the swarm has a tendency to move around the obstacle as a whole. Otherwise, the swarm most likely will flock towards the target while avoiding the obstacle along the way. When a large swarm is flocking towards a target in an environment filled with multiple obstacles, if the size of the obstacles is larger than d_u, the swarm will possibly adapt its shape to fit the gap between the obstacles and pass through. If the size of the obstacle is much smaller than the distance d_u, the swarm will maintain its shape and directly flock through the obstacles. The control framework relies on the fact that a robot can identify the positions of other robots with respect to local coordinates and distinguish them from obstacles. As calculations involve only three other neighbors' position, the control strategy has less computation and becomes less influenced by other robots.

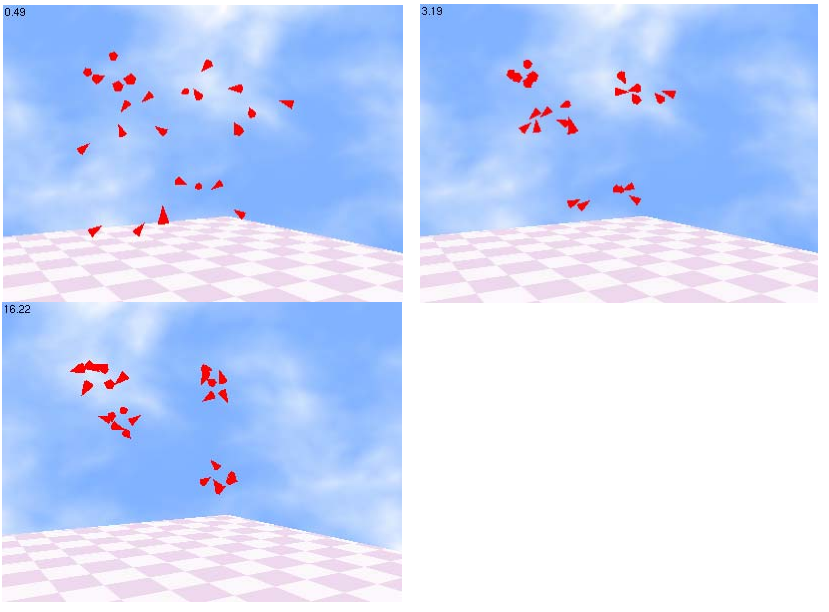


Fig. 6. Aggregation of twenty four robots: smaller d_u and wide initial space

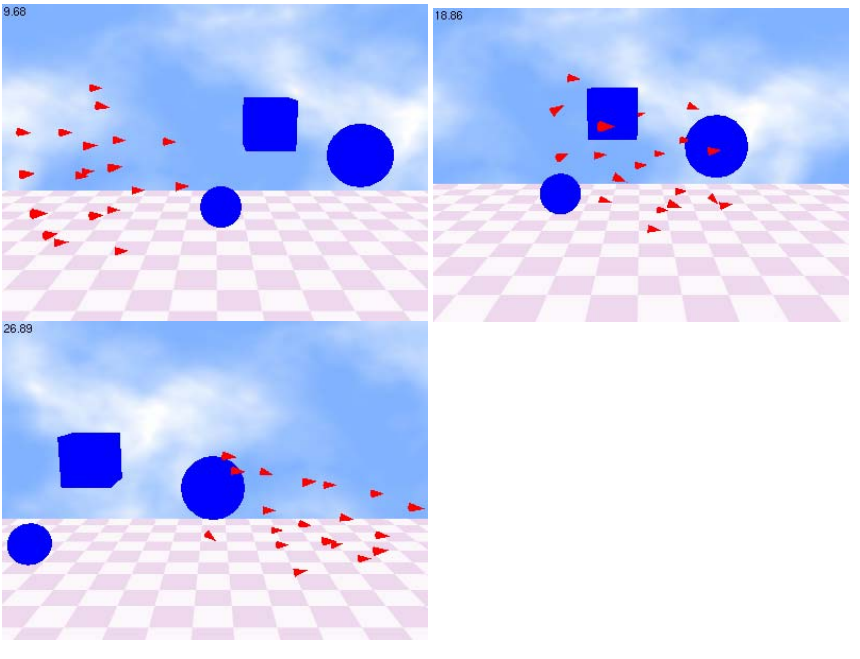


Fig. 7. Constrained swarm flocking

5 Conclusions

This paper presents a distributed control strategy based on artificial spring for a swarm of robots to demonstrate aggregation and flocking behaviors. Robots negotiate the environment and dynamically adapt a collective behavior. The basic control strategy enables four neighboring robots to form a RT configuration regardless of their initial positions. By integrating a neighbor selection procedure, this control method is extended to a large swarm. There are only two requirements for this control strategy to operate properly. Firstly, at start each robot should be able to sense at least three other robots nearby and secondly the value of R_s should be greater than d_u. Aggregation and unconstrained flocking behavior are also achieved for the large swarm. For the constrained flocking and obstacle avoidance, we present a simplified virtual physics mechanism which results in a practical flocking framework. The control method basically relies on the information about the neighbors and it is scalable. It is simple and can be realized with actual robots.

Acknowledgement

This work was supported by National Science Foundation of China (60874070), by National Research Foundation for the Doctoral Program of Higher Education of China (20070533131) and it is partly supported by The Hong Kong Polytechnic University (Project 4-ZZ9H).

References

1. Trianni, V., Labella, T.H., Dorigo, M.: Evolution of direct communication for a swarm-bot performing hole avoidance. In: Dorigo, M., Birattari, M., Blum, C., Gambardella, L.M., Mondada, F., Stützle, T. (eds.) ANTS 2004. LNCS, vol. 3172, pp. 130–141. Springer, Heidelberg (2004)
2. Bayindir, L., Şahin, E.: A review of studies in swarm robotics. Turk. J. Elec. Engin. 15(2), 115–147 (2007)
3. Şahin, E.: Swarm robotics: from sources of inspiration to domains of application. In: Şahin, E., Spears, W.M. (eds.) Swarm Robotics 2004. LNCS, vol. 3342, pp. 10–20. Springer, Heidelberg (2005)
4. Lochmatter, T., Martinoli, A.: Tracking odor plumes in a laminar wind field with bio-inspired algorithms. Springer Tracts in Advanced Robotics 54, 473–482 (2009)
5. Jatmiko, W., Sekiyama, K., Fukuda, T.: A particle swarm-based mobile sensor net-work for odor source localization in a dynamic environment. Distributed Autono-mous Robotic System 7, 71–80 (2007)
6. Baxter, J.L., Burke, E.K., Garibaldi, J.M., Norman, M.: Multi-robot search and res-cue: a potential field based approach. Studies in Computational Intelligence 76, 9–16 (2007)
7. Pearce, J.L., Powers, B., Hess, C., Rybski, P.E., Stoeter, S.A., Papanikolopoulos, N.: Dispersion of a team of surveillance and reconnaissance robots based on repellent pheromones. In: Proc. of the 11th Mediterranean Conference on Control and Automation (MED 2003) (June 2003)

8. Correll, N., Cianci, C., Raemy, X., Martinoli, A.: Self-Organized Embedded Sensor/Actuator Networks for "Smart" Turbines. In: Proc. of the IEEE/RSJ IROS 2006 Workshop on Networked Robotic Systems, Beijing, China (October 2006)
9. Chaimowicz, L., Cowley, A., Gomez-Ibanez, D., Grocholsky, B., Hsieh, M.A., et al.: Deploying air-ground multi-robot teams in urban environments. In: Multi-Robot Systems. From Swarms to Intelligent Automata, vol. III, pp. 223–234. Springer, Netherlands (2005)
10. Chaimowicz, L., Kumar, V.: Aerial shepherds: coordination among UAVs and swarms of robots. In: Distributed Autonomous Robotic Systems Part VI, pp. 243–252. Springer, Japan (2007)
11. Kalantar, S., Zimmer, U.: Distributed shape control of homogenous swarms of autonomous underwater vehicles. Journal Autonomous Robots 22, 37–53 (2007)
12. Reynolds, C.: Flocks, herds and schools: A distributed behavioral model. In: Proc. of the 14th annual conference on Computer graphics and interactive techniques, pp. 25–34. ACM Press, New York (1987)
13. Turgut, A.E., Çelikkanat, H., Gökçe, F., Şahin, E.: Self-organized flocking with a mobile robot swarm. Technical Report METU-CENG-TR-2008-01, Middle East Tech-nical University (January 2008)
14. Lee, G., Chong, N.Y.: Flocking controls for swarms of mobile robots inspired by fish schools. In: Lazinica, A. (ed.) Recent advances in multi robot systems. I-Tech Education and Publishing, Vienna (2008)
15. Yang, Y., Xiong, N., Chong, N.Y., Défago, X.: A decentralized and adaptive flocking algorithm for autonomous mobile robots. In: Proc. of the 3rd International Conference on Grid and Pervasive Computing Workshops, Kunming, China, May 2008, pp. 262–268 (2008)
16. Kim, D.H., Wang, H., Shin, S.: Decentralized control of autonomous swarm systems using artificial potential function-analytical design guidelines. J. Intell. Robot Syst. 45, 369–394 (2006)
17. Olfati-Saber, R.: Flocking for multi-agent dynamic systems: algorithm and theory. IEEE Transactions on automatic control 51(3), 401–420 (2006)
18. Bahceci, E., Soysal, O., Sahin, E.: A review: pattern formation and adaptation in multi-robot systems. Technical Report CMU-RI-TR-03-43, Carnegie Mellon University, Pittsburgh, PA, USA (2003)
19. Li, X., Ercan, M.F., Fung, Y.F.: A triangular formation strategy for collective behaviors of robot swarm. LNCS, vol. 5592, pp. 897–911. Springer, Heidelberg (2009)
20. Shucker, B., Bennett, J.K.: Scalable control of distributed robotic macrosensors. In: Proc. of the7th international symposium on distributed autonomous robotc systems, DARS 2004 (2004)
21. Shucker, B., Murphey, T., Bennett, J.K.: A method of cooperative control using occasional non-local interactions. In: Proc. of IEEE conference on robotics and automation (ICRA), Orlando, Florida (May 2006)
22. Shucher, B., Bennett, J.K.: Virtual spring mesh algorithms for control of distributed robotic macrosensors, Technical Report CU-CS-99, Department of Computer Sci-ence University of Colorado at Boulder (May 2005)
23. Klein, J.: Breve: a 3D simulation environment for the simulation of decentralized systems and artificial life. In: Proc. of Artificial Life VIII, the 8th International Conference on the Simulation and Synthesis of Living Systems. MIT Press, Cambridge (2002)

Nonlinear Analysis and Application of Servo Control System Based on Relay Feedback

Donglin Pu, Jianhua Wu, Zhenhua Xiong, Xinjun Sheng, and Han Ding

State Key Lab. of Mechanical System and Vibration,
School of Mechanical Engineering, Shanghai Jiao Tong University
1954 Hua Shan Road, Shanghai 200030, P.R. China
{pudonglintt,wujh,mexiong,xjsheng,hding}@sjtu.edu.cn

Abstract. In this paper, a PID controller tuning technology under different loads and velocities based on relay feedback for motor servo systems is proposed. Furthermore, the study on the relationship between relay parameters and plants' identification processes and results is novelly explored in depth. Through the theoretical analysis and experimental verification on the control loops of motor servo systems, we found that the major reason for the influence on relay feedback identification results is the nonlinear factors such as nonlinear friction in control loops. According to the analysis, the advices for optimizing the selection of relay parameters are given. Finally, a series of experiments on the general *XY* motion table and Zero-friction high precision air-cushion *XY* motion table have been conducted to prove the high performance of the proposed method. These experiments also validated the inference that the nonlinear factors have impacts on the results and processes of relay feedback identification.

Keywords: PID controller tuning, Relay feedback, Nonlinear analysis, Motion control.

1 Introduction

In the real industrial production, system parameter identification and controller tuning are key to production's efficiency and quality, in which time and efforts are also deeply involved. Generally speaking, we can derive a system's analytical model according to the laws of physics, or get the model parameters through experiments such as sweep-frequency. However, the former method is very difficult to use in complicated systems due to the complex to solve high-order differential equations, while the latter needs a lot of experiments to obtain useful data. Therefore, looking for a faster and easier way to get system parameters and tune controller is highly preferred in application. The relay feedback technology first proposed by Astrom and Hagglund [1] can obtain the parameters of system's model in a fast and accurate way, which is widely applied in controller tuning afterward. Some further extensions of relay feedback application can be found in literatures such as Shin.et.al [2] and Schei [3] doing.

Most of the research work mentioned above mainly concentrates on process control area. Distinguished from these literatures, we propose an offline PID parameters tuning

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 755–764, 2009.

method, which applies the relay feedback to motor servo control systems [4]. This method realizes the optimal PID parameters tuning under different loads and velocities.

In experiments, we found some interesting phenomenon. We got different results of identification if different relay parameters were chosen. However, theoretically for the same system, identification results should not be impacted by whichever relay parameters we select. Unfortunately, there has been seldom research work focused on this problem, although many scholars have done much work on the application of relay feedback technology. In this paper, an in-depth novel study has been done to deal with the problem. Some valuable advices are also presented to optimize the choice of relay parameters.

Through the analysis and experiments, we found that nonlinear factors in servo system are the main reasons for the identification results varying with the relay parameters. In addition, a further study on nonlinear factors in servo system control loops is done, which indicates: 1) system friction with strong nonlinear in position loop and velocity loop is the main factor to influence relay feedback; 2) the nonlinear factor in current loop has impact on the identification results. A series of experiments, especially the experiment on the zero-friction air-cushion motion table in contrast with general *XY* motion table, have been done to convectively prove the above conclusions.

This paper consists of six parts. Part 2 introduces the PID controller tuning with relay feedback. Part 3 is the kernel of the paper, which analyzed the nonlinear factors of the servo system control loops for relay feedback application. The detailed experiments are implemented in Part 4. The conclusion and acknowledge are declared in Part 5 and Part 6 respectively.

2 Relay Feedback Identification and PID Controller Tuning

The block diagram of the servo system with relay feedback is shown in Fig.1. The approximate simplified model of the plant explored in this paper is denoted as:

$$G_p(s) = \frac{Ke^{-Ds}}{\tau s + 1} \tag{1}$$

where K is the gain, and τ is the time constant. Now, the relay feedback is used here to identify the two parameters. The saturation element relay shown in Fig. 2 has been popularly used to improve the relay feedback's performance [5]. There is a switch between the control output and the plant. When the switch is connected to the relay feedback, the system starts to oscillate with P_u period.

The critical information like critical period P_u can be easily read from the system critical oscillation curves. Therefore, the normal critical frequency of the system can be denoted as:

$$\hat{\omega}_u = \frac{2\pi}{P_u} \tag{2}$$

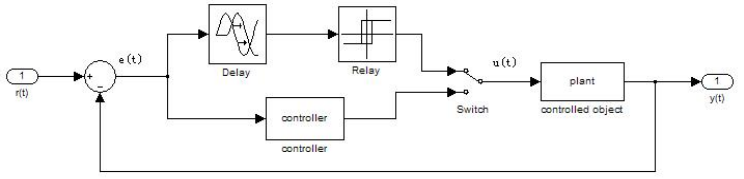


Fig. 1. Block diagram of control system based on relay feedback

and the system normal critical gain can be expressed as follows:

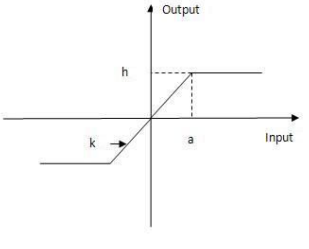
$$\hat{K}_u = \frac{2h}{\pi \bar{a}}\left[\left(\sin^{-1}\frac{\bar{a}}{a}\right)+\left(\frac{\bar{a}}{a}\sqrt{1-\left(\frac{\bar{a}}{a}\right)^2}\right)\right] \tag{3}$$

where h, a, and $\bar{a}$ are the amplitude, output amplitude and maximum input of the saturation relay, respectively. According to the definition and the information of critical point, K and τ of the controlled object can be derived as follows:

$$\tau = \frac{\tan[\pi-(D+d)\omega_u]}{\omega_u} \tag{4}$$

$$K = \frac{\sqrt{1+(\tau\omega_u)^2}}{K_u} \tag{5}$$

where ω_u and K_u can be obtained from formula (2) and (3). Thus, the model (1) of the motor servo systems can be fixed, which is useful to PID controller design.



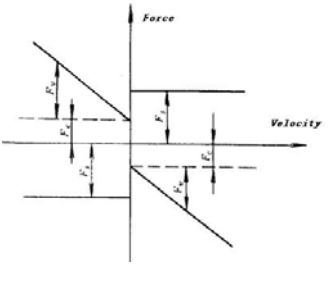


Fig. 2. Saturation element relay function **Fig. 3.** Friction diagram in motor servo system

In industry, PI controller is always applied in motor servo system's velocity loop. According to the integral of time-weighted absolute value of the error (ITAE) method [6], a popular PI calculation method, K_p and K_i can be described as function (6) and (7), where the K and τ are obtained from (4) and (5).

$$K_p = \frac{0.586}{K}\left(\frac{D}{\tau}\right)^{-0.916} \tag{6}$$

$$K_i = \frac{\tau}{1.03 - 0.165\left(\frac{D}{\tau}\right)} \tag{7}$$

Obviously, compared with the traditional model derivation and sweep-frequency experiment methods, relay feedback technology is a more convenient and faster way to identify the controlled object's parameters. In particular, it can be easily combined with other algorithms like ITAE to realize the PID tuning. Some experiments on the 'relay feedback + ITAE' method will be implemented in Part 4, which proves the effect of the PID controller tuning under different loads and velocities.

As our study focuses on the nonlinear analysis of the servo control system with relay feedback, we won't describe the PID tuning method here in detail.

3 Nonlinear Analysis on Servo Control System Based on Relay Feedback

There are many nonlinear factors and disturbances in real motor servo systems, which impact the control processes and results. Therefore, it is possible that the nonlinear factors will also impact the relay feedback identification in the motor servo control system. There are two important parameters in the relay feedback system: *h*, the output amplitude, and *D*, the delay. Theoretically, *h* determines the output amplitude of the identified system only and will not impact the identification results. However, in our experiments, *h* has influences on the identification results directly (Fig.12 and 13), and the characteristics of *D* shown in experiments are also different from those in the theoretical simulation. As a next step, we will analyze the nonlinear factors in the position/velocity loop and current loop.

3.1 Nonlinear Analysis in the Position/Velocity Control Loops

As we know, friction, especially the Coulomb friction, is the primary nonlinear factor for position and velocity loops in general motor servo systems. The output amplitude of relay, *h*, determines the velocity and displacement of identified object (motor). And the friction in servo system also has influence on motor's movement. Therefore, the friction will probably influence the final identification results of relay feedback. To verify this assumption, we conducted the below theoretical analysis.

In motor motion control, the friction can be described in the formula (8)

$$F_f = F_c \operatorname{sgn}(\omega) + F_v \omega \tag{8}$$

where F_c is the Coulomb friction, and F_v is the viscous friction coefficient. In the controlled object, $1/\tau$ indicates the ratio of viscous friction coefficient and load inertia. The formula (9) can be derived when the load inertia is constant:

$$\frac{1}{\tau} = \frac{\hat{F}_v}{m} = \frac{F_f}{m\omega} \tag{9}$$

$$\tau = \frac{m}{F_v + F_c / |\omega|} \tag{10}$$

where $\hat{F}_v$ is the estimated value of F_v, and m is the load inertia. The formula (10) is derived from formula (8) and (9). According to Fig.3, we know that F_v is relative to ω and F_c is the main nonlinear element. Therefore, τ varies with $/\omega/$ if m is fixed. Since ω is determined by the output amplitude h in relay feedback system, the selection of h will impact the results of identification. This means that nonlinear friction has an impact on the identification results.

To further verify the above analysis, we designed an experiment on the zero-friction air-cushion motion table in contrast with the general XY motion table with friction. The implement of the experiment will be described in details in Part 4.

3.2 Nonlinear Analysis in the Current Control Loop

Current loop, another important control loop, is the innermost loop of servo system. The nonlinear factor in this loop was also analyzed here for relay feedback. We did a simulation relay feedback experiment based on the $G_p(s)_{simulation}$, a general motor model in industry. The simulation results are illustrated in Fig.4.

$$G_p(s)_{simulation} = \frac{9000}{0.112s+1} e^{-0.001s} \tag{11}$$

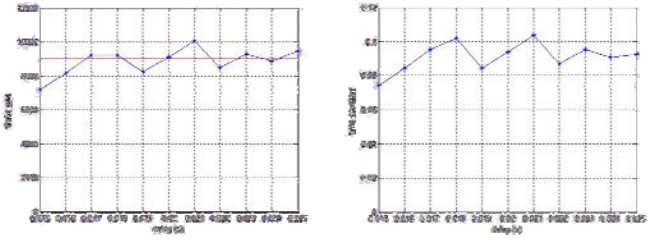


Fig. 4. Theoretical simulation results: the left diagram is the relationship between D and static gain; the right diagram is the relationship between D and time constant

Correspondingly, the real experiments results are illustrated as in Fig. 5, which have eliminated the friction disturbance by using zero-friction air-cushion table.

Obviously, there is a large variance between the results of theoretical simulation and real experiments. According to our experience as described in Section 3.1, the reason is probably the nonlinear factors in the current loop.

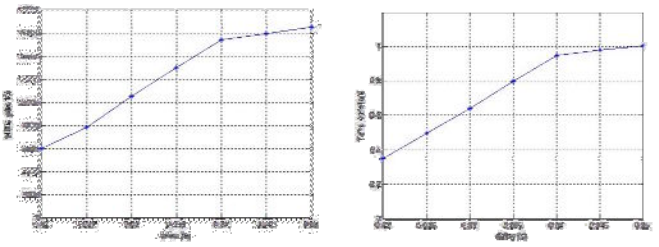


Fig. 5. Real experiments results: the left diagram is the relationship between D and static gain; the right diagram is the relationship between D and time constant

Regardless of the outside disturbance, the dynamic model of the plant can be expressed as follows, as the motion table is driven by linear motors.

$$\begin{aligned} M\ddot{x} + B\dot{x} + F_{load} &= F_m \\ K_e\dot{x} + L_a\frac{dI_q}{dt} + R_aI_q &= u \\ F_m &= K_fI_q \end{aligned} \tag{12}$$

Where I_q is the coil current and L_a is the coil inductance. Let's denote $I_{nonlinear}$ as:

$$I_{nonlinear} = L_a\frac{dI_q}{dt} \tag{13}$$

This describes the armature current changing's influence in the loop. Fig.6 illustrates that there are two rapid changes in each armature current oscillation period after bringing relay feedback to the system. Here $I_{nonlinear}$ is a constant C in the slope phase while is zero in the wave top. Then, the influence of $I_{nonlinear}$ can be denoted as:

$$C\frac{t1}{t1+t2} \tag{14}$$

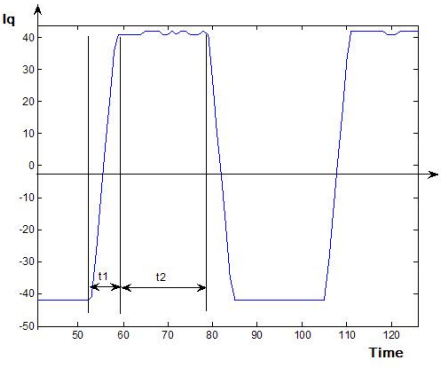


Fig. 6. The diagram of armature current in relay feedback

where *t1* is the time of slope phase and *t2* is the time of wave top. According to formula (14), the influence of $I_{nonlinear}$ is inversely proportional to *t2*. As *t2* increases with *D*, the results of relay feedback identification are impacted by *D* finally (as illustrated in Fig.5).

According to the analysis above, it is certain that *D* must be selected reasonably to overcome the nonlinear factors influence on relay feedback. Here we provide two advices for the selection of *D*.

(1) A larger delay should be used in relay feedback identification in order to reduce the influence of the nonlinear factors.
(2) As the critical oscillation amplitude grows with *D*, the choice of *D* must ensure that the system runs within the safety range.

4 Experiments

All the experiments in this study were implemented on two motion platforms (as shown in Fig.7): the general *XY* motion table driven by YASKAWA rotary servo motors (SGMGH 10ACA21), and the zero-friction air-cushion motion table driven by Kollmorgen linear servo motors (IL18-100A1P1). Both platforms are controlled by dSPACE controller (DS1103).

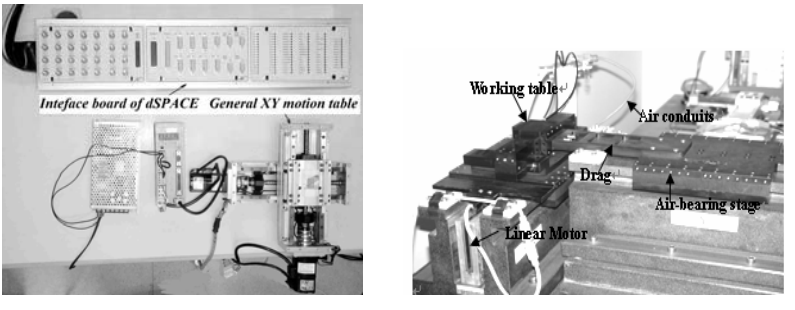


Fig. 7. The pictures of experiments platforms: the left is general XY motion table and the right is zero-friction air-cushion motion table

4.1 Experiments for PID Controller Tuning Based on Relay Feedback

To verify the effectiveness and applicability of the relay feedback technology, the step response experiments under different velocities and loads were conducted in this section. Firstly, we implemented the experiment to realize 1000 rmp step response without loads on YASKAWA motor. In accordance with the method described in Part 2, the model of the motor without loads was identified as:

$$\hat{G}(s) = \frac{3055}{0.05558s + 1} e^{-0.002s} \tag{15}$$

and the PI parameters can also be designed as K_p=0.004, K_i=*0.0743* by formula (6) and (7). The response curve of the system based on relay feedback is shown in Fig.8 as curve *B*. Curve A is created by another popular PID tuning method Z-N. It is clear that the response curve B is much better than A, which proves the advantages of the 'relay feedback + ITAE' PID tuning method.

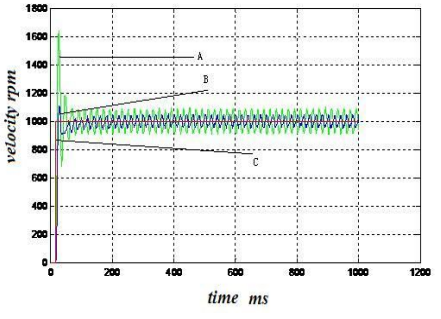


Fig. 8. The experiment results without load: A is created by Z-N method; B is created by the 'relay feedback + ITAE' method; C is the input instruction curve

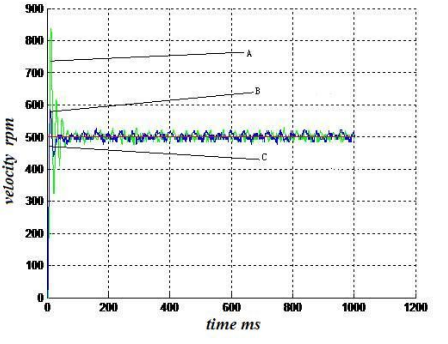


Fig. 9. The experiment results with loads: A is created by Z-N method; B is created by the 'relay feedback + ITAE' method; C is the input instruction curve

Next, we implemented the experiment to realize 500 rmp step response with loads on the general XY motion. The model with loads was identified as:

$$\hat{G}(s) = \frac{329.7}{0.01976s + 1} e^{-0.002s} \tag{16}$$

and PI parameters can be designed as: K_p=0.0145, K_i=0.7429 by formula (6) and (7). The control effectiveness (Fig.9) of 'relay feedback + ITAE' is still much better, which proves that the explored method is suitable for different loads and velocities.

4.2 Experiments for Nonlinear Analysis of the Motor Servo System with Relay Feedback

As described in Section 3.1, fiction is the main nonlinear factors to impact relay feedback identification in the position and velocity loops. Here, the experiments on the motion tables with and without friction are conducted to verify the theoretical analysis and inference. The identification results shown in Fig.10 and11 came from the experiments on the motion table with friction. It is clear that the identification results are influenced by the selection of *h*.

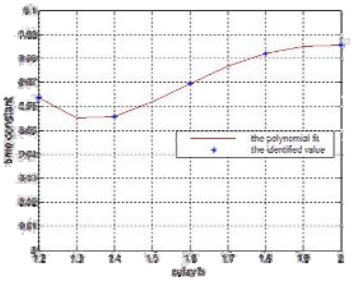


Fig. 10. Relationship between *h* and identified time constant with friction

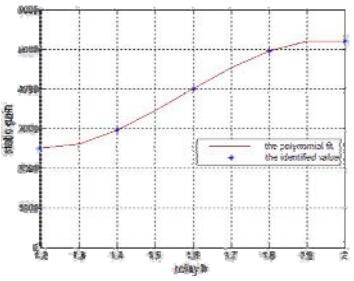


Fig. 11. Relationship between *h* and identified static gain with friction

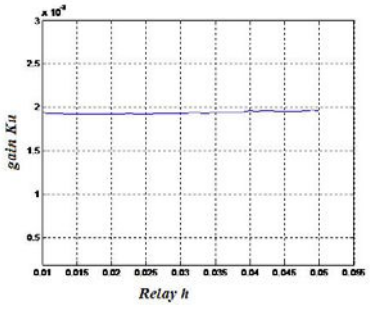


Fig. 12. Relationship between h and identified time constant without friction

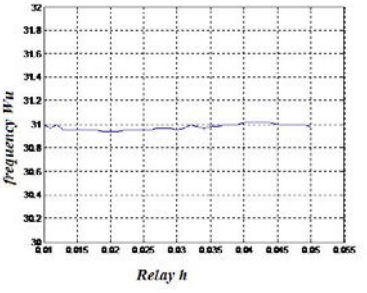


Fig. 13. Relationship between h and identified static gain without friction

The identification results shown in Fig.12 and 13 came from the zero-friction air-cushion motion table. The identified parameters' curves are nearly horizontal, which means that the selection of *h* won't impact the identified result if the friction is eliminated. These comparison experiments prove the conclusion in Section 3.1.

5 Conclusion

The application of relay feedback technology combined with ITAE in motor servo system is explored in this paper. Furthermore, the nonlinear analysis of the servo control loops based on relay feedback is novelly investigated, which provided the meaningful advices and strong supports for relay feedback applications. The convictive experiments on the general XY motion table and air-cushion XY motion table driven by linear motors are also described in detail in the end.

Acknowlegements

This work is supported in part by the National Natural Science Foundation of China under Grant 50805095, National High-tech Research and Development Program (863

Program):2007AA04Z316 and the Open Research Fund Program of the State Key Laboratory of Fluid Power Transmission and Control under Grant number: GZKF-2008002.

References

1. Åström, K.J., Hägglund, T.: Automatic tuning of simple regulators specifications on phase and amplitude margins. Automatica 20(5), 645–651 (1984)
2. Shin, C.H., Yoon, M.H., Park, I.S.: Automatic Tuning Algorithm of the PID Controller Using Two Nyquist Points Identification. In: SICE, Tokushima, pp. 1225–1228 (July 1997)
3. Schei, T.S.: Automatic Tuning of PID Controllers Based on Transfer Function Estimation. Automatica 30(12), 1983–1989 (1994)
4. Yongqiang, H., Jianhua, W., Zhenhua, X., Han, D.: Research on PID Controller Tuning of Servo Systems Based on Relay Feedback, Machinery (December 2008)
5. Shen, S.H., Yu, H.D., Yu, C.C.: Use of saturation relay feedback for autotune identification. Chemical Engineering Science 51(8), 1187–1198 (1996)
6. Su, W.S., Lee, I.-B., Lee, B.-K.: On-line process identification and automatic tuning method for PID controllers. Chemical Engineering Science 53(10), 1847–1859 (1998)

Kalman Estimator-Based State-Feedback High-Precision Positioning Control for a Micro-scale Air-Bearing Stage

ZeGuang Dong, PinKuan Liu*, and Han Ding

State Key Laboratory of Mechanical System and Vibration, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
Pkliu@sjtu.edu.cn

Abstract. Due to the high positioning precision requirement and the low damping surface effect, the design of high-quality controller for a micro-scale air-bearing positioning stage is a challenging for micro-scale positioning system. Furthermore, the strict synchronization requirement of the linear motors on both sides of the gantry beam even intensifies the difficulty of the controller design. An optimized state-feedback gain has been designed by implementing a suitable Kalman estimator to minimize both the tracking error and the control efforts. Significantly, the merits of such a state-feedback control are the capabilities of handling the coupling of the two motors on the two outputs and compressing the overshooting as well. Finally, the experimental results of the proposed state-feedback controller on the air-bearing stage are displayed in comparison with the traditional PID control law.

Keywords: High-precision positioning, air-bearing stage, state-feedback control, Kalman estimator, micro-scale.

1 Introduction

Micro-scale and even nano-scale high-precision and long travel-length motion control technology are widely used in industry fields such as semiconductor manufacturing systems. Among the abundant micro-scale motion control equipments, H-typed gantry stage is one of the most popular benchmark systems aiming at developing and polishing advanced high-precision positioning control methods. In this stage, two motors are parallelly mounted along a gantry beam in a single feed axis, which is equipped with a high-density power resources to fulfill the consumption of the dual actuators. In real applications, the asynchronous dynamics of dual motor actuators can produce a rotary motion, which are undesirable or even dangerous for the system and hence intensify the challenge faced by the micro-scale motion control technology. Significantly, the target of the controller design is not only to achieve high-precision with sufficiently short settling time but also to eliminate the synchronization error between the two parallel motors. To this end, the two mainstream control approaches widely used in gantry stage positioning control are master/slave scheme and set-point

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 765–775, 2009.

coordinated scheme. In the former one, the master motor executes the reference command directly whose dynamics will be followed by the slave motor [1]. Nevertheless, due to the time-delay embedded in such master/slave time-response dynamics, the synchronization error can not be compressed into a sufficient low level. Bearing in mind of such a problem, Hsieh *et al.* [2] multiplied the command to the slave motor by a proportional gain to compensate the different characteristics between two motors and hence somewhat reduce the synchronization error. As to the set-point coordinated control approaches, each motor drives the actual trajectory of its servo loop adhere as closely as possible to the reference trajectory. More promisingly, some Lyapunov function-based nonlinear control methods, such as sliding mode control strategy [3], are developed to further reduce the synchronization error in [4]. However, due to the unavoidable chattering phenomenon [3], the oscillations can not be sufficiently reduced. Therefore, developing a suitable advanced control strategy to *greatly reduce the synchronization error* is quite desirable for the gantry beam's high-precision positioning.

On the other hand, friction is another essential issue in micro-scale motion control as it causes uncertainties. Many efforts have been developed to address such unexpected friction effects, in which the most representative approaches are some kind of state-feedback controller combined with a feed-forward compensator. More precisely, a zero phase error tracking controller (ZPETC) based on inversion approximation of the closed-loop system was introduced for the feedforward compensator [5,6], and some disturbance estimators were developed to prohibit the low-frequency disturbances [7,8]. Apart from the soft analytical methods handling uncertainties and external noise, some other mechanical-electronic scholars resort to better mechanical structures to greatly reduce the friction, such as air bearings platforms, which provides frictionless-like motion [9] and hence the sliding guideways do not show any undesirable frictional effects like stick-slip. However, air-bearing mechanism also has its inherent disadvantage, i.e. too small damping in the surfaces between the gantry beams and the guideways, which makes it hard for PID controller to quickly drive the load back when exceeding the set-point. Consequently, it is quite necessary and promising to turn to some advanced control methods based on state-space model, which can provide sufficient degrees of freedom and thus be beneficial to *seek a balance between steady-state and transient performances* [10]. Taking the above problems into consideration, a state-space controller of air-bearing positioning stage will be designed in this paper to address the two main problems mentioned above.

The rest of the paper is organized as follows. First, our experimental equipment, i.e. the micro-scale air-bearing positioning stage with large inertia is introduced, and then the two essential theoretical problems are distilled from the numerical high-accuracy control applications in Section 2. Then, in Section 3, the minimal state-space model is realized by transforming the transfer function matrix produced by frequency-domain analysis. Afterwards, in Section 4, a suitable Kalman estimator and an optimal LQR state-feedback controller gain vector are developed independently to minimize the overshooting, oscillations and the settling time, and improve the tracking accuracy and synchronization performances

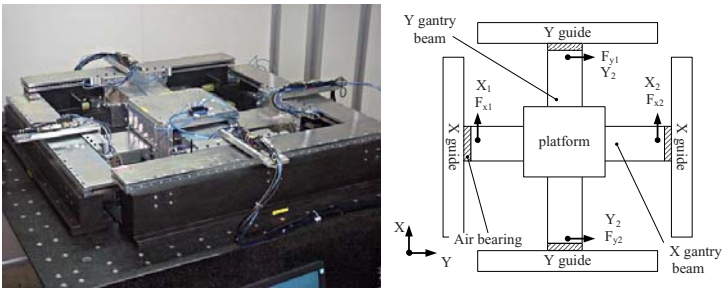
as well. Intensive experimental results are demonstrated in Section 5 to verify the feasibility, reliability and superiority of the present Kalman estimator-based state-feedback controller. Finally, the concluding remark is drawn in Section 6.

2 System Description

The micro-scale air-bearing positioning stage with large inertia is shown in Fig. 1, which consists of two gantry (X and Y-axes) beams, an upper platform and an underneath basement table. The X and Y-axes beams are constrained to move on X and Y directions by two fixed X and Y-axes sliding guideways, respectively. The platform sliding freely on both two gantry beams, is also constrained in the horizontal plane to guarantee the planar and smooth motion. Note that each gantry beam is driven by two permanent magnet linear motors, in which two pairs of coil windings are mounted on both ends of the gantry beams while the permanent magnets are fixed on the base as the stators. In order to achieve high-precision positioning control performance, the positioning stage adopts air bearings and hence all of the stage motion is almost free from friction. Here, the driven forces are denoted by $F_{X_1}, F_{X_2}, F_{Y_1}, F_{Y_2}$ in Fig. 1, and the movements of each gantry beam X_1, X_2 (or Y_1, Y_2) are measured by two linear optical encoders located at both ends of the beam. The scanning period, resolution and maximal working speed of the encoder are $20\mu s$, $0.5\mu m$ and $7.2m/s$. The technical design parameters of the stage are as follows: the movement range of the stage is $300mm \times 300mm$, the maximal acceleration is $1.5g$, maximal stable velocity is $600mm/s$, the oscillation of stable velocity is 1%, and the maximal load is $40Kg$.

Now, we distill the two essential problems concerning the positioning control of such an air-bearing positioning stage.

– *Motor synchronization*
Basically, each gantry beam has just one degree-of-freedom, say, $X_1 = X_2$, $Y_1 = Y_2$. However, due to the different dynamics and load inertia of the motors on both sides of the beam, the synchronized movement of these two motors can not be guaranteed in practice. Moreover, such unexpected



(a) (b)

Fig. 1. The Prototype(a) and Configuration(b) of the precision gantry stage

synchronization errors will slightly rotate the gantry beam around Z-axis perpendicular to the $X-Y$ plane, which may damage both sides of the air-bearings and the sliding guideways. Note that the conventional PID controller based on single-loop analysis can not address the severe coupling and interactions between the motors, and hence an suitable advanced control method is expected to effectively reduce the synchronization errors.

– *Low damping effect compensation*
As shown in Fig. 1, the air-bearing is employed to eliminate the friction, making the surface damping between the sliding guideways and the moving parts quite low. Due to such a low damping characteristics, it will take a long time for the outputs to settle down upon exceeding the set-point. Although in conventional PID controllers, raising D-gain will cut the overshooting, it will also amplify external high-frequency noises. On the other hand, decreasing I-gain can also shorten the setting time, but it will inevitably impair the tracking accuracy. Thus PID controller can not yield satisfactory performance due to its lacking of sufficient adjusting parameters and analytical dynamics tuning method.

Thereby, to solve the above-mentioned two urgent yet serious problems, it is very natural to seek assistance from the state-space description, which is the common internal model of numerous advanced control approaches having exerted significant impacts on manufacturing industry. This kind of state-space model has the potential to solve the two control-relevant problems since it extracts the internal dynamics, i.e., the state x containing much more plentiful dynamic information than the outputs $y_{1,2}$, from the input/output sequence, which sharply increases the degrees of freedom for later controller design while greatly facilitates the closed-loop performance analysis as well.

3 Identification of the Air-Bearing Positioning Stage

For conciseness, we fix the Y-axis beam and just use two linear motors to control the displacements of the X-axis beam. Clearly, now the positioning stage system shown in Fig. 1 is a two inputs (i.e. the two control voltages u_1 and u_2 of the two motors) and two outputs (i.e. the displacements y_1 and y_2 of the two ends of the X-axis beam) system. We use linear identification methods to generate a concise linear time invariant control-oriented model of the system which will be described in two stages: external transfer function and internal state-space expression.

The idea behind our frequency-domain identification experiments was to excite the system with sinusoids of different frequencies and measure the resulting magnitude and phase of the response so as to gain the Bode diagram. We would then fit these points in the frequency domain with the response of a rational transfer function, and use this transfer function matrix as the external model of the system. Hereby, we began by selecting roughly twenty sample frequencies ranging from $0.1 rad/s$ to $200 rad/s$. Having completed a scan over the entire

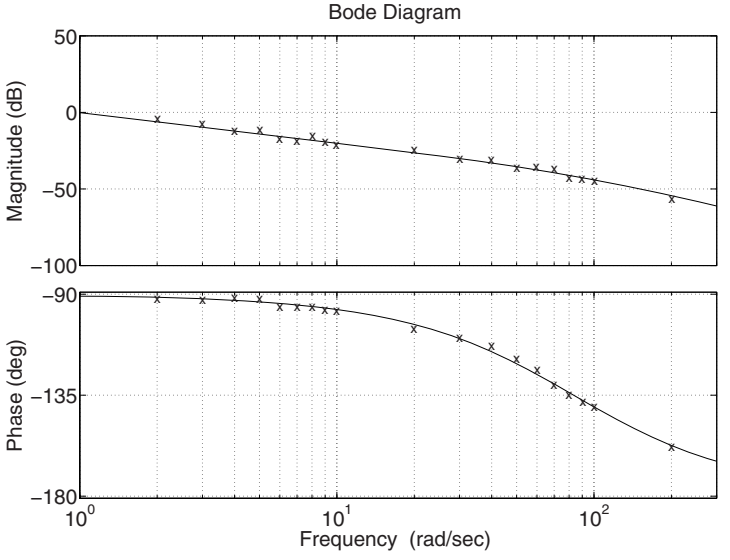


Fig. 2. Bode diagram of $u_1 \rightarrow y_1$ channel of the two-actulator/two-displacement system. Here, the bode diagrams of channels $u_2 \rightarrow y_1$, $u_2 \rightarrow y_2$ and $u_1 \rightarrow y_2$ are omitted for conciseness.

range of frequencies, we then repeated the experiments 50 times to acquire sufficient data to make meaningful fits and reduce the impact of experimental error. Note that, in the identification procedure, to identify the influences of u_1 or u_2 on $y_{1,2}$, only one of the motors needs to be activated while the other is switched off.

Fig. 2 shows the 20-point measurements obtained in one of the 50 experiments. Noticing that the system's magnitude slopes begin at $20db/dec$ and then roll off at $40db/dec$, the simplest model to fit the data would be a second-order self-integrating system, which well explains the mechanism of the motor-displacement dynamics since the linear relationships between the input voltages $u_{1,2}$ and the intermediate variables (or motor output velocities) $v_{1,2}$ are one-order inertia transfer functions while integrating $v_{1,2}$ yields the displacements $y_{1,2}$, respectively. Fitting such a system, we obtain

$$G(s) = \begin{bmatrix} \dfrac{0.9837}{(0.012s+1)s} & \dfrac{0.097}{(0.047s+1)s} \\ \dfrac{0.08}{(0.05s+1)s} & \dfrac{0.756}{(0.014s+1)s} \end{bmatrix}. \tag{1}$$

It is found that the gains of the four input/output channels are self-consistent as $|G_{1,1}(j\omega)|$ and $|G_{2,2}(j\omega)|$ are much larger than the coupling magnitudes $|G_{1,2}(j\omega)|$ and $|G_{2,1}(j\omega)|$. However, in order to gain high-precision positioning control performance, these coupling magnitudes can not be ignored, that

is why the single-loop PID methods can not yield satisfactory performances. By discretizing the transfer function matrix (1), one can obtain the minimal realization of the discrete state-space model as below:

$$x(k+1) = Ax(k) + Bu(k) \tag{2}$$

$$y(k) = Cx(k) \tag{3}$$

with $u(k) \in \mathbb{R}^2, x(k) \in \mathbb{R}^6, y(k) \in \mathbb{R}^2$,

$$A = \begin{bmatrix} 0.4138 & 0.3315 & 0.2374 & -0.1919 & -0.05936 & 0.02906 \\ 0.5516 & 0.3671 & -0.4931 & 0.3986 & 0.1233 & -0.06035 \\ -0.4417 & 0.3617 & 1.237 & -0.3463 & 0.1429 & -0.06996 \\ 0.3571 & -0.2923 & -0.2653 & 1.023 & 0.3163 & -0.1549 \\ 0.1104 & -0.09041 & 0.0371 & -0.02999 & 1.48 & -0.485 \\ -0.05406 & 0.04426 & 0.2485 & -0.2009 & 0.9379 & 0.03042 \end{bmatrix},$$

$$B = \begin{bmatrix} 0.01723 & -0.001315 & -0.04717 & 0.03813 & 0.01179 & -0.005773 \\ 0 & 0 & 0.01562 & 0 & 0.0625 & 0 \end{bmatrix}^T,$$

$$C = \begin{bmatrix} -0.0209 & 0.01711 & -0.03814 & 0.04155 & 0.01108 & -0.005422 \\ 0.004204 & 0.004969 & 0.00295 & -0.002385 & 0.03378 & 0.02759 \end{bmatrix}.$$

Clearly, the two coupling channels $\{u_1 \to y_2\}$ and $\{u_2 \to y_1\}$ have been naturally embedded in the state-space expression (2) and (3). Hereby, with the above discrete-time state-space models of the air-bearing positioning stage, we are now prepared to launch our systematic validation experiments and controller design.

4 Controller Design

Since the state x of the system can not be directly measured, one has to design an observer to gain the state from the input/output sequence $\{u(k), y(k)\}_{k=1,\cdots,\infty}$. Thus, to achieve a suitable estimate state $\hat{x}$ of x, we resort to Kalman estimator [12] which is the optimal solution to the following discrete estimation problem:

$$x(k+1) = Ax(k) + Bu(k) + Gw(k) \tag{4}$$

$$y(k) = Cx(k) + Du(k) + Hw(k) + v(k) \tag{5}$$

with known inputs $u(k)$ and process and measurement white noise $w(k)$ and $v(k)$ satisfying $E(w) = E(v) = 0$, $E(ww^T) = Q$, $E(vv^T) = R$, $E(wv^T) = N$. Here, Q and R are positive-definite symmetric matrices. The Kalman estimator L can minimize the steady-state error covariance $\lim_{k\to\infty} E((x - \hat{x})(x - \hat{x})^T)$. More precisely, the state space expression of Kalman estimator L is given as follows:

$$\hat{x}(k+1) = A\hat{x}(k) + Bu(k) + L(y(k) - C\hat{x}(k)) \tag{6}$$

Now, with the state estimate $\hat{x}(k)$, one can design the state-feedback controller

$$u(k) = K\hat{x}(k) \tag{7}$$

to stabilize the closed-loop system, where L and K denote the observer gain and state-feedback gain vectors, respectively. According to the separation principle, each of K and L can be designed independently without effecting the performance of the other [11]. In detail, subtracting Eq. (6) from Eq. (2) yields

$$e(k+1) = (A + LC)e(k), \tag{8}$$

with the state estimate error $e(k) \triangleq x(k) - \hat{x}(k)$. As a consequence, the observer and controller designing methods aim at stabilization of $A + LC$ and $A + BK$, respectively. Note that, in order to shorten the settling time, the poles of the observer $A + LC$ are usually selected to dissipate faster than those of $A + BK$. Thereby, by selecting $Q = I_6$, $R = 0.5 \cdot I_2$ and $N = 0$, we have gained a Kalman estimator L for the state-space model (2) and (3) as

$$L = \begin{bmatrix} 0.1286 & 0.1026 & 0.1638 & -0.1309 & -0.0391 & 0.0215 \\ -0.0016 & -0.0014 & 0.1086 & 0.1108 & 0.1711 & 0.1687 \end{bmatrix}^T. \tag{9}$$

Certainly, one can adjust all the entries of Q and R to further improve L.

Now, let us investigate the design of the state-feedback gain K. First, we address the regulator problem and optimally select such a vector gain K by using standard linear quadratic regulator (LQR) method [11] which minimizes the following index $J(u(k)) = \sum_{k=1}^{\infty}(x(k)^T Q x(k) + u(k)^T R u(k) + 2x^T(k) N u(k))$. Actually, this LQR method first calculate the infinite horizon solution S of the associated discrete-time Riccati equation $A^T SA - S - (A^T SB + B)(B^T SB + R)^{-1}(B^T SA + N^T) + Q = 0$, and then K is derived from S by $K = -(B^T SB + R)^{-1}(B^T SA + N^T)$. Therefore, by choosing $Q = 0.1 \cdot I_6$, $R = I_2$ and $N = 0$, we have obtained the LQR solution of K as

$$K = \begin{bmatrix} 0.4578 & -0.3488 & -0.6095 & 0.4927 & 0.1524 & -0.0746 \\ 0 & 0 & 0.1852 & -0.1268 & 0.9311 & -0.4489 \end{bmatrix}. \tag{10}$$

Of course, one can polish the design of K by selecting better Q, R and N. Significantly, it can be verified that the eigenvalues of $A + LC$ are much closer to the origin of the complex plan than those of $A + BK$, making the transient process of the estimator settle down much quicker than that of the controller. By this means, a sufficiently large stability region of the closed-loop system is ensured.

Based on the LQR solution K shown above, we are now ready to address the tracking problem for the gantry beam positioning control application by modifying the input in (7) to

$$u(k) = K\hat{x}(k) + \rho r, \tag{11}$$

where r and ρ are the set-point and offset value to be determined afterwards, respectively. Now, substituting Eq. (11) into Eq. (6) yields $\rho = C(I - A - BK)^{-1}B$. In order to demonstrate the present controller more vividly, we provide the control system structure as shown in Fig. 3 composed of the plant, the

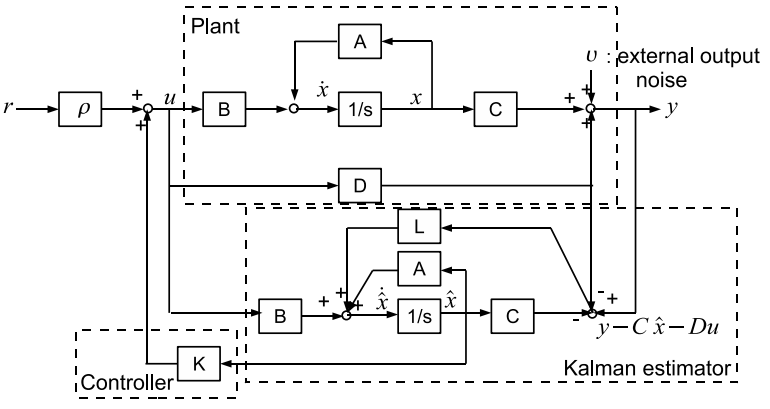


Fig. 3. State-feedback controller structure based on Kalman estimator

Kalman estimator and the state-feedback controller. More precisely, first the input signal $u(k)$ is fed into the plant to yield state $x(k)$ and output $y(k)$. Afterwards, by inputing $y(k)$ and feed-forward signal $Du(k)$ into the Kalman estimator (6), the state estimate $\hat{x}$ is yielded which is imported into the state-feedback controller gain K to settle down the response of the initial conditions. For tracking problem, the offset signal ρ is added into $K\hat{x}(k)$ to drive $y(k)$ to the set-point $r(k)$ asymptotically.

5 Experimental Results

The dSPACE software including ControlDesk Standard and MATLAB/Simulink Real-Time toolbox are used to establish the controller software architecture. In real applications, the dSPACE DS-1103 DSP board forms the kernel of the closed loop system, which processes the measured output signal and then executes the control algorithm to produce the control law. As mentioned in Section 2, in order to focus our investigation on actuators' synchronization, high tracking precision and low damping surface compensation, we fix Y-axis gantry beam and just control the movement of X-axis gantry beam. As shown in Fig. 4, we demonstrate its control performances of tracking $r = 20mm$ step set-point curve in comparison with conventional PID approaches with different parameter settings. Here, the sampling period is set as $1ms$, and optimal parameters of the present state-feedback controller and the PID controller are provided in Sections 3, 4 and Fig. 4, respectively.

The statistical results, such as the expected and optimal settling time, overshoot, steady-state tracking errors and synchronization errors are shown in Table 1. These experimental results show the advantages of the present state-feedback controller over PID, including improved transient performance and better displacements synchronization performance. The two underlying reasons of these merits are given as follows:

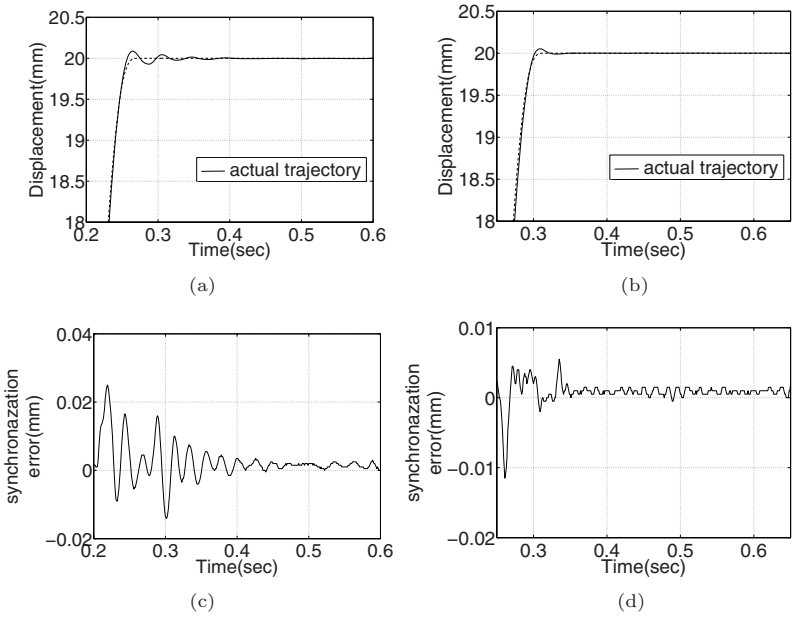


Fig. 4. The comparison of tracking and synchronization performances of PID controller and the present state-feedback controller. (a) PID control: tracking set point $r = 20mm$, (b) state-feedback control: tracking set point $r = 20mm$, (c) PID control: synchronization error curve, (d) state-feedback control: synchronization error curve. Here, the parameters of the PID controller are set as: $k_p = 0.5$, $k_i = 0.001$ and $k_d = 0.004$.

Table 1. Statistical comparison of *state-feedback controller* and *PID controller* on the air-bearing positioning stage

	Indexes			
	$t_s(ms)$	$\sigma_p(\mu m)$	$e_p(\mu m)$	$e_s(\mu m)$
E(PID)	468	127	± 2.95	± 27
O(PID)	403	105	± 2.82	± 25
E(state-feedback)	213	54	± 2.87	± 15
O(state-feedback)	187	43	± 2.70	± 12

t_s: settling time, σ_p: overshooting, e_p: steady-state error,
e_s: max synchronization error;
E(state-feedback), E(PID): Expectation values of *state-feedback* and *PID controllers*, respectively;
O(state-feedback), O(PID): Optimal values of *state-feedback* and *PID controllers*, respectively.

– The state-space internal model can easily consider the actuators-outputs coupling, and thus endow the state-feedback controller the potential of improving the synchronization performances;
– The Kalman estimator and LQR state-feedback gain design have optimized the poles of the state matrix of the closed-loop system, which drives the system towards the optimal equilibrium of steady-state and transient performances.

6 Conclusion

In this study, we investigated the identification and optimal control of the a micro-scale air-bearing positioning stage. First, we used frequency-domain identification techniques to obtain the transfer function matrix describing the coupling nature of the system. Afterwards, such a transfer function matrix is transformed into a minimal realization of state-space expression, with which a suitable Kalman estimator is designed to extract the state. Finally, we applied an optimal LQR state-feedback gain to yield a stable control law to synchronize the two ends' motions which track the desired set-point. We then demonstrate the performance of the present state-feedback control in contrast to conventional PID controllers. It is observed that the overshooting, oscillation and settling time have been greatly cut by over 60% as well. More promisingly, the synchronization error have also been effectively cut down, therefore, the high-precision control bottleneck of PID positioning controllers can be expected to be broken through by the present Kalman estimator-based state-feedback control method. In brief, it is believed that advanced control approaches based on state-space model have the potential to outperform traditional PID controller in handling many complex mechanical-electronic characteristics often encountered in nano- and micro-scale motion control systems.

Acknowledgement

This work is supported by the National Natural Science Foundation of China (NSFC) under Grant No. 50675132 and the National high Technology Research and Development Program of China (863 Program) under Grant No. 2006AA4Z334.

References

1. Tan, K.K., Lim, S.Y., Huang, S., Dou, H.F., Giam, T.S.: Coordinated motion control of moving gantry stages for precision applications based on an observer-augmented composite controller. IEEE Trans. Control Systems Technology 12(6), 984–991 (2004)
2. Hsieh, M.F., Tung, C.J., Yao, W.S., Wu, M.C., Liao, Y.S.: Servo design of a vertical axis drive using dual linear motors for high speed electric discharge machining. Int. J. Machine Tools and Manufacture 47(3-4), 546–554 (2007)

3. Khalil, H.K.: Nonlinear systems. Pearson Education, NJ (2002)
4. Giam, T.S., Tan, K.K., Huang, S.: Precision coordinated control of multi-axis gantry stages. ISA Trans. 46(3), 399–409 (2007)
5. Tomizuka, M.: Zero phase error tracking algorithm for digital control. J. Dyn. Syst. Meas. Control 109(1), 65–68 (1987)
6. Endo, S., Kobayashi, H., Kempf, C.J., Kobayashi, S., Tomizuka, M., Hori, Y.: Robust digital tracking controller design for high-speed positioning systems. Control Engineering Practice 4(4), 527–536 (1996)
7. Kempf, C.J., Kobayashi, S.: Disturbance observer and feedforward design for a high-speed direct-drive positioning table. IEEE Trans. Control Systems Technology 7(5), 513–526 (1999)
8. Wang, Y., Xiong, Z., Ding, H., Zhu, X.: Nonlinear friction compensation and disturbance observer for a high-speed motion platform. IEEE Int. Conf. on Robotics and Automation 2004(5), 4515–4520 (2004)
9. Slocum, A., Basaran, M., Cortesi, R., Hart, A.J.: Linear motion carriage with aerostatic bearings preloaded by inclined iron core linear electric motor. Precision Engineering 27(4), 382–394 (2003)
10. Schmidt, C., Heinzl, J., Brandenburg, G.: Control approaches for high-precision machine tools with airbearings. IEEE Trans. Industrial Electronics 46(5), 979–989 (1999)
11. Chen, C.T.: Linear System Theory and Design. Oxford University Press, Oxford (1999)
12. Franklin, G.F., Powell, J.D., Workman, M.L.: Digital Control of Dynamic Systems, 2nd edn. Addison-Wesley, Reading (1990)

Dynamic Control and Analysis of a Nonholonomic Mobile Modular Robot

Jingguo Wang and Yangmin Li

Department of Electromechanical Engineering, University of Macau,
Macao SAR, China
{YA77403,YMLi}@umac.mo

Abstract. This paper presents a modeling method for a Mobile Modular Robot(MMR) which is composed of a mobile platform and a modular manipulator, the kinematics analysis is made for both the mobile platform and the upper manipulator. Then the dynamic equations are formulated considering the nonholonomic constraints and the controller is proposed with both the redundancy resolution and optimization of null space motion. A 5-DOF modular manipulator mounted on the mobile platform is built up to verify the proposed control law in the simulation and the results show the effectiveness of the proposed method.

1 Introduction

The mobile modular robot(MMR) is normally composed of a m-wheeled holonomic or nonholonomic mobile platform and a n-DoF onboard modular manipulator. In this paper, a MMR is constructed with a 5-DOF modular manipulator and the mobile platform, which is supported by two driving wheels and one caster wheel. Since the MMR is able to perform manipulation tasks in a much larger workspace than a fixed-base manipulator, it will provide more functions in many complicated human robot interaction environment such as opening the door [1] [2], pushing the objects, interacting with human beings and so on.

However, dynamic modeling and trajectory following become difficult to achieve due to their interactive motions. Therefore it is necessary and important to understand how to properly and effectively coordinate the motions of the mobile platform and the upper manipulators because of the dynamic interactions between them. Also the external force will be combined together when the MMR interacts with the environment.

Some related works can be found on studying the dynamics of mobile robots, which are highlighted in the follows. The impedance control method is applied to control MR Helper [3] and the control algorithm is implemented in the experimental system for handling an object in cooperation with the robot and the human. A systematic modelling of the nonholonomic MMR is proposed and derived to generalize the standard definition of manipulability to the case of mobile manipulators [4]. However, it is only actuated and operated in the horizontal plane. A general method is presented to establish the dynamic equation of a mobile manipulator system by utilizing forward recursive formulation for open-loop

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 776–791, 2009.

multi-body systems in [5]. A dynamic model is established for the multi-link mobile manipulator [6], which explicitly takes into account the dynamic interactions between the manipulator and the mobile platform. A nonlinear feedback controller is designed for the mobile manipulator that is capable of fully compensating the dynamic interactions [7]. However, their dynamic model is somewhat inaccurate since the dynamic influence of the caster wheel is neglected. The dynamic modeling and various control methods are investigated on the mobile manipulator in [8] [9], but redundancy is not considered.

In this paper, the whole system has one redundant DoF for the spatial motion with both desired position and orientation while the redundancy makes the manipulator be able to implement the multi-tasks, which can also be referred in previous related works [10]- [13]. The other parts of the paper are organized as follows. In Section II, the kinematics of the mobile platform atop the modular manipulator is given out. In Section III, the constraints of the MMR and dynamic modeling are developed with the dynamic equations of the entire system formulated by Lagrangian equations and the model-based controller is designed. Simulations are given out in section IV. Finally, the conclusion remarks are obtained in the last section.

2 Modeling of the Mobile Modular Robot

The configuration of the MMR is a 5-DOF modular manipulator mounted on a mobile platform, which is supported by two driving wheels and one caster wheel as shown in Fig. 1 and Fig. 2.

Three assumptions are adopted in the modeling of the MMR system.

- The 5-DOF modular manipulator is installed vertically on the planar platform.
- There are no slipping and no sideways between the wheels and the ground.
- Every link of the robot is rigid.

2.1 Kinematics of Mobile Platform

The kinematics analysis of the platform, driving wheels, castor wheel and the bracket had been discussed in another work [12] by a different method used.

Fig. 1. The mobile modular robot

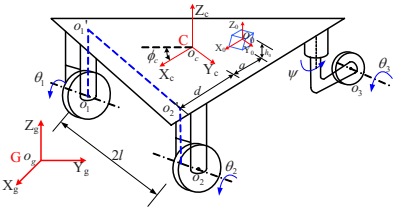


Fig. 2. The configuration of mobile platform

The coordinate systems are defined as follows: $G - X_g Y_g Z_g$ forms an inertial base frame(global frame)fixed on the motion plane and $C - X_c Y_c Z_c$ is a frame fixed on the mobile platform. O_c is selected at the midpoint of the line segment connecting the two fixed wheel centers; and Y_c is along the coaxial-line of the two fixed wheels. The heading angle ϕ_c determines the posture of the mobile platform, while θ_1 and θ_2 are the rotating angles of the left and right wheels as shown in Fig. 2.

The mobile platform can be treated as a special module, which can move both on the plane and rotate about the vertical axis. Here locate the point O_c with the coordinate of (x_c, y_c, z_c) and then the transformation matrix between the frame $G - X_g Y_g Z_g$ and the frame $C - X_c Y_c Z_c$ can be expressed by

$$
{}^{G}_{C}\mathrm{T} = \begin{bmatrix} cos\phi_c & -sin\phi_c & 0 & x_c \\ sin\phi_c & cos\phi_c & 0 & y_c \\ 0 & 0 & 1 & z_c \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{1}
$$

The frame $M_0 - X_0 Y_0 Z_0$ is attached to the first module of the manipulator as shown both in Fig. 2 and Fig. 3. Since the modular manipulator is mounted on the mobile platform, it can be assumed that the relative position of the frame $M_0 - X_0 Y_0 Z_0$ and the frame$C - X_c Y_c Z_c$ is unvaried. Then the transformation matrix between the frame $M_0 - X_0 Y_0 Z_0$ and the frame $C - X_c Y_c Z_c$ can be expressed by

$$
{}^{C}_{0}\mathrm{T} = \begin{bmatrix} 1 & 0 & 0 & -a \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & h_0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2}
$$

where a is the distance between the point of O_c and O_0 in x direction and h_0 is the distance in z direction.

2.2 Kinematics of the Whole Body

Since the kinematic analysis incorporates the interactive motions between the modular manipulator and the mobile platform, the position and pose of end-effector co-

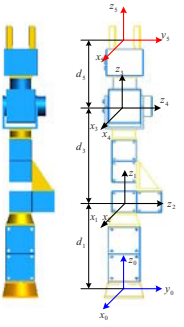


Fig. 3. The frame assignments of the 5-DOF manipulator

Table 1. LINK PARAMETERS OF THE MHR 5-DOF ARM

Frame	Link	α_{i-1}	a_{i-1}	d_i	q_i	$q_i min$	$q_i max$
Frame 0-Frame 1	1	0	0	d_1	q_1	−360	360
Frame 1-Frame 2	2	−90	0	0	q_2	30	150
Frame 2-Frame 3	3	90	0	d_3	q_3	−360	360
Frame 3-Frame 4	4	−90	0	0	q_4	30	330
Frame 4-Frame 5	5	90	0	d_5	q_5	−360	360

Note: The tool point depends on the tool type at hand.

ordinates can be derived with each known homogeneous transformation Denavit-Hartenberg matrix by forward kinematics method [10]. The D-H parameters are shown in Table 1.

$$ {}^{G}_{5}\mathrm{T} = {}^{G}_{C}\mathrm{T}\,{}^{C}_{0}\mathrm{T}\,{}^{0}_{1}\mathrm{T}\,{}^{1}_{2}\mathrm{T}\,{}^{2}_{3}\mathrm{T}\,{}^{3}_{4}\mathrm{T}\,{}^{4}_{5}\mathrm{T} \tag{3} $$

It can be expressed as

$$ {}^{G}_{5}\mathrm{T} = \begin{bmatrix} n & s & a & p \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{4} $$

where the column vector n, s, a, p can be easily obtained from the transform matrix and the detailed forms are neglected here.

3 Dynamic Control of Mobile Modular Robot

3.1 Constraints

The mobile platform can either move in the direction of X_c or rotate about the Z_C, but the instantaneous motion in the direction Y_C is constrained because of the nonholonomic behavior of the wheels. Then the first constraint restricts the velocity of the mobile platform to be zero in the direction ($\overrightarrow{o_1 o_2}$) through two axial center lines of driving wheels.

$$ -\dot{x}_c sin\phi_c + \dot{y}_c cos\phi_c - d\dot{\phi}_c = 0 \tag{5} $$

The other two constraints are included with assumption of no-slipping of each rolling wheel in the forward directions, and the velocity of the driving wheels can be expressed as:

$$\dot{x}_c cos\phi_c + \dot{y}_c sin\phi_c - l\dot{\phi}_c = r\dot{\theta}_1 \tag{6}$$

$$\dot{x}_c cos\phi_c + \dot{y}_c sin\phi_c + l\dot{\phi}_c = r\dot{\theta}_2 \tag{7}$$

Let $q_m = \begin{bmatrix} x_c & y_c & \phi_c & \theta_1 & \theta_2 \end{bmatrix}^T$ represent the basic coordinates (q_m) of the MMR platform, so the above three constraints of velocities can be given in the matrix form

$$A(q_m)\dot{q}_m = 0 \tag{8}$$

where $A(q_m) = \begin{bmatrix} -sin\phi_c & cos\phi_c & -d & 0 & 0 \\ cos\phi_c & sin\phi_c & -l & -r & 0 \\ cos\phi_c & sin\phi_c & l & 0 & -r \end{bmatrix}$ and r is the radius of the driving wheels.

The independent velocities of two driving wheels are chosen as the generalized coordinates ($\dot{q}_w$) because the nonholonomic constrained mobile platform owns only 2-DOF under planar motion assumption and there is such a relation

$$\dot{q}_m = S(q_m)\dot{q}_w \tag{9}$$

where $\dot{q}_w = \begin{bmatrix} \dot{\theta}_1 & \dot{\theta}_2 \end{bmatrix}^T$, and$S(q_m)$ is a 5×2 full rank matrix, whose columns are in the null space of $A(q_m)$. One choice of $S(q_m)$ is defined as:

$$S(q_m) = \begin{bmatrix} \rho(lcos\phi_c + dsin\phi_c) & \rho(lcos\phi_c - dsin\phi_c) \\ \rho(lsin\phi_c - dcos\phi_c) & \rho(lsin\phi_c + dcos\phi_c) \\ -\rho & \rho \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Here $\rho = \frac{r}{2l}$.

With the additional DoFs of the upper modular manipulator, the extended coordinates of the whole system is composed of the wheeled mobile platform and the modular manipulator, whose form is $q_e = \begin{bmatrix} q_m^T & q_1 & \cdots & q_5 \end{bmatrix}^T$.

Now it can be parameterized in terms of the independent velocities $q = \begin{bmatrix} q_w^T & q_1 & \cdots & q_5 \end{bmatrix}^T$

$$\dot{q}_e = S\dot{q} \tag{10}$$

where

$$S = \begin{bmatrix} S(q_m) & 0_{5\times3} \\ 0_{5\times2} & I_{5\times5} \end{bmatrix}$$

Combining the Eq.8 and Eq. 9 together, the constraint equation can be easily obtained as:

$$A(q_m)S(q_m) = 0 \tag{11}$$

3.2 Dynamics Equation

For the whole mobile modular robot, the dynamics can be obtained using energy-based Lagrange method, which is based on the principle of virtual work function.

$$L(q, \dot{q}) = T - V \tag{12}$$

where T denotes the kinetic energy of the whole system, as the sum of the individual kinetic-energy expressions. V denotes the potential energy of the whole system, as the sum of the individual expression, for every element storing potential energy.

The kinetic energy of the whole system can be written as the sum of two parts of mobile platform and upper manipulator, $T = T_p + T_m$, whose former part contains the kinetic energy of the cart, two driving wheels and the castor wheel. The potential energy of the whole system contains only the part of modular manipulator, $P = \sum_{i=1}^{5} P_i = \sum_{i=1}^{5} m_i g^T r_{ci}$,(where g is the vector at the direction of gravity in the inertial frame and the vector r_{ci} gives the coordinates of the center of ith module) for the mobile platform will not produce the change of potential energy during the planar motion.

As discussed above, $q_e = \left[q_m^T \; q_1 \cdots q_5 \right]^T$ denotes the extended coordinates, with considering the external forces/torques for the whole system with constraints($A(q_m)\dot{q}_m = 0$) in three dimensional operational space, the generalized dynamic equation of motion can be written as

$$\begin{aligned} M(q_e)\ddot{q}_e + C(q_e, \dot{q}_e)\dot{q}_e + G(q_e) &= E\tau + E_2 F - A^T \lambda \\ A(q_m)\dot{q}_m &= 0 \end{aligned} \tag{13}$$

where $M(q_e)$ is the nonsingular symmetric inertia matrix, $C(q_e, \dot{q}_e)$ is the vector that implicitly includes centrifugal, Coriolis and viscous friction, $G(q_e)$ is the gravity terms, E is the matrix maps the actuator torque to the joint torques, E_2 is the matrix that maps the task space F to the joint space, A^T is the above constraint matrix and λ is the vector of Lagrange multipliers corresponding to the aforementioned constraint equations.

$$S^T(M(q_e)\ddot{q}_e + C(q_e, \dot{q}_e)\dot{q}_e + G(q_e)) = S^T E\tau + S^T E_2 F - S^T A^T \lambda \tag{14}$$

With the equation of (11) and the derivative of equation of (10),

$$\ddot{q}_e = \dot{S}\dot{q} + S\ddot{q} \tag{15}$$

The motion equation of MMR can be written as

$$\bar{M}\ddot{q} + \bar{C}\dot{q} + \bar{G} = \bar{\tau} + \bar{F} \tag{16}$$

where $\bar{M} = S^T M S$, $\bar{V} = S^T V S + S^T M \dot{S}$, $\bar{G} = S^T G$, $\bar{\tau} = S^T E\tau$, and $\bar{F} = S^T E_2 F$).

3.3 Dynamic Redundancy Resolution

The whole MMR is a kinematically redundant system as there are two DoFs for the mobile platform and five DoFs for the upper modular manipulator. A nonempty null space $N(J)$ exists which is available to set up systematic procedures for an effective handling of redundant degrees of freedom and one inverse solution based on inertia-weighted pseudo-inverse can be written as

$$\dot{q} = J^{\dagger}(q)\dot{x} + (I - J^{\dagger}(q)J(q))\dot{q}_0 \tag{17}$$

where $\dot{x}$ is the task space velocity, $J^{\dagger}$ is the inertia-weighted pseudo-inverse($J^{\dagger} = M^{-1}J^T(JM^{-1}J^T)^{-1}$) and the matrix ($N = I - J^{\dagger}J$) is a projector of the arbitrarily chosen joint vector $\dot{q}_0$ onto $N(J)$.

Differentiating(17), the accelerations of joints can be found:

$$\ddot{q} = J^{\dagger}(q)(\ddot{x} - \dot{J}\dot{q}) + N\ddot{q}_0 \tag{18}$$

With the similar method of redundancy resolution, the forces of redundant systems can be written as

$$\tau = J^T(q)F + N^T\tau_0 \tag{19}$$

From the (16), the constrained torque vector has the form

$$\bar{\tau} = \bar{M}\ddot{q} + \bar{C}\dot{q} + \bar{G} - \bar{F} \tag{20}$$

Combining the (19) and (20) together with the above equation $N^T = I - J^T J^{\dagger T}$,

$$J^T(q)F + N^T\tau_0 = (J^T J^{\dagger T} + N^T)(\bar{M}\ddot{q} + \bar{C}\dot{q} + \bar{G} - \bar{F}) \tag{21}$$

Substituting (18) in (21), as the main task corresponds to the task-space motion while the secondary task denotes the null-space motion [14], the equation can be decoupled

$$J^T(q)F + N^T\tau_0 = \tau_T + \tau_N + F_C \tag{22}$$

Among them, τ_T denotes the forces in the task space, $\tau_T = (J^T J^{\dagger T})(\bar{M}(J^{\dagger}(\ddot{x} - \dot{J}\dot{q})) + \bar{C}\dot{q} + \bar{G} - \bar{F})$, τ_N denotes the torques in the null space, $\tau_N = N^T(\bar{M}N\ddot{q}_0 + \bar{C}\dot{q} + \bar{G} - \bar{F})$, and F_C represents the coupling torques and forces, $F_C = J^T J^{\dagger T}\bar{M}N\ddot{q}_0 + N^T\bar{M}J^{\dagger}(q)(\ddot{x} - \dot{J}\dot{q})$.

Since the inertia matrix M is identity, the inertia-weighted pseudo-inverse $J^{\dagger}$ can be reduced to the pseudo-inverse, $J^{\#} = J^T(JJ^T)^{-1}$, thus it can be easily known that $F_C = 0$.

3.4 Controller Design

With the decoupling of the task space and null space, one can choose the appropriate controller for both subspaces. There are always some uncertainties in the system dynamics, such as nonlinear frictions, payload variations, flexibilities of links and so on, and then the inexact dynamic parameters can not guarantee to

cancel the nonlinear terms. A hybrid impedance controller in the task space is applied here [17], which can also be found in our related work [15],

$$u = \ddot{x}^d + B_d\dot{e} + K_d e + K_f(F_E^d - F_E) \tag{23}$$

where $\ddot{x}^d$ is the desired acceleration, e and $\dot{e}$ are the position error and velocity error, B_d, K_d and K_f are the gains chosen appropriately with some techniques, F_E^d and F_E are the desired and actual forces of the end-effector. Then the desired impedance relation can be written as

$$\ddot{e} + B_d\dot{e} + K_d e = -K_f(e_f) \tag{24}$$

From (22), it can easily get the equation,

$$\tau = \tau_P + \tau_N + C\dot{q} + G \tag{25}$$

where τ_P and τ_N are the vectors of task space and null space joint torques respectively, C represents the coriolis and centripetal matrix and G denotes the gravity force vector.

The aim of the end effector control law is to track the desired generalized task coordinates x_d, which includes position and force. Let the control law [19] be given as

$$\tau = J^T\{\Lambda(u - \dot{J}\dot{q}) + F_e\} + C\dot{q} + G + \tau_N \tag{26}$$

where $\Lambda \triangleq [JM^{-1}J^T]^{-1}$ denotes the pseudo kinetic energy matrix [18].

The primary task is assumed to control the motion and force interactions of the end-effector with the attached payload/external environment while the secondary task is assumed to control the system surplus DoF. This nonlinear control method separates the equations of motion into the task-space motion and the null-space motions and allows them to be controlled independently. Combining (23), (24), (25) and (26), the task space joint torque has the form

$$\tau_P = J^T\{\Lambda(\ddot{x}_d - \dot{J}\dot{q}) + B_d\dot{e}_p + K_d e_p + K_f e_f + F_e\} \tag{27}$$

A proper choice of τ_N enables the manipulator to achieve secondary sub-task without affecting the task space dynamics, like force and trajectory tracking of the manipulator's end-effector. The null space torque vector will be obtained as

$$\tau_N = J_N^T\{\Lambda_N(\ddot{x}_{Nd} - \dot{J}_N\dot{q}) + B_N\dot{e}_N\} \tag{28}$$

where $\dot{e}_N$ is the velocity tracking error of the null space motion, which can be written as $\dot{e}_N = \dot{x}_{N_d} - \dot{x}_N$. One has the freedom to choose any control routine in the null-space. The control scheme is designed as shown in Fig. 4.

3.5 Optimization of Null Space Motion

In addition to the primary task, the secondary task is to minimize a desired cost function. Since the null space has many forms, a proper choice of $N(J)$ will achieve the avoidance of rank deficient matrix.

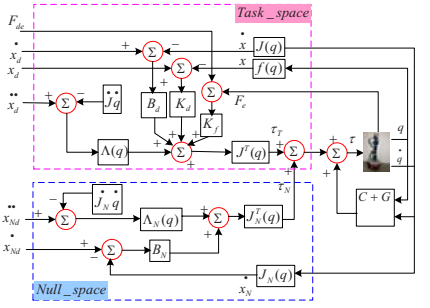


Fig. 4. The control scheme

The additional task is represented as

$$Z(q) = N(q)^T \nabla \Phi(q) \tag{29}$$

Therefore, the Jacobian matrix of null space can be written as

$$J_N = \frac{\partial Z_i(q)}{\partial q_i} = \frac{\partial N(q)^T \nabla \Phi(q)}{\partial q} \tag{30}$$

where $\Phi(q)$ is a scalar objective function of the joint variables and $\nabla \Phi(q)$ is the vector function representing the gradient of Φ.

The kinematic constraint function considered [16] [15] is the joint limit avoidance as $\Phi(q)=\frac{1}{n}\sum_{i=1}^{n}(\frac{q_i-\bar{q}_i}{q_{iM}-q_{im}})^2$. Here $q_{iM}(q_{im})$ denotes the maximum(minimum) limit for q_i and $\bar{q}_i$ the middle value of the joint range, and thus the redundancy resolution problem is to define a joint trajectory which optimizes this objective function subject to the end-effector position.

4 Cases Study and Simulations

4.1 Simulation Setup

A 5-DOF spatial modular manipulator is built up with the same mass and inertia as the real Powercube modular manipulator in our lab and the frame assignment and its transformations are given out in Fig. 3. The corresponding D-H parameters of joints and the maximum and minimum joint values are shown in Table 1. The link length parameters are $d1 = 0.305(m)$, $d3 = 0.308(m)$ and $d5 = 0.265(m)$. The mass and inertia parameters of the four links are listed in the Table 2. The mobile platform is also built up with the same mass as our real Activmedia mobile robot and some parameters are $xc = 0.5; yc = 0.5; zc = 0.24; l = 0.215; d = 0.065; a = 0.045; h0 = 0; r = 0.095$; as shown in Fig. 2. The maximum and minimum limits for the joints range of five modules can be found in Table 1.

Table 2. The mass and inertia parameters of the MMR(Unit: kg, m)

Link	m_i	I_{xx}	I_{yy}	I_{zz}
(cart and wheels) 0	8.0	0.9013	0.5141	0.6453
(1st module) 1	2.8	0.0438	0.0121	0.0468
(2nd module) 2	1.8	0.0199	0.0147	0.0287
(3rd module) 3	1.6	0.0159	0.0052	0.0159
(4th module) 4	1.8	0.0193	0.0096	0.0229
(5th module) 5	1.4	0.0158	0.0074	0.0186

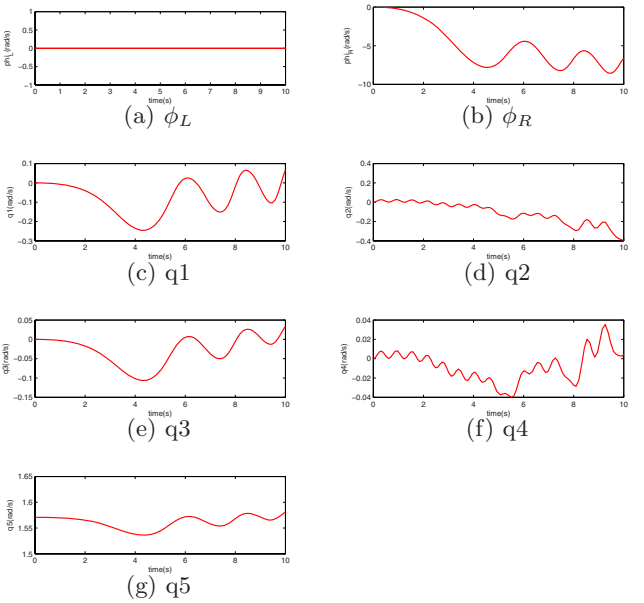


Fig. 5. The rotating angles of two driven wheels and five joints angles of modular manipulator

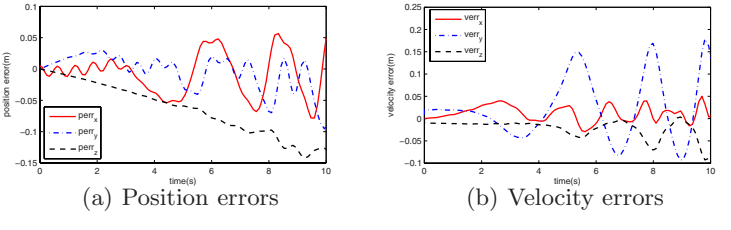


Fig. 6. The position errors and velocity errors between the desired and the actual trajectory

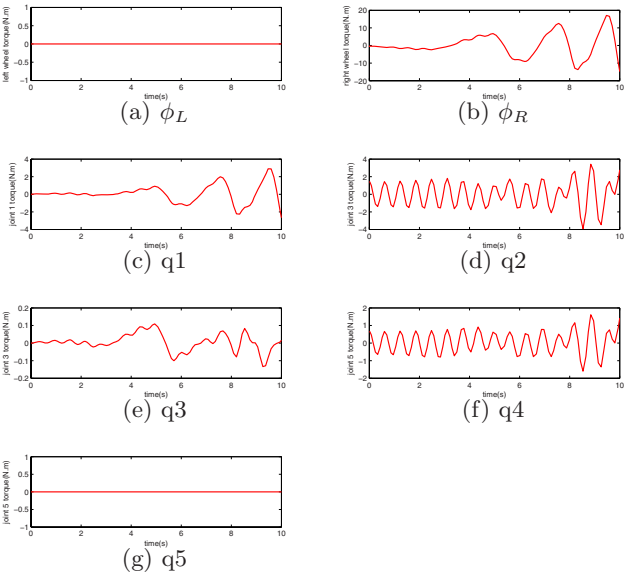


Fig. 7. The torques of two driven wheels and five joints of modular manipulator

4.2 Simulation Analysis

The robot is asked to track the desired trajectory considering the feedback of contact forces between the robot end-effector and the environment. Here a trajectory is defined as $p_d(t) = \begin{bmatrix} 0.560 - 0.1cos(t/10) \\ 0.500 + 0.1sin(t/10) \\ 1.218 - 0.1e^{t/10} \end{bmatrix}$ to simulate some kinds of task while the angle posture keeps unvaried. The initial posture of the upper modular manipulator is set with joint initial values $q_0 = \begin{bmatrix} 0 & 0 & 0 & 0 & \pi/2 \end{bmatrix}$ and the initial state of the mobile platform is still, which corresponds to the initial position $p = \begin{bmatrix} 0.560 & 0.500 & 1.118 \end{bmatrix}^T$ and both the joints initial velocities and wheels velocities are set to be zero at the very beginning. The Jacobian matrix of the manipulator is given out in the Appendix. For many choices of the null space Jacobian matrix [15], we choose one here as, $J_N = \begin{bmatrix} 1/2d3(1/c3)(-(c1/(ac)) + (2s1)/r)s4 \\ (d3(1/c3)(rc1 + 2acs1)s4)/(2acr) \\ ((-d3c1 + a(1/s2))(1/c3)s4)/a \\ 0 \\ -c4 - (c2/s2)(1/c3)s4 \\ -s4(s3/c3) \\ 1 \end{bmatrix}^T$.

The stiffness matrix is $K_d = diag\begin{bmatrix} 500 & 500 & 500 & 200 & 200 & 200 \end{bmatrix}$, the damping matrices are $B_d = diag\begin{bmatrix} 50 & 50 & 50 & 10 & 10 & 10 \end{bmatrix}$ and $B_N = diag\begin{bmatrix} 30 \end{bmatrix}$, and the force error gain is $K_f = diag\begin{bmatrix} 2 & 2 & 2 & 2 & 2 & 2 \end{bmatrix}$.

During the simulation, the left driven wheel is set as still as shown in Fig. 5(a) and then it is easy to know the rotation of the mobile platform from the difference between left and right wheels velocities. The rotating angles of the right wheel is shown in Fig. 5(b) and the joints variables of modular manipulator are shown in Fig. 5(c) to Fig. 5(g) respectively. The position errors and velocity errors between the desired and the actual trajectory are shown in Fig. 6(a) and Fig. 6(b). The torques on the two driven wheels and the five module joints are shown in Fig. 7. Note that the rotating angle (ϕ_c) of the mobile platform is the relation of the left and right wheel velocities, and then ϕ_c is used when solving the dynamics as shown in the Inertia Matrix, the Coriolis Matrix and the Gravity Matrix in the Appendix.

5 Conclusion

This paper performs a detailed kinematic analysis on both the mobile platform and the atop modular manipulator of the MMR. The dynamic equations are formulated based on the Lagrangian equations with consideration of the nonholonomic constraints of the mobile platform. The dynamic redundancy resolutions are given out and the model-based controller is designed considering both the secondary task and optimization function. The simulations are made and the results and analysis are followed. However, because of many unexpected factors and inexact dynamic parameters(nonlinear frictions etc.), the dynamic equations and controller need to be improved in next step work and adaptive control method will be used in future as well.

Acknowledgement

This work is supported by the Research Committee of University of Macau under grant no. UL016/08-Y2/EME/LYM01/FST and Macao Science and Technology Development Fund under Grant no. 016/2008/A1.

References

1. Nguyen, H., Kemp, C.C.: Bio-inspired Assistive Robotics: Service Dogs as a Model for Human-Robot Interaction and Mobile Manipulation. In: The 2nd Biennial IEEE/RAS Int. Conf. on Biomedical Robotics and Biomechatronics, Scottsdale, AZ, USA, pp. 542–549 (2008)
2. Nagatani, K., Yuta, S.: Designing strategy and implementation of mobile manipulator control system for opening door. IEEE Int. Conf. on Robotics and Automation 3, 2828–2834 (1996)
3. Kazuhiro, K., Manabu, S., Norihide, K.: Mobile robot helper. In: Proc. of the IEEE Int. Conf. on Robotics and Automation, San Francisco, April 2000, pp. 583–588 (2000)
4. Bayle, B., Fourquet, J.Y., Renaud, M.: Manipulability of Wheeled Mobile Manipulators: Application to Motion Generation. Int. J. Robotics Research 22, 565–581 (2003)

5. Yu, Q., Chen, I.-M.: A general approach to the dynamics of nonholonomic mobile manipulator systems. Journal of Dynamic Systems, Measurement, and Control, Trans. of the ASME 124, 512–521 (2002)
6. Yamamoto, Y., Yun, X.: A Modular Approach to Dynamic Modeling of a Class of Mobile Manipulators. IEEE Transactions of Robtics and Automation 12(2), 41–48 (1997)
7. Cheng, M., Tsai, C.: Dynamic modeling and tracking control of a nonholonomic wheeled mobile manipulator with two robotic arms. In: IEEE Conf. on Decision and Control, Hawaii, vol. 3, pp. 2932–2937 (2003)
8. Li, Y., Liu, Y.: Dynamic modeling and adaptive neural-fuzzy control for nonholonomic mobile manipulators moving on a slope. Int. J. Contl. Autom. Systems 4(2), 1–7 (2006)
9. Li, Y., Liu, Y.: Real-time Tip-over Prevention and Path Following Control for Redundant Nonholonomic Mobile Modular Manipulators via Fuzzy and Neural-Fuzzy Approaches. Journal of Dynamic System, Measurement, and Control, Trans. of ASME 128, 753–764 (2006)
10. Wang, J.-G., Li, Y.: Inverse Kinematic Analysis for the Arm of a Mobile Humanoid Robot Based on the closed-loop algorithm. In: IEEE Int. Conf. on Information and Automation, Zhuhai/Macau, pp. 516–521 (2009)
11. Wang, J.-G., Li, Y.: Comparative Case Studies for the Inverse Kinematics of Redundant Manipulator based on Repetitive Tracking Tasks. In: IEEE Int. Conf. on Automation and Logistics, Shenyang, China, pp. 164–169 (2009)
12. Wang, J.-G., Li, Y.: Dynamic Modeling of a Mobile Humanoid Robot. In: IEEE Int. Conf. on Robotics and Biomimetics, Bangkok, Thailand, pp. 639–644 (2009)
13. Wang, J.-G., Li, Y.: Kinematics and tip-over stability analysis for a mobile humanoid robot moving on a slope. In: IEEE Int. Conf. on Automation and Logistics, Qingdao, China, pp. 2426–2431 (2008)
14. White, G.D., Bhatt, R.M., Krovi, V.N.: Dynamic redundancy resolution in a nonholonomic wheeled mobile manipulator. Robotica 25, 147–156 (2007)
15. Wang, J.-G., Li, Y.: Impedance Control of a Spatial Redundant Manipulator Used for Relaxing Muscle Fatigue. In: IEEE Int. Conf. on Mechatronics and Automation, Changchun, Jilin, China, pp. 2799–2804 (2009)
16. Liegeois, A.: Automatic Supervisory Control of the Configuration and Behavior of Multibody Mechanisms. IEEE Trans. Systems, Man, and Cybernetics 7(12), 868–871 (1977)
17. Spong, M.W., Hutchinson, S., Vidyasagar, M.: Robot Modeling and Control. John Wiley & Sons, Inc., Chichester (2006)
18. Khatib, O.: A Unified Approach for Motion and Force Control of Robot Manipulators: The Operational Space Formulation. IEEE Journal of Robotics and Automation 3, 43–53
19. Nemec, B., Zlajpah, L.: Null space velocity control with dynamically consistent pseudo-inverse. Robotica 18(5), 513–518 (2000)

A Appendix

A.1 The Jacobian Matrix

$$J_{11} = (1/2)rc0 - c(-d5c01s3s4 + as0 - ((d3 + d5c4)s2 + d5c2c3s4)s01);$$
$$J_{12} = (1/2)rc0 + c(-d5c01s3s4 + as0 - ((d3 + d5c4)s2 + d5c2c3s4)s01);$$

$J_{13} = -d5c01s3s4 - ((d3 + d5c4)s2 + d5c2c3s4)s01;$
$J_{14} = c01(c2(d3 + d5c4) - d5c3s2s4); J_{15} = -d5s4(c2c01s3 + c3s01);$
$J_{16} = d5c01(c2c3c4 - s2s4) - d5c4s3s01; J_{17} = 0;$
$J_{21} = (1/2)rs0 - c(-ac0 + c01((d3 + d5c4)s2 + d5c2c3s4) - d5s3s4s01);$
$J_{22} = (1/2)rs0 + c(-ac0 + c01((d3 + d5c4)s2 + d5c2c3s4) - d5s3s4s01);$
$J_{23} = c01((d3 + d5c4)s2 + d5c2c3s4) - d5s3s4s01;$
$J_{24} = (c2(d3 + d5c4) - d5c3s2s4)s01; J_{25} = d5s4(c3c01 - c2s3s01);$
$J_{26} = d5c4c01s3 + d5(c2c3c4 - s2s4)s01; J_{27} = 0; J_{31} = 0; J_{32} = 0; J_{33} = 0;$
$J_{34} = -(d3 + d5c4)s2 - d5c2c3s4; J_{35} = d5s2s3s4; J_{36} = -d5(c3c4s2 + c2s4);$
$J_{37} = 0; J_{41} = 0; J_{42} = 0; J_{43} = 0; J_{44} = -s01; J_{45} = c01s2;$
$J_{46} = -c2c01s3 - c3s01; J_{47} = c01(c4s2 + c2c3s4) - s3s4s01; J_{51} = 0;$
$J_{52} = 0; J_{53} = 0; J_{54} = c01; J_{55} = s2s01;$
$J_{56} = -s1(c2c0s3 + c3s0) + c1(c3c0 - c2s3s0); J_{57} = c01s3s4 + (c4s2 + c2c3s4)s01;$
$J_{61} = -c; J_{62} = c; J_{63} = 1; J_{64} = 0; J_{65} = c2; J_{66} = s2s3; J_{67} = c2c4 - c3s2s4;$ (31)

A.2 The Inertia Matrix

$$
\begin{aligned}
M_{11} = {} & Iz1 + Iz2 + Iz3 + Iz4 + Iz5 + Iz6 + a^2m2c0^2 + a^2m3c0^2 \\
& + m4(-ac0 + d3c01s2)^2 + m5(-ac0 + d3c01s2)^2 + a^2m2s0^2 + a^2m3s0^2 \\
& + m4(as0 - d3s2s01)^2 + m5(as0 - d3s2s01)^2 + m6(-ac0 + c01((d3 + d5c4)s2 \\
& + d5c2c3s4) - d5s3s4s01)^2 + m6(-d5c01s3s4 + as0 - ((d3 + d5c4)s2 \\
& + d5c2c3s4)s01)^2; \\
M_{12} = {} & Iz2 + Iz3 + Iz4 + Iz5 + Iz6 + d3m4c01s2(-ac0 + d3c01s2) \\
& + d3m5c01s2(-ac0 + d3c01s2) - d3m4s2s01(as0 - d3s2s01) \\
& - d3m5s2s01(as0 - d3s2s01) + m6(c01((d3 + d5c4)s2 + d5c2c3s4) \\
& - d5s3s4s01)(-ac0 + c01((d3 + d5c4)s2 + d5c2c3s4) - d5s3s4s01) \\
& + m6(-d5c01s3s4 - ((d3 + d5c4)s2 + d5c2c3s4)s01)(-d5c01s3s4 + as0 \\
& - ((d3 + d5c4)s2 + d5c2c3s4)s01); \\
M_{13} = {} & d3m4c2(-ac0 + d3c01s2)s01 + d3m5c2(-ac0 + d3c01s2)s01 \\
& + d3m4c2c01(as0 - d3s2s01) + d3m5c2c01(as0 - d3s2s01) \\
& + m6(c2(d3 + d5c4) - d5c3s2s4)s01(-ac0 + c01((d3 + d5c4)s2 + d5c2c3s4) \\
& - d5s3s4s01) + m6c01(c2(d3 + d5c4) - d5c3s2s4)(-d5c01s3s4 \\
& + as0 - ((d3 + d5c4)s2 + d5c2c3s4)s01); \\
M_{14} = {} & Iz4c2 + Iz5c2 + Iz6c2 + d5m6s4(c3c01 - c2s3s01)(-ac0 \\
& + c01((d3 + d5c4)s2 + d5c2c3s4) - d5s3s4s01) - d5m6s4(c2c01s3 \\
& + c3s01)(-d5c01s3s4 + as0 - ((d3 + d5c4)s2 + d5c2c3s4)s01); \\
M_{15} = {} & Iz5s2s3 + Iz6s2s3 + m6(d5c01(c2c3c4 - s2s4) - d5c4s3s01) \\
& (-d5c01s3s4 + as0 - ((d3 + d5c4)s2 + d5c2c3s4)s01) \\
& + m6(-ac0 + c01((d3 + d5c4)s2 + d5c2c3s4) - d5s3s4s01) \\
& (d5c4c01s3 + d5(c2c3c4 - s2s4)s01);
\end{aligned}
$$

$$
\begin{aligned}
M_{16} = {} & Iz6(c2c4 - c3s2s4);\ M_{21} = M_{12};\\
M_{22} = {} & Iz2 + Iz3 + Iz4 + Iz5 + Iz6 + d3^2m4c01^2s2^2 + d3^2m5c01^2s2^2\\
& + d3^2m4s2^2s01^2 + d3^2m5s2^2s01^2 + m6(c01((d3 + d5c4)s2 + d5c2c3s4)\\
& - d5s3s4s01)^2 + m6(-d5c01s3s4 - ((d3 + d5c4)s2 + d5c2c3s4)s01)^2;\\
M_{23} = {} & m6(c2(d3 + d5c4) - d5c3s2s4)s01(c01((d3 + d5c4)s2 + d5c2c3s4)\\
& - d5s3s4s01) + m6c01(c2(d3 + d5c4)\\
& - d5c3s2s4)(-d5c01s3s4 - ((d3 + d5c4)s2 + d5c2c3s4)s01);\\
M_{24} = {} & Iz4c2 + Iz5c2 + Iz6c2 + d5m6s4(c3c01 - c2s3s01)\\
& (c01((d3 + d5c4)s2 + d5c2c3s4) - d5s3s4s01)\\
& - d5m6s4(c2c01s3 + c3s01)(-d5c01s3s4 - ((d3 + d5c4)s2 + d5c2c3s4)s01);\\
M_{25} = {} & Iz5s2s3 + Iz6s2s3 + m6(d5c01(c2c3c4 - s2s4) - d5c4s3s01)\\
& (-d5c01s3s4 - ((d3 + d5c4)s2 + d5c2c3s4)s01) + m6(c01((d3 + d5c4)s2\\
& + d5c2c3s4) - d5s3s4s01)(d5c4c01s3 + d5(c2c3c4 - s2s4)s01);\\
M_{26} = {} & Iz6(c2c4 - c3s2s4);\ M_{31} = M_{13};\ M_{32} = M_{23};\\
M_{33} = {} & Iy3c01^2 + Iy4c01^2 + Iy5c01^2 + Iy6c01^2 + d3^2m4c2^2c01^2 + d3^2m5c2^2c01^2\\
& + d3^2m4s2^2 + d3^2m5s2^2 + m6((-d3 - d5c4)s2 - d5c2c3s4)^2\\
& + m6c01^2(c2(d3 + d5c4) - d5c3s2s4)^2 + Ix3s01^2 + Ix4s01^2 + Ix5s01^2\\
& + Ix6s01^2 + d3^2m4c2^2s01^2 + d3^2m5c2^2s01^2\\
& + m6(c2(d3 + d5c4) - d5c3s2s4)^2s01^2;\\
M_{34} = {} & d5m6s2s3s4((-d3 - d5c4)s2 - d5c2c3s4) - Ix4c01s2s01 - Ix5c01s2s01\\
& - Ix6c01s2s01 + Iy4c01s2s01 + Iy5c01s2s01 + Iy6c01s2s01\\
& - d5m6c01s4(c2(d3 + d5c4) - d5c3s2s4)(c2c01s3 + c3s01)\\
& + d5m6s4(c2(d3 + d5c4) - d5c3s2s4)s01(c3c01 - c2s3s01);\\
M_{35} = {} & -d5m6(c3c4s2 + c2s4)((-d3 - d5c4)s2 - d5c2c3s4) + Iy6c01(-s1(c2c0s3\\
& + c3s0) + c1(c3c0 - c2s3s0)) - Ix5s01(-c2c01s3 - c3s01) - Ix6s01(-c2c01s3\\
& - c3s01) + Iy5c01(c3c01 - c2s3s01) + m6c01(c2(d3 + d5c4) - d5c3s2s4)\\
& (d5c01(c2c3c4 - s2s4) - d5c4s3s01) + m6(c2(d3 + d5c4)\\
& - d5c3s2s4)s01(d5c4c01s3 + d5(c2c3c4 - s2s4)s01);\\
M_{36} = {} & -Ix6s01(c01(c4s2 + c2c3s4) - s3s4s01)\\
& + Iy6c01(c01s3s4 + (c4s2 + c2c3s4)s01);\ M_{41} = M_{14};\ M_{42} = M_{24};\\
M_{43} = {} & M_{34};\ M_{44} = Iz4c2^2 + Iz5c2^2 + Iz6c2^2 + Ix4c01^2s2^2 + Ix5c01^2s2^2\\
& + Ix6c01^2s2^2 + d5^2m6s2^2s3^2s4^2 + Iy4s2^2s01^2 + Iy5s2^2s01^2\\
& + Iy6s2^2s01^2 + d5^2m6s4^2(c2c01s3 + c3s01)^2 + d5^2m6s4^2(c3c01 - c2s3s01)^2;\\
M_{45} = {} & Iz5c2s2s3 + Iz6c2s2s3 - d5^2m6s2s3s4(c3c4s2 + c2s4) + Iy6s2(-s1(c2c0s3\\
& + c3s0) + c1(c3c0 - c2s3s0))s01 + Ix5c01s2(-c2c01s3 - c3s01)\\
& + Ix6c01s2(-c2c01s3 - c3s01) + Iy5s2s01(c3c01 - c2s3s01)\\
& - d5m6s4(c2c01s3 + c3s01)(d5c01(c2c3c4 - s2s4)\\
& - d5c4s3s01) + d5m6s4(c3c01 - c2s3s01)(d5c4c01s3 + d5(c2c3c4 - s2s4)s01);
\end{aligned}
$$

$$
\begin{aligned}
M_{46} = \; & Iz6c2(c2c4 - c3s2s4) + Ix6c01s2(c01(c4s2 + c2c3s4) - s3s4s01) \\
& + Iy6s2s01(c01s3s4 + (c4s2 + c2c3s4)s01); \\
M_{51} = \; & M_{15}; M_{52} = M_{25}; M_{53} = M_{35}; M_{54} = M_{45}; M_{55} = Iz5s2^2s3^2 \\
& + Iz6s2^2s3^2 + d5^2m6(c3c4s2 + c2s4)^2 + Iy6(-s1(c2c0s3 + c3s0) \\
& + c1(c3c0 - c2s3s0))^2 + Ix5(-c2c01s3 - c3s01)^2 + Ix6(-c2c01s3 - c3s01)^2 \\
& + Iy5(c3c01 - c2s3s01)^2 + m6(d5c01(c2c3c4 - s2s4) - d5c4s3s01)^2 \\
& + m6(d5c4c01s3 + d5(c2c3c4 - s2s4)s01)^2; \\
M_{56} = \; & Iz6s2s3(c2c4 - c3s2s4) + Ix6(-c2c01s3 - c3s01)(c01(c4s2 + c2c3s4) \\
& - s3s4s01) + Iy6(-s1(c2c0s3 + c3s0) + c1(c3c0 - c2s3s0))(c01s3s4 \\
& + (c4s2 + c2c3s4)s01); M_{61} = M_{16}; M_{62} = M_{26}; M_{63} = M_{36}; M_{64} = M_{46}; \\
M_{65} = \; & M_{56}; M_{66} = Iz6(c2c4 - c3s2s4)^2 + Ix6(c01(c4s2 + c2c3s4) - s3s4s01)^2 \\
& + Iy6(c01s3s4 + (c4s2 + c2c3s4)s01)^2; \qquad (32)
\end{aligned}
$$

A.3 The Coriolis Matrix

The Coriolis and centrifugal forces are computed directly from the inertia matrix via the formula

$$C_{ij}(q, \dot{q}) = \sum_{i=1}^{n} \Gamma_{ijk}\dot{q}_k = \frac{1}{2}\sum_{i=1}^{n}(\frac{\partial M_{ij}}{\partial q_k} + \frac{\partial M_{ik}}{\partial q_j} - \frac{\partial M_{kj}}{\partial q_i})\dot{q}_k \qquad (33)$$

A.4 The Gravity Matrix

$$
G = \begin{bmatrix}
0 \\
0 \\
(m2(-r2s2) - m3((r2 + r3)s2) + m4(-s2d3) + m5(-s2(d3 + r5c4) - r5c3c2s4))g0 \\
m5(r5s3s2s4)g0 \\
m5(c2(d3 - r5s4) - r5c3s2c4)g0 \\
0
\end{bmatrix} \qquad (34)
$$

where $c = r/(2 * l)$,
$c0 = cos(\phi_c), s0 = sin(\phi_c), c01 = cos(\phi_c + q1), c1 = cos(q1), \cdots, c5 = cos(q5)$,
$g0 = 9.81m/s2$, $r2 = 0.049$, $r3 = 0.124$, $r5 = 0.183$.

Quantitative and Qualitative Evaluation of Vision-Based Teleoperation of a Mobile Robot

Luca Brayda, Jesús Ortiz, Nicolas Mollet,
Ryad Chellali, and Jean-Guy Fontaine

Istituto Italiano di Tecnologia, TERA Dept, Via Morego 30, Genova, Italy
luca.brayda@iit.it
http://www.iit.it

Abstract. This paper analyzes how performance of a basic teleoperation task are influenced by the viewpoint of the video feedback, using a remote mobile robot. Specifically, the viewpoint is varied in terms of height and tilt and the influence on a basic task, such as following some pre-defined paths, is analyzed. The operators are able to control one motor degree of freedom and up to two perceptive degrees of freedom. It is shown that performance vary depending both on the viewpoint and on the amount of perceptive freedom; in particular, the chosen metrics give better results when more perspective and, surprisingly, a more constrained perception is deployed. Furthermore, the contrast between the actual performance and the performance perceived by the operators is shown, which allows to discuss about the need of quantitative approaches in measuring the efficiency of a teleoperation task.

1 Introduction

Teleoperating a mobile robot through its camera is intuitively the most natural way to drive the robot. The user takes the viewpoint of the robot, so he/she could feel present in the remote environment. Such is the role of an efficient immersion in general, which uses sensory information and associated displays technology that are combined to obtain a feeling of being present in an environment different from the real one [She92]. According to Helmholtz' doctrine of unconscious inference in 1882 and a lot of other studies [Pfa96], even with an ideal sensory stimulation that matches the original environment, the perceived world is different from the veridical world, because it is seen through certain and chosen sensory channels and is interpreted and re-constructed by our own experience. So the efficiency of a system to achieve a task can be different from the perception that teleoperators could have, and systems need to be evaluated with neutral but quantified criteria. The case of vision feedback is, as a consequence, much more complicated than evaluations of comfort or sensation of immersion.

The purpose of this paper is twofold: first, we present preliminary results in studying quantitative metrics to evaluate operators' performance while guiding

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 792–801, 2009.
Springer-Verlag Berlin Heidelberg 2009

a remote robot through visual feedback, particularly when such feedback is given from some very different points of view; second, we compare such metrics to a more qualitative evaluation and show how difficult may be to infer the operators' performance from a mere qualitative study.

The utilization of robot's camera for teleoperation has been widely spread. However, this interface is still lacking in supporting perfect visual feedbacks. This is mainly due to the known drawbacks such as the reduced Field Of View (FOV), the video transfer latency, the unintuitive interface for camera's control or the loss of directional sense. Several techniques have been employed according to the literature to improve user's comfort and efficiency through the visual feedback. In order to compensate the Head Mounted Display (HMD) latency (due to this minimal procedure: obtaining HMD position, controlling the robot's camera, and finally obtaining video's point of view), authors of [Fia05] proposed the use of panoramic images that cover the whole potential FOV of the user. On the other hand, and about the loss of directional sense, we can find systems that display pitch and roll information of robot's camera [Nie07][La03], where experiments show an efficient compensation with such displays in addition to the real view of the camera. Depth perception is a major problem, and has been partially solved through the use of stereoscopic cameras [LP06][SN00] with adjusted parallax that improve perception feeling and comfort.

Generally speaking , such solutions *compensate* known problems, but it's still possible to improve the initial system itself, through the accurate evaluations of human's characteristics, limits and needs. In [TC06] the minimum video frame rate is evaluated, while in [Cho05] and [Ra05], the authors underline the importance of the FOV and the resolution of displays like a HMD for distortion perception: region warping are more perceptible with higher (and quantified) FOV and resolutions. Other studies underline the internal conflicts felt by human and provided by teleoperation systems [BR98], as only a few sensory channels are stimulated. [DED98] highlights the human's mental construction of spatial knowledge and understanding of a remote environment through VR, reality, and maps. It also highlights the differences of perception and understanding according to the methods of exploration and the restrictions posed by the displays like the FOV or the displacements through a mouse. On the contrary, the visual channel can be distorted in a controlled way to improve task efficiency [CMA03].

However, to the best of our knowledge there is still a lack of effort in evaluating how the different points of view of the video feedback, for a given task, can affect performance and, second, how this performance can be measured. Specifically, it is not yet clear how the relative amount of task-based or user-centered metrics can contribute to infer the goodness of a teleoperation task, and how these metrics relate to the effort being demanded to the operator. Furthermore, given that distant world inherently cause distorted perception, there is also lot to do in research on the way such perception can bias, positively or negatively, the operator's judgement about their own performance, thus leading qualitative feedback far from the quantitative results.

2 Performance Evaluation of Teleoperation Task

In this work we aim at finding ways to measure the capability of a teleoperator to achieve a simple task. This can be a path following task that the operator must perform: the path can be depicted on the ground and the user must drive the robot as close as possible to this path. The evaluation is done by comparing the path traced by the mobile robot and the original path. This allows us to drive some conclusions concerning the behavior of the operator. Specifically, one way to measure the degree of accuracy of the task is to compute the surface between the theoretical (T) and the experimental (E) path. Each path is modeled as a curve, approximated by a piecewise linear segments joined by points in a 2D space: the approximation comes from the fact that the position and orientation of the robot is sampled, e.g by a camera acquisition system. By considering that the T and E frequently cross each other, the in-between surface S can be computed as:

$$S = \frac{1}{2} \sum_{i \in I} \sum_{p \in P_i} \begin{vmatrix} x_p & x_{p+1} \\ y_p & y_{p+1} \end{vmatrix} \tag{1}$$

where $I \in \{T \cap E\}$ is the set of points in which the two paths intersect, $P_i \in \{T \cup E\}$ is a subset of points between two consecutive intersection, p and $p+1$ are two consecutive points in each subset and x, y are the 2D coordinates of a point. The inner sum is the known Surveyor's formula for the calculus of the area of a polygon. S can be interpreted as a surface-based error. Furthermore, because we make test across different paths of different lengths, we can normalize by the theoretical path lengths by defining a Normalized Average Distance (NAD):

$$NAD = \frac{S}{\sum_{p \in T} \sqrt{\Delta x_p^2 + \Delta y_p^2}} \tag{2}$$

With such metric, the operators with a high/low NAD will be likely to have experienced a higher/lower deviation in following the main path. Such deviations are related to the degree of ability people have to change mentally their point of view (POV) or, on the contrary, it may represent the distortion the teleoperation system imposes to them. In other words, the deviation depend (at least partially) on the fidelity of the perception of space that each operator can feel. Figure 2(d) depicts an example of surface S, where the area is depicted in gray. The relationship is partial because other ingredients are missed such as the motor transformation between the hand actions and the robots rotations.

3 Experimental Setup

3.1 Description of the Protocol

In the experiments, the users had to follow as best as they could a stained path, by carrying out the teleoperation of an Unmanned Ground Vehicle (UGV) using

a joystick for motor control output and a Head Tracking System (HTS) for perceptive control input. The users didn't have any previous knowledge about the UGV or the path, and during the experiments they could rely on the sole subjective vision by teleoperating in a separated room. To reduce the experiment variability, the speed of the vehicle was fixed to 0.15 m/s (25% of the maximum speed). This way, the user only had to care about one degree of freedom of the UGV, i.e. the steering, and two degrees of freedom for the HTS (pan & tilt): this way the comparisons can be simpler and clearer.

The experiment was carried out by 7 people (3 women and 4 men), with an age range from 22 to 46 years old. Every user made a total number of 9 trials, i.e. 3 paths by 3 POV configurations. The use of the HTS was alternated between trials, so there is an average of 3.5 users for every possible combination of paths, POV and pan & tilt. The total amount of trials is then 63 (7 users times 9 trials). To avoid the influence between experiments, the user never made two trials in a row (the user distribution is interlaced): rather, we tried to maximize the time between two successive trials.

The scene could be observed via three different POV, each of them corresponding to a different [tilt, height] pair (see Table 1(a)). The height is referred to the ground level and the tilt angle is referred to the horizon: the higher the value, the more the camera is looking down. Note that the users could not perform "self-observation", thus they were not able to develop any possible new proprioceptive model. After every trial, the users were asked to draw the shape of the path. Finally, once all trials were finished, the users filled a short form with questions regarding to the subjective perception of the experiment.

3.2 Unmanned Ground Vehicle (UGV) and User Interface

The UGV used during testing was a small vehicle ($0.27m$ length x $0.32m$ width) which was built using a commercial platform. This base has four motored wheels without steering system. The speed control of each wheel is used to steer the vehicle. Figure 1(a) shows a picture of the UGV. The pan & tilt camera system was placed in a vertical guide in order to change the height of the camera. This systems uses a manual configuration since the height was only changed between experiments and not during them. The webcam has a standard resolution of 640x480 pixels and a horizontal FOV of 36 degrees. For the experiments the frame capture was made at 15 frames per second.

The user interface is composed by three main elements:

Table 1. Experimental constraints (a) and paths data (b)

(a) Points of view

	1	**2**	**3**
Height (m)	0.073	0.276	0.472
Tilt angle (deg)	1.5	29.0	45.0

(b) Paths

	1	**2**	**3**
Length (m)	19.42	16.10	9.06
Width (m)	0.28	0.28	0.42

(a) General view of the UGV, with Vicon markers

(b) Operator during the experiments

Fig. 1. Experimental setup

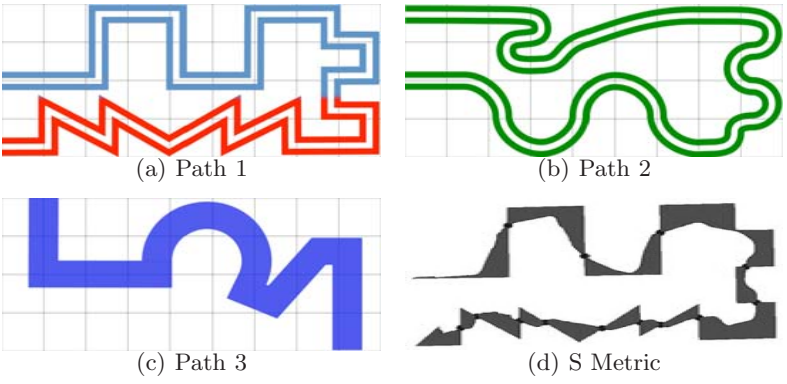
- Head Mounted Display. The user watched the images acquired by the UGV's webcam through a HMD system (see figure 1(b)).
- Joystick. The user only controlled the steering of the UGV, since it travels at constant speed. To make the control as natural as possible, the vertical rotation axis of the joystick was chosen (see figure 1(b)). The joystick orientation was recorded during the experiments.
- Head Tracking System. To acquire the user's head movement when controlling the pan & tilt movement of the camera, a wireless inertial sensor system was used (see figure 1(b)). The head orientation was also recorded during the experiments.

3.3 Data Acquisition System and Paths Description

During the experiments, the position and rotation of the UGV as well as the movement of the UGV's webcam was recorded at $50Hz$ using an optical motion capture system (Vicon[1]). Such system acquires the position of seven markers placed on the UGV (see Figure 1(a)) by means of 10 infra-red cameras (8 x 1.3Mpixel MX cameras and 2 x 2Mpixel F20 cameras). The raw data coming from this system was then properly reconstructed and filtered to extract the robot center. The user's input (joystick and HTS) was recorded with a frequency of $10Hz$, since that is the rate of the UGV's commands. To analyze the data, this information was resampled to $50Hz$ with a linear interpolation.

Three different paths were used in the experiments because we intend to compare the results in different conditions and across different styles and path complexities. They were placed under the Vicon system, covering a surface of about 13 square meters. The first path (figure 2(a)) is characterized by merlon

[1] http://www.vicon.com



(a) Path 1 (b) Path 2

(c) Path 3 (d) S Metric

Fig. 2. Paths used for the experiments and the S metric applied to Path 1

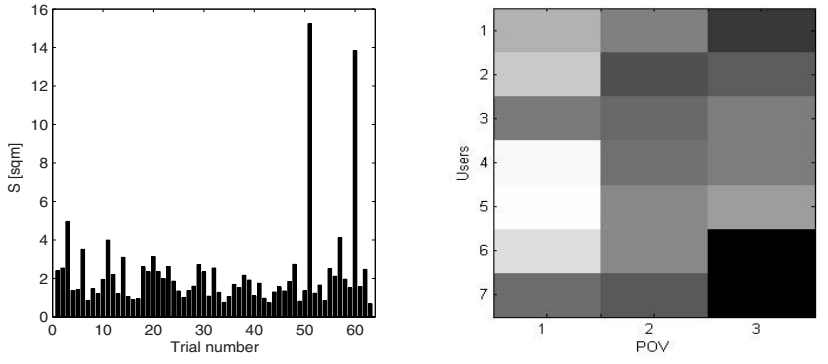
and sawtooth angles. The second path (figure 2(b)) has the same main shape of the first but is covered CCW by the robot and has rounded curves of different radius. The thid (see figure 2(c)) is simpler with wider curves and with rounded and sharp curves. The table 1(b) shows a measure comparison between paths.

4 Results

In this section we organize results according to our proposed metrics in a quantitative way, while results coming from questionnaires are commented in a qualitative evaluation. We then discuss about how these two approaches comply.

4.1 Quantitative Evaluation

We analyzed the performance of teleoperators across the 63 trials: Figure 3(a) shows that, according to the metric S, users spanned error surfaces in the range of [0.67,4.96] square meters. However, two outliers appear (largest S values): we verified that these correspond to trials in which the users got lost from the beginning, and whom we let pursue their trial until they arbitrarily decided to stop. We then observe the metric NAD in function of both users and POV (Figure 3(b)): brighter cells correspond to lower NAD, i.e. the experimental path was closest to the theoretical one. Clearly the POV closer to the ground and giving more perspective entails a better performance. The blackest cells, i.e. the biggest errors, are influenced by the two outliers: note that both outliers come from two different users, unaware of the mutual performance, who experienced the second path (this cannot be inferred from pictures) watched from the highest point of view (POV=3) with HTS activated on the camera. This is interesting, since they were exactly in the same condition: we will see that, even without outliers, this is the condition of average worst performance, and that this result provides a hint for the role of adding degrees of freedom. In order to analyze data more consistently, we derive a reduced test set without outliers (61 trials): Figure 4(a)



(a) (b)

Fig. 3. Global per-trial (a) and per-user (b) performance on the whole test set. In (a) we evidence the presence of two outliers as well as the order of magnitude of the surface S, in (b) the brighter the color, the lower the NAD, i.e. the better the performance.

plots, for each POV, the bar related to the performance in NAD across the three paths, as well as their mean and standard deviation depicted as circle and vertical segments: performance get worse as the POV increases in height and tilt. In fact the middle and highest POV gives the worst performance, and while their mean is similar in absolute (NAD =0.140 and 0.142 meters respectively), the third POV compels users to a higher standard deviation (0.047 and 0.065 meters respectively). This denotes that with less perspective there is a higher tendency to deviate from the correct path. This is intuitive, since less perspective implies less chance given by the system to the operator to plan the future path and to anticipate the mid-long term controls. We will see that this goes against the user perceived performance. For the sake of completeness, we note that including the outliers would make the bars increase monotonically for increasing heights, thus make the phenomenon more evident. The best performance is, in contrast, reached at the lowest POV (0.10 mean and 0.032 meters std), in which the perspective played an active role in keeping the mobile robot tight to the theoretical path. It is interesting to note that the global NAD mean on the 61 trials is 0.127 meters, which is less than half the width of any path: this means that users, on average, stayed in the corridor or within the lines (with a std of 0.052 meters), and globally complied with the protocol requests.

4.2 Qualitative Evaluation

As anticipated in Section 3, users were asked a few questions about their global performance: specifically, they were asked 1) how many paths they perceived during the whole session, 2) whether the HTS was useful to achieve the task or to 3) globally perceive the environment, 4) whether a POV with more perspective was more useful and 5) to list the paths in decreasing order of self-evaluated performance. Table 2 shows the given answers for each user number. Answers to

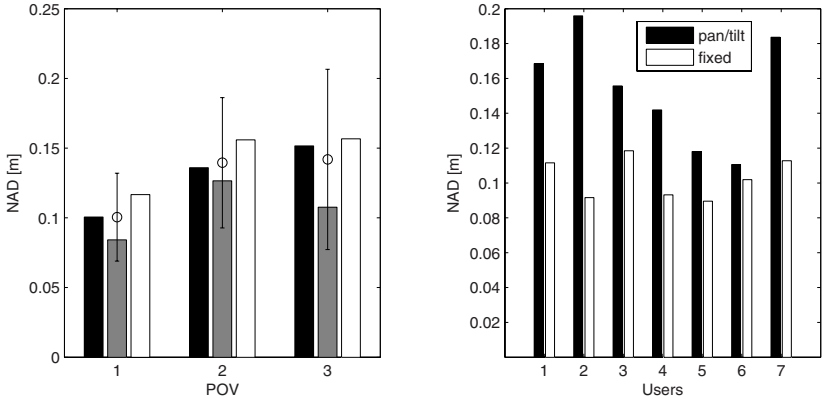


Fig. 4. Average and per-POV (a) and per-user (b) performance on the reduced test-set. In (a) we show the per-POV *NAD* mean (circle) and standard deviation (segments), as well as how they distribute across paths. In (b) we show that for this task the pan & tilt feature was detrimental for performance.

q.2 shows that the three POV were so different that some users thought to experience more than three paths. Given that each path had its own color, the result is surprising. Answers to q.2 and 3 were different across users: they show that users judged the HTS detrimental in 4 out of 7 cases, but rather good for the global perception of the environment (5 out of 7). By comparing such answers to Figure 4(b) we derive that pan & tilt is generally an advanced feature which, nonetheless, makes performance drop, which is a counter-intuitive result. This may indicate that global exploration and the task achievement required different kind of level of attention. This fact is confirmed by the extreme difficulty in drawing the geometrical shape of the path right after the experiment. Interestingly, the two users (5 and 6) who indicated pan & tilt useful for good task achieving were actually wrong (performance were worse, we claim because of distorted perception), but experienced the smallest differences with and without pan & tilt (see the relative height difference between white and black bars).

More interesting are answers to questions 4 and 5: we know from Figure 4(a) that performance (both mean and std) are decreasing with increasing height. This goes against the users evaluation. In fact a "no" / "yes" *should* imply better performance with POV=3 / POV=1 and vice versa. It is enough to look at Figure 3(b) to realize that self-evaluated performance corresponds to the actual performance (in parenthesis) only for two users. We consider not a chance that these two users are also the best performing ones. This demonstrates first that the perception is on average distorted and second that distorted perception is related to a worst performance. Finally, the Table shows that there is rare correspondence between preferred paths and paths ordered by descending actual performance. In particular, path 1 was underestimated, while path2 and path3

Table 2. Users' evaluation: number of perceived paths, HTS judged useful for task achievement and global perception, perspective judged useful (actual result), preferred paths (actual result)

user	paths	HTS for task	HTS for global	persp.useful (perf)	pref.paths (perf)
1	3	no	no	no(yes)	321 (321)
2	3	no	yes	no(yes)	231 (213)
3	6	no	no	yes(no)	321 (213)
4	4	no	yes	yes(yes)	231 (213)
5	6	yes	yes	yes(yes)	123 (213)
6	3	yes	yes	no(yes)	321 (132)
7	4	no	no	no(yes)	312 (132)

overestimated. This indicates a sort of distance between the perceived comfort and the actual performance.

5 Discussion and Conclusion

In this preliminary work we found that performance of a basic teleoperation task are influenced by the viewpoint of the video feedback. Future work will investigate how the height and the fixed tilt of the viewpoint can be studied separately, so that the relative contribution can be derived. The metric we used allows us to distinguish between a tightly and a loosely followed path, but one limitation is that we still know little about the degree of anticipation and the degree of integration of the theoretical path that an operator can develop.

Furthermore, we have shown that, non intuitively, the effects of a HTS were detrimental for performance: we speculate that results with an active HTS could be negative because we constrained velocity to be fixed. On one side, in fact, we added two degrees of freedom and approached a human-like behavior, on the other side we forced the user to take decisions at an arbitrary, fixed, and as such unnatural speed, thus conflicting with the given freedom. However, we point out that the operators who positively judged an active HTS also spontaneously used the first seconds of the experiment to watch the global path, then concentrated on the requested task. The results coming from HTS could also be biased by the absence of an eye-tracking system, as the true direction of attention is not uniquely defined by the head orientation. From the questionnaire, the post-experiments drawings and from further oral comments, we can conclude that operators cannot concentrate both on following and remembering a path. This is a constraint and a precious hint for future considerations about possible multi-tasking activities. Globally speaking, our evaluations show that good performances imply that self-judgement about performance can be reliable, while the sole judgements are misleading and cannot be used as a measure of performance and no implications can be derived from them. This confirms the motivation of our study about the need of quantitative measures for teleoperation purposes.

We also confirmed users' sensation are heterogeneous, i.e. that there is no preferred mode of using the mobile robot (POV, HTS), but clearly there is a mode which improves performance which is rather independent on the path shape and that this mode is not evaluated by operators as the best. This confirms, as expected in many applications, an inter-operator variability. We believe that such variability needs to make the design of a teleoperation system adaptive and self-compensating according to quantitative, pre-defined but possibly evolving metrics.

Acknowledgements

We would like to thank Stefano Saliceti for customizing the UGV for this study. We also would like to thank Marco Jacono, Ambra Bisio and Thierry Pozzo from the IIT RBCS dept. for their strong technical support and their availability, as well as Nick Dring for the graphical design and all our teleoperators.

References

[BR98] Biocca, F.A., Rolland, J.P.: Virtual eyes can rearrange your body: adaptation to visual displacement in see-trough HMD. Presence 7(3) (1998)

[Cho05] Chow, Y., et al.: The effects of head-mounted display attributes on human visual perception of region warping distortions. In: Proc. of IVCNZ (2005)

[CMA03] Casals, A., Muiioz, L., Amat, J.: Workspace deformation based teleoperation for the increase of movement precision. In: IEEE ICRA (2003)

[DED98] Waller, D., Hunt, E., Knapp, D.: The transfer of spatial knowledge in virtual environment training. Presence 7(2), 129–143 (1998)

[Fia05] Fiala, M.: Pano-presence for teleoperation. In: Proc. of IROS (2005)

[La03] Lewis, M., et al.: Experiments with attitude: attitude displays for teleoperation. In: Proc. IEEE Int. Conf. on Systems, Man, and Cybernetics (2003)

[LP06] Livatino, S., Privitera, F.: 3d environment cognition in stereoscopic robot teleguide. In: Int. Conf. Spatial Cognition (2006)

[Nie07] Nielsen, C.W.: Ecological interfaces for improving mobile robot teleoperation. IEEE transaction on robotics 23 (2007)

[Pfa96] Pfautz, J.D.: Distortion of depth perception in a virtual environment application. PhD thesis (1996)

[Ra05] Ryu, J., et al.: Influence of resolution degradation on distance estimation in virtual space displaying static and dynamic image. In: IEEE I.C. on Cyberworlds (2005)

[She92] Sheridan: Defining our terms. Presence 1(2), 272–274 (1992)

[SN00] Siegel, M., Nagata, S.: Just enough reality: Comfortable 3-d viewing via microstereopsis. IEEE Trans. Circ. and Sys. for Video Tec. (2000)

[TC06] Thropp, J.E., Chen, J.Y.C.: The effects of slow frame rates on human performance. U.S. Army Research Laboratory Technical report (2006)

An Embedded Vision System for a Power Transmission Line Inspection Robot

Weiyang Lei, En Li, Guodong Yang, Changchun Fan, Fengshui Jing, and Zize Liang

Laboratory of Complex Systems and Intelligence Science,
Institute of Automation,
Chinese Academy of Sciences, 100190 Beijing, China
{weiyang.lei,en.li,guodong.yang,changchun.fan,
fengshui.jing,zizi.liang}@ia.ac.cn

Abstract. Surrounding environment information is significant for an inspection robot to plan its behavior to stride over obstacles. This paper describes a compact embedded monocular vision system for a power transmission line inspection robot. This system involving a DSP processor has superior hardware design and excellent performance on image processing. Emphasis of the image processing software design is the algorithm of image segmentation and straight line extraction. Base on the structure of the transmission lines, an algorithm for obstacle distance estimation is proposed. Experimental results show that good capability on recognition and location can be achieved with this proposed vision system.

Keywords: Embedded Vision System, Power Line Inspection Robot, Monocular Vision, Obstacle Recognition.

1 Introduction

In order to ensure the quality and reliability of electric power supplied to customers, power transmission lines must be inspected regularly. Unfortunately, most of current inspection work is performed artificially. An intelligent mobile robot for inspection is eagerly required on power transmission lines, which can crawl along the transmission line to accomplish some necessary operations, such as capturing and recording the videos of the environment around the power lines and towers. Several kinds of electric equipments settled on the power transmission lines are acting as obstacles to the inspection robot during continuous inspection work. Consequently, a navigation system which can recognize and locate the obstacles is needed to help the robot to plan its motion for stride over these obstacles autonomously. An effective and real-time vision system is an excellent choice to acquire the information of the obstacles in advance and inform the robot controller to perform the obstacle-avoidance operations before the collisions happen. Zhang *et al.* [1] developed a binocular vision system based on PC104 hardware platform for a power line inspection robot. Binocular vision system has high precision on obstacle location, but the complexity of the object matching algorithm is not appropriate to a low power consumption real-time system.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 802–811, 2009.

Pouliot *et al*. [2] developed an inspection robot for power transmission lines with a complicated vision system composed of four cameras. Images are sent to ground control center for analysis which makes the robot unable to move autonomously along the line once it can not communicate with ground center.

In this paper, we design an embedded monocular vision system to recognize and estimate the distance of the current obstacle for a special kind of power line robot which runs in partite 220KV transmission lines environment. This vision system uses high-speed DSP micro-processor to capture images and carry out real-time image processing algorithms. According to the specific characteristic of the partite lines structure, an obstacle distance estimating arithmetic is proposed. The rest of the paper is organized as follows. In Section 2, a brief introduction of our inspection robot and the structure of the partite transmission lines is given. Section 3 introduces the monocular vision system and the algorithms of the image processing. Section4 describes the obstacle recognition method and the distance estimating arithmetic. Conclusions are given in the end.

2 Power Line Inspection Robot and Its Working Environment

Considering the requirements in mechanical design and the structure of the partite lines and obstacles, our inspection robot, as shown in Fig. 1, is equipped with two arms and a counter weight box where the controller and the vision system are installed in. Each arm has a driving wheel and an up-down design, and is designed to be able to loose and grasp the line. Hence, the robot can fluently achieve automatic motions on transmission lines, such as rolling, crawling or spanning obstacles, generally at a 200m/h working velocity.

Fig. 1. The power line robot and the structure of its arm

The navigation system of the inspection robot need to recognize and locate the obstacles with its sensors, and communicate with control system which plans the motions of the robot to crawl along the line autonomously. The navigation system is composed of two sections: electromagnetic sensor network and the embedded vision system. The electromagnetic sensor network is installed at the arms of the robot, very close to the grasp claws. Different electric equipments installed on the power lines will lead to different inflection to the electromagnetic field intensity, so it is realizable for the electromagnetic sensor network to analyze which and where the obstacle it is

currently. However, the electromagnetic sensor network can only get information extremely near to the arms so the embedded vision system is extraordinary significant to supply necessary information of the distant environment in front of the robot. With this information, the robot can make estimations and judgments in advanced and decelerate for the preparations of the obstacle-detouring actions, to avoid any probable collisions.

In China, high-voltage transmission lines are usually arranged in bundles of 2 or 4 cables, supported by transmission towers. Generally, the distance between two adjacent towers varies from 300m to 500m. Though the entire line as long distance between two towers is actually shaped as a parabola curve, the local segment that the robot is crawling along can be recognized as straight line because the robot is too short relative to the distance between two towers. In the span between two towers, spacers are placed to keep a fixed distance between different cables and avoid them from touching each other. The distance between two parallel cables is usually an invariable distance of 400mm. Fig. 2 gives a sketch of the partite transmission lines and the equipments installed on. Spacers and other kinds of equipments such as counterweights, suspension clamp and strain clamp with isolator strings are obstacles for the robot. All of these obstacles have specific industrial normalized size, and also have distinctly differences in configuration and appearance. These characteristics can help us to figure out a simple and rapidly algorithm for vision-based obstacle recognition and distance estimation.

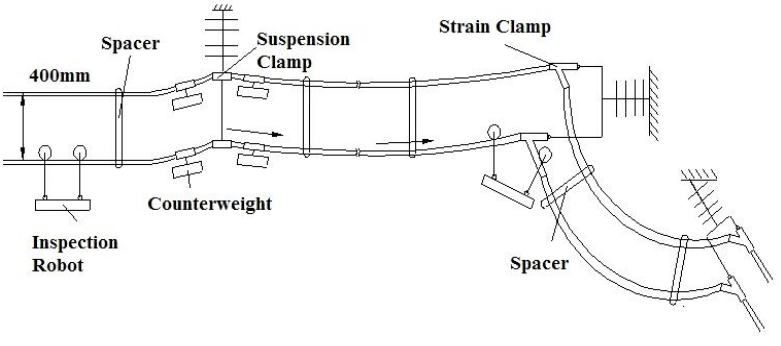


Fig. 2. Structure of the partite transmission lines

3 The Vision System of the Power Line Robot

Compared with PC-based vision system, the DSP-based embedded system is more suitable for an inspection robot, because the control box has a strictly occupation space limit and the power supply is also limited for the robot. Our vision system is designed with high-speed DSP micro-processor and characterized with compact size, low power cost and high computational performance. DSP is the most popular image processor because of its high speed data processing capacity. Base on the DSP hardware platform, specialized image processing algorithms have been developed and can steadily capture and analyze the images and recognize the obstacles in front of the robot.

3.1 Hardware Design

The vision system consists of a camera, an image processing module, and video transmission devices, as we can see in Fig. 3. The camera is installed in the front of the robot, precisely under the transmission line. The optical axis of the camera intersects with the transmission line by 20°. This configuration ensures the most background of the image is sky, which reduces the complexity of image analysis. Image processing algorithms are running on the DSP-based image processing module. Once an obstacle is recognized, information about its category and location will be sent to the robot controller via the RS-232 interface. Wireless video transmission equipment can send desired images to the ground surveillance center.

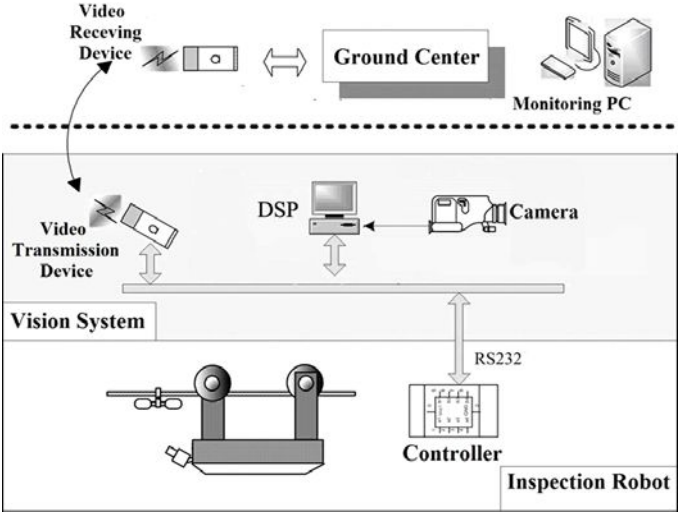


Fig. 3. Working scheme of the vision system

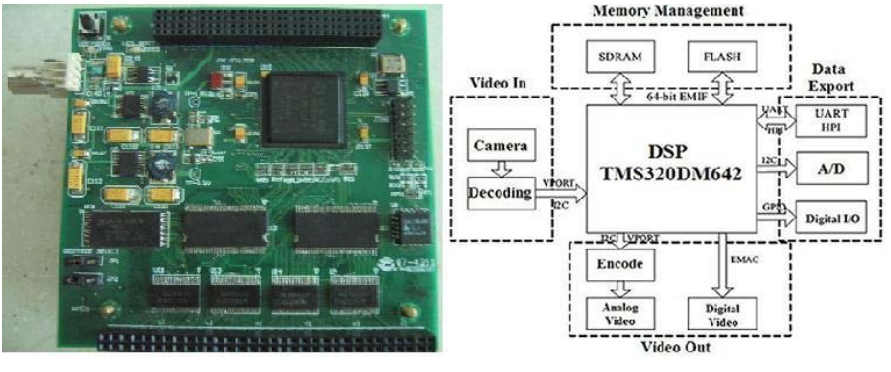


Fig. 4. DSP-based image processing board and its framework

DSP acts as the core of the whole system and is responsible for reading data from memory to implement image processing and information extraction. The professional digital media processor TMS320DM642 is selected for its excellent image processing performance. Necessary auxiliary hardware modules have been developed, including the video capture module, the video encoding module, the communication module, and the memory management module. Fig. 4. shows the DSP board we designed and its internal structure. The board has compact size (100×100mm) and low power consumption (1.2w). The availability and effectiveness of the hardware system have been proved in experiments. Modular structure on hardware design and sufficient peripheral interfaces make it easy to be applied for other different requirements with few modifications.

3.2 Software Design for Image Processing

The software system is composed of device drivers, application program, and image processing algorithm package. The superiority of the embedded vision system is its rapidly processing capacity especially for real-time image. So the image processing algorithms must have high-speed characteristics and be fit for the features of the image contents.

The key point of image processing is to locate the transmission lines. Firstly two transmission lines need to be exacted from the image, and then the parameters of the lines can be calculated. With the geometry information of the transmission lines, obstacle searching can be confined to the areas along the lines on the image, for the purpose of reducing searching areas and improving the efficiency of the obstacle recognition algorithm. The major procedure of the image processing algorithm is given as Fig. 5.

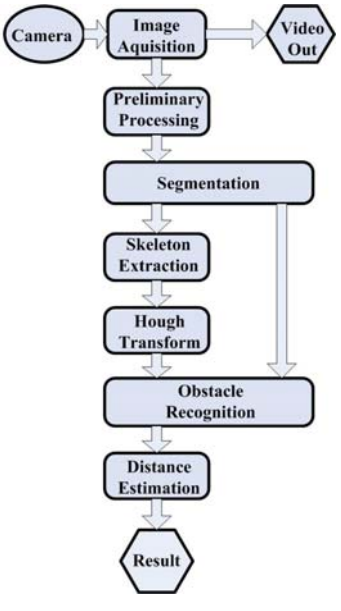
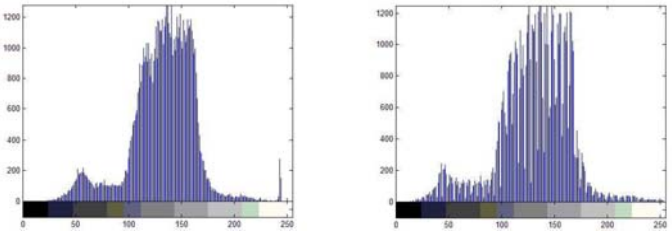


Fig. 5. Flow chart of the image processing algorithm

Step1: Preliminary processing
In the procedure of the image acquisition and sampling, noise points are inevitable. Changes of the illumination density will affect the brightness difference between the object and the background. The purpose of the preliminary processing is to reduce these negative influences in order to guarantee the intact contour of the obstacle after segmentation. Firstly, a median filter is used as an image smoothing algorithm for noise suppression. Then, considering the negative influence brought by the illumination changes, an adaptive histogram equalization algorithm is used for image enhancement to strengthen the object we interested. The algorithm is a localization of the histogram equalization algorithm and has rapid running speed. It calculates the new gray-scale of the pixel only by its neighborhood histogram rather than the global histogram. The processing result is shown in Fig. 6. As we can see, the transmission lines and obstacles are more prominent and the histogram has been improved after the preliminary processing.

(a) Original suspension clamps image. (b) Preliminary processing result.



(c) The histogram of the original image. (d) The histogram after processing.

Fig. 6. Preliminary processing result

Step2: Image segmentation and skeleton extraction
The purpose of the image segmentation is to separate the interest objects from the background. In this step, an optimal thresholding algorithm is used to distinctly separate the transmission lines and current obstacles from the background. The optimal threshold is calculated according to the histogram, by seeking the maximal gray-scale difference between the objects and the background. As a result, the image is transformed to a binary image. This algorithm is easy to carry out and the result is satisfactory for skeletonization, as shown in Fig. 7.

(a) Counterweight (b) Spacer (c) Suspension clamps

Fig. 7. Image segmentation results

A robust binary image skeletonization algorithm based on the medial axis transformation of the upright maximal squares can extract the skeletons from the image. By computing the numbers of the maximal neighboring pixels in an object, a medial axis transformed image is obtained, and then skeleton can be extracted. With an initialized threshold, this method can eliminate the noise. Fig. 8 shows that the transmission lines can be excellently extracted in this step and it will be convenient to calculate the parameters of the transmission lines in step3.

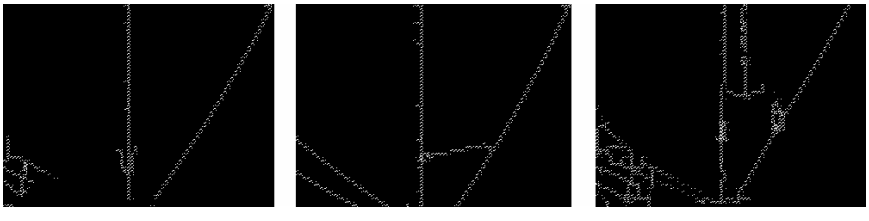


Fig. 8. Image skeletonization results

Step3: Straight-line extraction

The geometry information of the transmission lines is very important for obstacle's recognition and distance estimation. In this step, a fast Hough transform algorithm is used to extract straight lines from the skeletons extracted in step2, and calculate the slope and the intercept of the lines. With these geometry parameters we can determine which lines on the image are the transmission lines, and confirm the locations of the transmission lines.

After the above steps of the image processing, a binary image is acquired on which the obstacles has been separated from background. And the geometry information of the transmission lines on the image is known. With these image processing results, the recognition and location of the obstacles are feasible.

4 The Algorithm of the Obstacle Recognition and Distance Estimation

The contour and the position of the obstacles are clear enough on the binary image after segmentation. According to the parameters of the transmission lines obtained from the previous frame, partial searches are carried out along and between the transmission lines. If pixels block exists along or in the middle of the two lines, it will fill the searching window at certain percent. If the percentage exceeds a threshold, there is a big chance of the existence of an obstacle. Whether it is an obstacle or not can be inferred according to the features of the industrial installation standard on transmission lines. Fig. 9. gives the obstacle detection results.

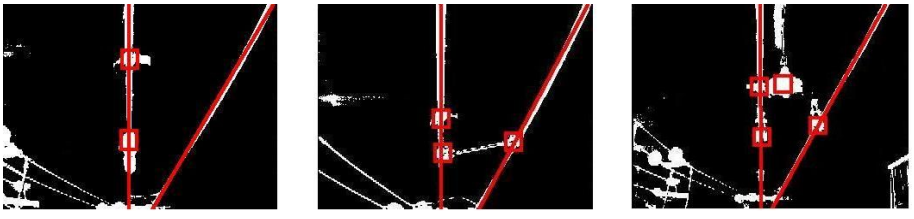


Fig. 9. Obstacle recognition

Combined with the imaging principles, the obstacle's distance can be estimated according to the geometry structures of the transmission lines. As we can see in Fig. 10, in order to make the calculation formula simple and convenient to understand, the origin of the camera reference frame (x_c, y_c, z_c) is settled at the camera's optical center, z-axis is defined along the camera's optical axis and the x-axis is defined vertical to the parallel transmission lines. The origin of the world reference frame (x_w, y_w, z_w) is settled at the intersection point of the transmission line and its vertical line through the camera's optical center, the z-axis is defined along the transmission line and the x-axis is defined vertical to the transmission lines.

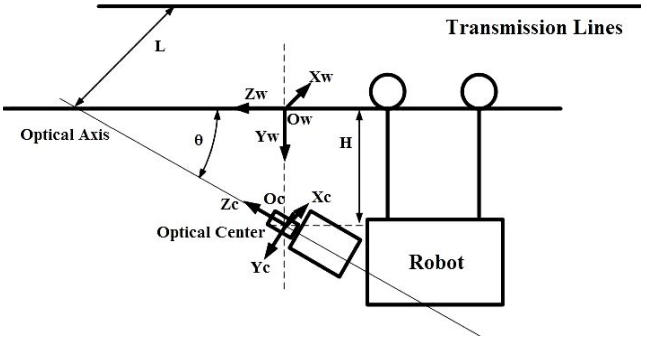


Fig. 10. Obstacle distance estimation

The coordinate transformation from the world reference frame to the camera reference frame is described as:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}, \text{ or } \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \mathbf{T} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}. \tag{1}$$

Depending on geometry structures of the transmission lines we have mentioned, the parameters of the formula (1) can be concluded as:

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix} \quad \mathbf{T} = H \begin{bmatrix} 0 \\ -\cos\theta \\ \sin\theta \end{bmatrix} \tag{2}$$

Based on the imaging principle, the coordinate transformation formula from the world reference frame to the image reference frame can be described as:

$$z \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{T} \\ \mathbf{0}^T & \mathbf{1} \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = M_1 M_2 X = MX. \tag{3}$$

In the formula (3), u_0, v_0 is the coordinate of the center point of the image. The parameters of M_1 are essentially the intrinsic parameters of the camera, which can be obtained from the camera calibration. The parameters of M_2, which can be regarded as the extrinsic parameters, have been calculated by formula (2). Therefore, the actual location of the obstacle's coordinate can be estimated according to its position on the image.

This arithmetic is tested on indoor simulation transmission lines. The imaging distortion has been taken into account for more precise results. The distances from the obstacles to the camera are measured in Table 1. The obstacle's apex towards the camera is regarded as the representative location of the obstacle.

Table 1. Comparison between the actual and measured distance

Measured Distance (cm)	Actual Distance (cm)	Error Percent
115	107	6.9%
163	156	4.3%
232	218	6.0%
289	263	8.8%

The estimating errors are approximate 9%, which could be accepted for applications. The accuracy of the distance measurement result mainly depends on the image processing precision and the stability of the geometry structure of the running environment.

If a dubious obstacle on or between the transmission lines is detected in several frames, then its distance to the robot will be constantly calculated to estimate its current velocity. If this velocity is close to the current velocity of the robot, there is a large

probability of the existence of this obstacle. This information combined with the current obstacle distance will be sent to the robot controller for further judgment.

The situation of the bundles of 4-wires structure of the transmission lines is similar. The relative position of each cable is settled, and the parameters of their geometry structure are given. According to the imaging principle, it will be easily to deduce which lines on the image are the bottom parallel cables. Only considering the two bottom parallel cables, the arithmetic we proposed above is still applicable.

In most cases of real outdoor environment, the transmission line inclines in a slight angle due to the gravity of the robot and itself. And the robot may swing right-and-left on the transmission line as the result of the action of the counter weight adjustment and the wind force. The extrinsic parameters will be variable parameters for the vision system during the robot's running process. Only by a real-time modification of the extrinsic parameters of the camera, the accuracy of our vision system can be greatly improved in the real-world applications

5 Conclusions

This paper presented an embedded vision system for an inspection robot to detect and locate the typical obstacles on the power line. This embedded vision system is characterized with compact size, low power cost and excellent performance in image acquisition and processing operations. The algorithm of the obstacle's recognition and distance estimation is based on the structure of the transmission lines, so the calculation complexity is considerably reduced. Experiments show promising results on image processing and obstacles location. In future work, complex background environment will be taken into account, and more researches on improving the real-time processing ability in real-world applications are necessary.

Acknowledgments. The authors also wish to acknowledge the help and suggestions from Gongping Wu of the College of Power and Mechanics Engineering, Wuhan University. This research has been supported in part by the Hi-Tech R&D Program (863) of China (Grant No. 2006AA04Z202).

References

1. Yunchu, Z., Siyao, F., Xiaoguang, Z., Zize, L., Min, T., Yongqian, Z.: Visual Navigation for a Power Transmission line Inspection Robot. In: 2006 IEEE International Conference on Intelligent Computing, Kunming, YunNan Province, China (2006)
2. Paulo, D., Michele, G., Kensuke, T.: Expliner – Robot for Inspection of Transmission Lines. In: 2008 IEEE International Conference on Robotics and Automation, Pasadena, CA, USA (2008)
3. Nicolas, P., Serge, M.: Geometric Design of the LineScout, a Teleoperated Robot for Power Line Inspection and Maintenance. In: 2008 IEEE International Conference on Robotics and Automation, Pasadena, CA, USA (2008)
4. Texas Instruments Incorporated: TMS320C64x DSP External Memory Interface (EMIF) User's Guide. Literature Number: SPRU711B (2006)
5. Texas Instruments Incorporated: TMS320C64x Technical Overview Literature Number: SPRU395 (2000)

What Visual Cues Do We Use to Perceive Depth in Virtual Environments?

Abdeldjallil Naceri[1,2], Ryad Chellali[2], Simone Toma[1,2], and Fabien Dionnet[2]

[1] University of Genoa, via Balbi 5, 16126 Genoa, Italy
[2] Italian Institute of Technology, via Morego 30, 16163 Genoa, Italy
{abdeldjallil.naceri,ryad.chellali,simone.toma,fabien.dionnet}@iit.it

Abstract. The main objective of this work is to investigate human depth perception in virtual reality (VR). Specifically, we attempt to give a first step that towards finding the relationship between size-distance in depth perception in virtual environment. Depth perception has been shown to be key element and a major factor either for simple navigation tasks or for complex and dexterous manipulation tasks. However, in the history of psychology of perception few matters have been of more continuous interest than the issue of the relationship between perceived size and perceived distance. Most studies focused on such questions have converged upon a hypothesis named Size-Distance Invariance. This hypothesis is often stated in the following terms: "A retinal projection or visual angle of given size determines a unique ratio of apparent size to apparent distance" [1]. The relationship between distance and size perception remains unclear in a virtual environment. The effectiveness of virtual environments has often been linked to the sense of presence that users feel in the virtual world. Moreover, Depth perception is one major factor among many believed to underlie presence for teleoperation and virtual environments. Our findings suggest that the strategy based on imagination of motor tasks could have a major effect on users' accurate depth perception in virtual reality.

1 Introduction

One of the key issues in virtual reality (VR) environments is immersion, namely, one needs to give users the complete sensations of being somewhere else and feeling a realistic environment. In other words, immersive VR-based systems must generate stimulus that might be interpreted as ones coming from real world. For the visual channel, VR systems must provide adapted cues allowing the integration of accurate spatial information. Indeed, humans perceive the world in three dimensions to understand it and to achieve goal-directed motors actions while performing physical interactions. This complex task, e.g., 3D geometry integration and specifically depth perception, is derived from the integration of several visual and non-visual cues [2]. For depth perception, people use mainly binocular cues (exploitation of parallax, oculomotor convergence, etc.), monocular cues (motion parallax, accommodation, occlusion, relative size, etc.) and motor activity.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 812–823, 2009.

Many variations were proposed to understand the integration process and to see what parameters people rely on to estimate egocentric and allocentric distances. For instance, some studies showed that the size and the perceived egocentric-distance of a real-world object are strongly related by what is called the size-distance invariance hypothesis [1,3,4,5,6]. Other studies investigated in real environment conditions the differences between verbal estimation paradigms and open-loop motor behavior such as pointing, reaching or walking [7]. For immersive virtual environments, many analogous experiments has been also done [8]. The main finding in these studies is the fact that within VR environments, distances appear to be compressed. Many hypotheses were proposed to explain this fact such as lack of presence (e.g., subjects feel a difference between the real environment where the experiment is performed and the virtual one suggested by displayed images) or low rendering fidelity, etc. That is to say, one can easily see that for VR-based depth perception, there is a combination of unknowns and uncertainties: real-based and virtual-based ones.

Thus, accurate and effective analysis of how depth perception occurs in virtual contexts is still needed. This open issue may center upon finding a method to assess properly VR systems in allowing realistic perception of distances.

The first interest of this preliminary study is to design a valid and reliable experimental setup to investigate individual variables like perceived size, stereoscopic disparity, apparent motion and self-occlusions, context and backgrounds, etc. on which one may rely when perceiving depths. The next section presents the protocol we built and the expectations of the experiments to be performed. The third section presents the results of the experiments. We finish by discussing these results and by giving some directions toward our next step for deeper human depth perception investigations.

2 Protocol and Expectations

In designing the protocol we wanted firstly to verify if the data found in real context should have appeared in virtual environment conditions. Specifically, [9] found that an object which was consistently underestimated in relative size was consistently overestimated in relative distance (size-distance paradox), supporting the strong influence of the size-distance covariance in distance estimation. However, recently Berryhill and colleagues [10] reported a high rate of accuracy in performances of normal healthy subjects judging correctly either the size or the distance of objects even though both these perceptive variables did not covary and they were the only ones available. With the aim of reproducing similar conditions reported in [9,10] we designed a paradigm with four different ways of stimulus state and maintained it for all the experiments. In detail, states I and IV were characterized by a sequential presentation of two stimuli (sphere or cube) which could have either different or identical size at different positions (Figure 1). Such Berryhill-like experiment were introduced with the expectation that users should perform the task as well as those in real context either in state I, where a clear size distance co-variation was presented, and in state IV where

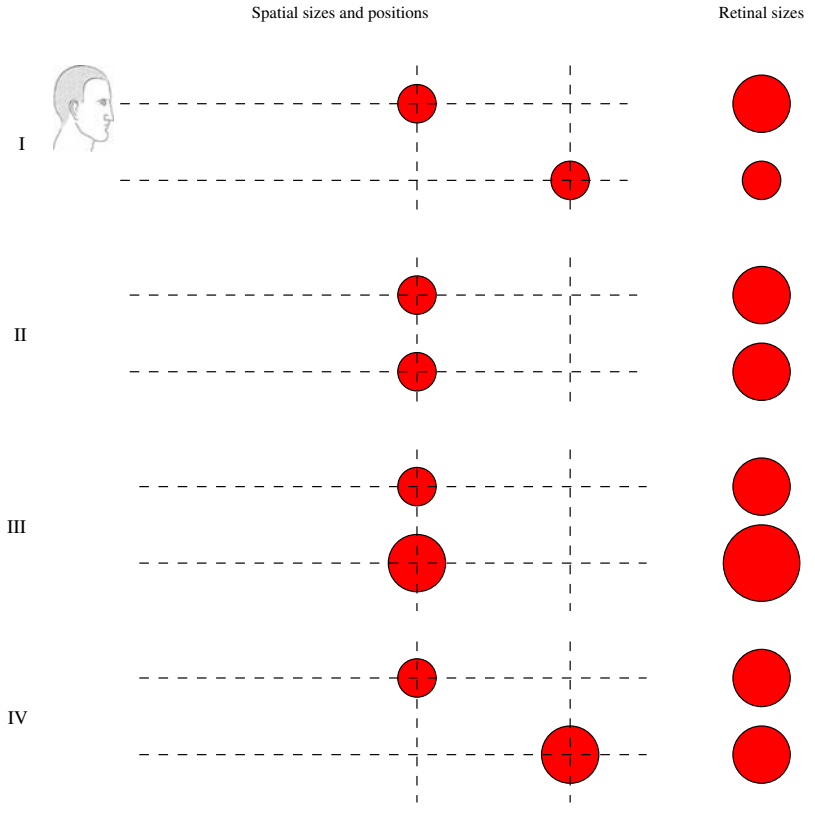


Fig. 1. Experimental states

only distance varied. On the other hand, we displayed the two other states (II and III) more similar with Gruber's experiment [9]. The aim here is to verify whether or not varying only the size could affect users' performances. For state II both stimuli (spheres and cubes) have the same size and the same position. In state III, stimuli have the same position but different sizes. Although, by means of the adjustment method, one could extract quantitative degrees of over- and underestimation in distance ratio judgments [9]. We expect that also that the size variation effect should diminish users' performances.

3 Experimental Setup and Method

3.1 Virtual Reality Hardware

The experiment took place in TERA department's virtual reality room. This room is equipped with a pair of stereo projectors based on the polarization effect, and a large projection screen (2.7×2 m^2). We used also head mounted device (HMD) to display the virtual objects. Images on the screen were projected

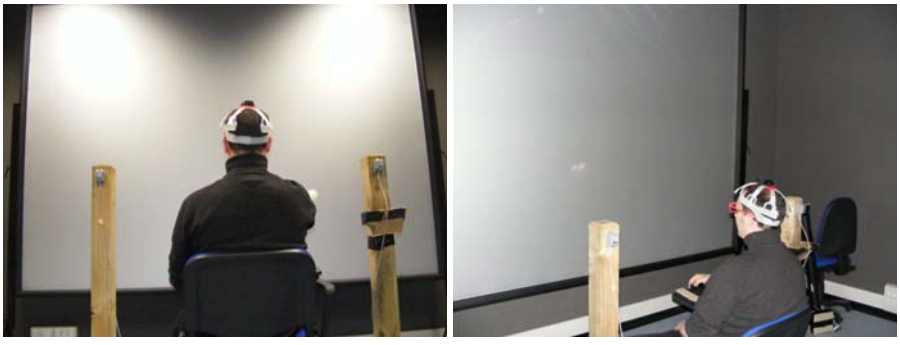


Fig. 2. Subject in the experimental site, seated at 1.6 m in front of a big screen wearing passive glasses. The subject wore a helmet for a head tracking that was not used in this experiment because the stimulus was presented at eye level with the same angle of the subject view.

simultaneously at 120 Hz. The subjects were sat about 160 cm in front of the projection screen (figure 2), and we maintained the same subject distance for big screen or HMD conditions. For the big screen condition, the subjects wore passive glasses containing two opposite filters to separate between the right and the left images. The subjects were asked to maintain a fixed position during the trials in order to have a constant retinal image during the experiment.

3.2 Task and Conditions

The subject carried out a verbal task in all the experiments and there were no motor tasks for perceiving the virtual spheres' distance. The experiment was performed on the big screen and on the head mounted device. We used two spheres, one of 7 cm and the other of 10 cm diameter. The spheres were displayed in the horizontal plane containing subjects' eyes. The two virtual spheres were projected to be 60 cm and 80 cm away from the subject for the closer and farther presentation respectively. Each sphere was presented alone for 3 sec, with 2 sec between presentations in order to avoid the apparent motion that could alter subjects' judgments. There were four experimental states (figure 1), e.g., spheres presentation.

- The same ball but at two different positions with a difference in apparent or retinal size,
- The same sphere was presented twice, the same size at the same position (identity),
- Two spheres of different sizes at the same position
- Two spheres, the largest one at 80 cm and the smaller at 60 cm from the subject in order to have a constant retinal size (a retinal size-based conflict).

Each state was divided in two others sub-states in order to check the effect of the presentation's order. The eight sub-states in total were presented randomly with

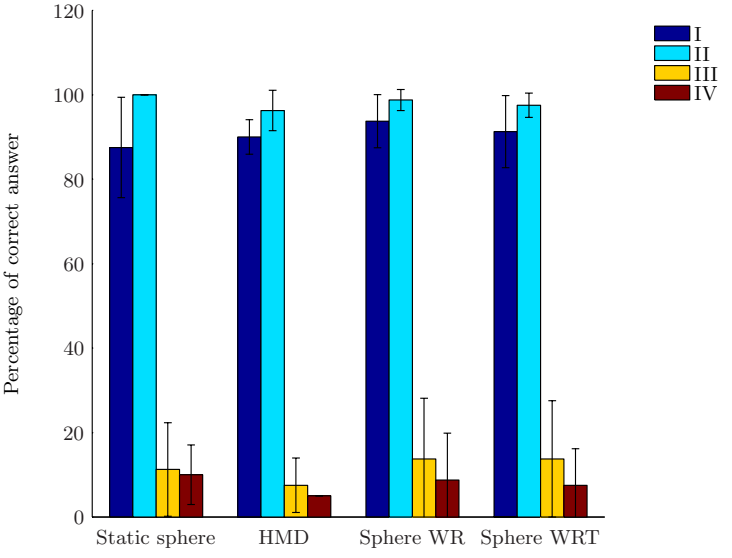


Fig. 3. Subjects' average percentage answers in different conditions of the group 1

ten repetitions for each one. In the beginning of the experiment, subjects carried out a rapid familiarization test (3 repetitions for each of the 8 sub-states).

Experiment 1. In the first experiment the subjects were asked to answer whether the two successive displayed spheres are at the same position or not; two alternative forced choice (2AFC). The first and the second conditions were with static virtual spheres displayed in big screen and then in HMD. In the third condition we added a self rotation for the spheres (Sphere WR). Finally, in the fourth condition, in addition of self motion, we added a tennis ball texture to the spheres (Sphere WRT).

Participants. Eight subjects (seven male and one female) took part in this first experiment. They all had normal or corrected to normal visual acuity.

Results. In this experiment, we had two categories of performances (figures 3, 4). In order to analyze deeply the result we have separated the data into two groups based on subjects' performances: bad and good performers.

First of all, we recorded that all the subjects of both groups were good at the states (I and II) in a big screen or in HMD condition. Subjects of the group with bad performance answered that the spheres were at the same position in retinal constant conditions, whether those ones were at the same or different positions (states II and IV). Similarly, they reported that the spheres were not at the same position in the physical constant or different (states I and III). On the contrary, the second group -those with good performance- were more accurate than the first group in all these states.

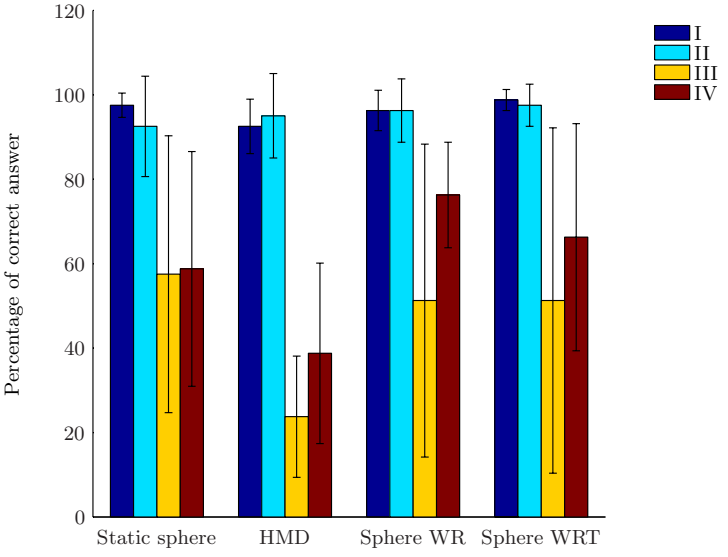


Fig. 4. Subjects' average percentage answers in different conditions of the group 2

As counterproof of the size's influence in detecting position's change we found, as predicted, that the majority of the subjects made incorrect distances estimation and their percentages plummeted when a change in size was not associated with a change in position (state III) in full screen as well as in HMD. We noticed, by using ANOVA test that there is no effect of presentations' order of the spheres (big/small or small/big) ($p > 0.48$). Subjects with good performance did not present a significant differences in their performance for different conditions ($p > 0.05$). Such performances could lead us to suppose that those subjects used different strategies to perform the task, a strategy not influenced at all by the size. For that purpose we carried out a second experiment with the subjects of the second group.

Experiment 2. We repeated the same tasks of the experiment 1 with the four subjects that showed good performance. They were asked which of the two spheres was easiest to reach by hand. They have to answer whether the first sphere or the second presented is the easiest to reach for different positions' states or both if they estimate that all of them are reachable.

Results. Although the experiment task was more difficult compared to the first one, the subjects had three possible answers instead of two alternative forced choice, the subjects showed good performance which confirmed their strategy used in the previous experiment was not based on the size as a dominant cue. Moreover, ANOVA test revealed no significant difference of group 2 performances in the experiment 1 and 2 with a static sphere in the four states ($p > 0.13$).

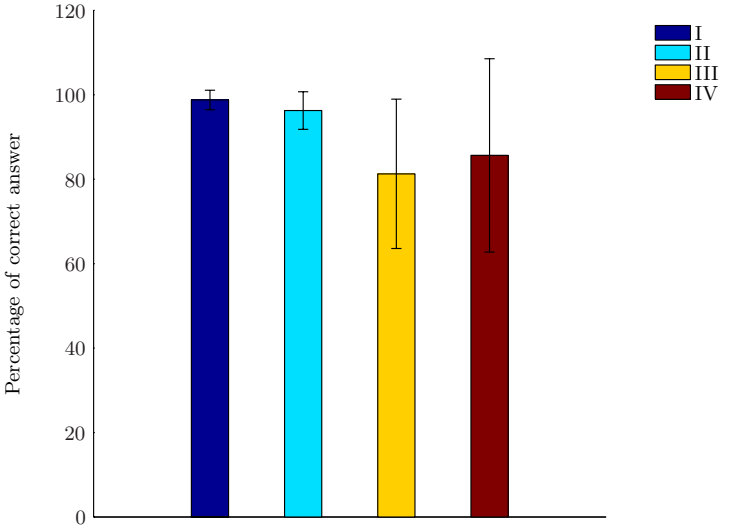


Fig. 5. Subjects' average percentage answers by state (I:IV) for the experiment 3

Experiment 3. We repeated the same task of the experiment one with 8 subjects, four of whom were new and naive to the task. The subjects of the group 1 (mentioned above) participated in this experiment. The subjects were asked the same question of the experiment 2; furthermore, we replaced the sphere by a cube with a self rotation.

Results. In this latter experiment, all the subjects performed their task with high rate of accuracy (figure 5). Comparing the performances of the first group (bad performer) in the experiments one and three, we noticed a significant difference of performance ($p < 0.00001$). Moreover, we noticed less variability in all subjects' performances, thus they were more accurate in this experiment than the previous one.

Experiment 4. We carried out the fourth experiment in order to see whether this enhancement in performance is due to the cube, the background displayed, or from the imagination of the motor task. In this experiment we used the same paradigm as the previous experiments. The subjects carried out the same task in three different conditions. The first condition used spheres with black background (Sphere BB). The second condition used spheres with white background (Sphere WB). Finally, the third condition used cube with white background (Cube WB). The order of conditions was mixed between subjects in order to avoid the learning process.

Participants. Ten subjects (7 males and 3 females) participants took part in the study. All of them did not participate in the experiments mentioned above. They all had normal or corrected to normal visual acuity.

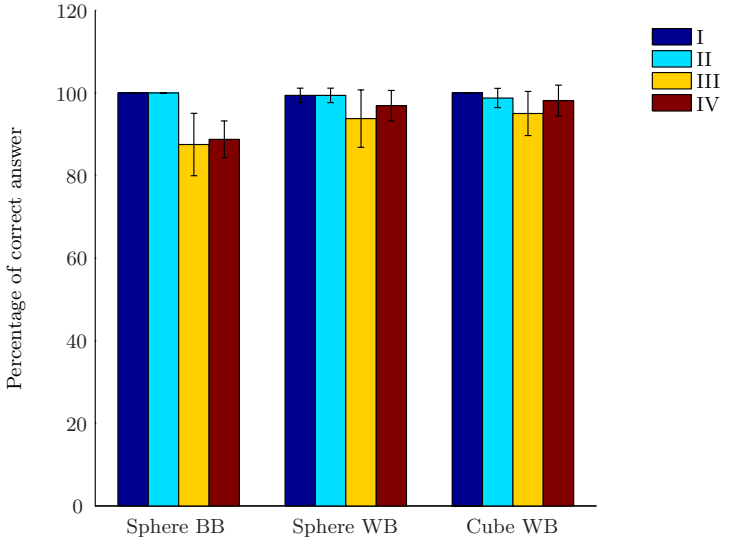
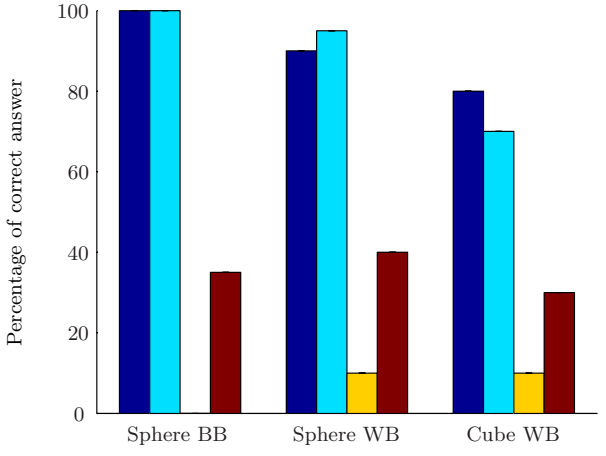


Fig. 6. Eight subjects with good performance average answers for the experiment 4

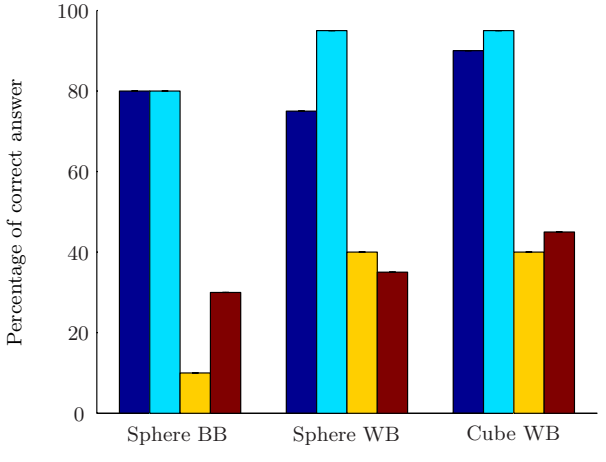
Results. In this experiment we recorded 8 subjects with good performance (figure 6) and two with bad ones (figure 7). All subjects were asked about which strategy that they have used in order to perceive the depth. Subjects with good performances have answered that the task was close to real for them and the virtual objects could be reachable by their hands. Basically they did not rely heavily on the size. On the contrary the other two with bad performances where they used only the size as a main cue in order to perceive the distances. For those subjects that showed a decorrelation between the size and distance there is no significant difference between the three states I, II and III ($p > 0.39$, $p > 0.35$ and $p > 0.08$) respectively. For the state IV there is a significant difference between the condition of the sphere with a black background compared with the conditions of a white background with a sphere and a cube ($p < 0.0002$ and $p < 0.0001$) respectively. This significance does not mean that the white background is the main variable that affect heavily the subjects' performance. In other words, the change of the background color is not the main parameter that could change the subject's performance from low to high rate of success. On the contrary there is no significant difference between the condition white background with a sphere or cube ($p > 0.51$). Thus, in our presentation the virtual objects (cube or sphere) have no effect in subjects' performances. Overall, these latter results are also seen in the other two subjects with bad performances.

4 Discussion

In this work, we wanted to investigate how perceptive variables (i.e., distance and size of the object, background, shape and task required) co-variances could af-



(a) subject 3



(b) subject 4

Fig. 7. Two subjects with bad performances for the experiment 4

fect distance change detections in a virtual environment. What we found in experiment 1 was an overall splitting of the subjects' performances into two groups: good and bad performers. The group with good performances in all conditions except the HMD one where the ANOVA test revealed a significant effect ($p < 0.0002$), whereas the former had good percentage of correct answers only for those states where stimulus size is physically constant but displayed in different positions (i.e., state I). Thus in the light of the data obtained we hypothesized two different strategies characterizing the two groups. Specifically we think that subjects defined as bad performer in the experiment 1 and 4, have performed their task by relying only on the size. In other words, these observers associated the change of size with a change of distance for perceiving the depth of the virtual ob-

jects. Likewise when they did not perceive any object's size changes they did not report any perception of change in position. Such strategy allowed this group to be more accurate only in states I and II, where size-distance constant hypothesis was respected and a change in size always accompanied a co-variation in position. On the contrary in states where either only size or distance was manipulated such size-related strategy led bad performers to a regular bias: reporting always no distance changes with same retinal size and different position, and reporting always the presence of distance changes in states of different size but same positions. On the other hand, different strategies lead good performers to be accurate also when size-distance coupling was decoupled. For those subjects we hypothesized an egocentric related strategy, in which they did not use the size as a main cue and they performed the task with estimating the difference (when occurred) in distance between the stimulus presented. An egocentric distance related strategy could explain first of all the good performers' success in all states and above all while in these relevant states also good performers failed in perceiving distances in the HMD condition. The less accuracy recorded in the subjects' performances of group 2 in the experiment 1 (states III and IV) by using the HMD coincides with what was founded by several studies on the distance perception in virtual environments using the HMD [11,12,13]. Moreover, the finding that in the HMD condition for the states (I and II) had no effect in the performances might us concludes that the reduction of accuracy of the good performers in using HMD in the other two states (III and IV) is not due to an instrumental artifact. Rather we deem that HMD, which increases the feeling of immersion but reduces egocentric estimations [12,13], simply played a critical role in those subjects who relied on a egocentric related strategy. To sum up, on the light of the obtained results one can divide the participants into the following categories

- Allocentric group: We suggest that the group of subjects of bad performance were perceived the stimulus as they were watching in 2D screen. In fact, they considered a spatial relations between two objects, and they relied on external environment to define distances [14,15]. Thus, they realized that they were far from the scene so as a consequence they have considered that the size is the only information that they could use in perceiving distance.
- Egocentric group: For the second group, on the contrary of the previous one, they estimated the distance of the spheres considering their body as reference such as the provided information of the size. By definition egocentric frames of reference define spatial positions using the body, or specific part of the body [14,15]. They believed that they were inside the scene and they felt the presence of the spheres as real ones [15]. However being the feel of presence magnified but egocentric distance perception decreased during HMD sessions this group reduced his accuracy level.

In the third experiment we noticed enhancement in subjects' performances. The fact of asking them to answer which cube was easiest to reach is to include the imagination of a motor task that leads to egocentric distance perception [16]. Indeed, this finding of high rate of accuracy when teh observers have used the imagination of motor task coincide with what was reported in [17,18]. Moreover, several

studies have shown that the subjects were more accurate in motor task on the contrary in the verbal one they were less accurate [19,20]. Indeed, it was shown that visuomotor control relies on representations which are centered on the individual, to execute the appropriate and accurate motor action; the individual needs to know where the object is regarding to them self and the effectors that will carry out the action [14,21]. According to Milner and Goodale's model the allocentric information is associated with perceptual tasks and egocentric information is associated with visuomotor tasks [14,21]. Finally in the experiment 4 we have seen that the variables of changing the sphere by a cube or the use of white background are not major parameters that could change completely the subject performance. Thus, this result could support more the current finding results.

5 Conclusion and Future Work

In this work, we investigated human depth perception in virtual reality. Our findings suggest that the verbal test for depth perception is not enough to make a complete evaluation of the human depth perception in VR. Our results have shown that the subjects were more accurate when imagining motor tasks. Furthermore, we found that there was not a large variability in subjects' performances through several experimental conditions in the experiment 4. On the contrary we noticed an increase in subjects' performances when we changed the question from verbal task to the imagination of movement tasks. By changing the question, we could help the subjects who showed bad performance in the experiment estimate correctly the egocentric-distances in the virtual environment. The subjects' performances in the HMD should be investigated deeper in order to situate what was obtained in [22] by limiting the subject's field of view in order to simulate the HMD real world.

Our future work will concern the same task with more variability. First of all, we will include in the stimulus presentation a fixation point to ensure that the observer does not rely on eyes movements in their task. The second point is to add the eye tracking to help to track eye movement from presentation to another. The third point will concern building up a protocol that can separate an allocentric task from the egocentric one. The fourth point is to make an evaluation of the depth perception in several angles of observers field-of-view and distances. Finally, we will build up an experimental setup that include a motor task in users' depth perception in addition of the verbal task for a global evaluation.

References

1. Kilpatrick, F.P., Ittelson, W.H.: The size-distance invariance hypothesis. Psychol. Rev. 60(4), 223–231 (1953)
2. Bradshaw, M.F., Rogers, B.J.: The interaction of binocular disparity and motion parallax in the computation of depth. Vision Res. 36(21), 3457–3468 (1996)
3. Gilinsky, A.S.: Perceived size and distance in visual space. Psychological Review 58, 460–482 (1951)

4. McCready, D.: On size, distance, and visual angle perception. Percept. Psychophys. 37(4), 323–334 (1985)
5. Loomis, J.M., Da Silva, J.A., Philbeck, J.W., Fukusima, S.S.: Visual perception of location and distance. Current Directions in Psychological Science 5, 72–77 (1996)
6. Louw, S., Smeets, J.B.J., Brenner, E.: Judging surface slant for placing objects: a role for motion parallax. Exp. Brain Res. 183(2), 149–158 (2007)
7. Loomis, J.M., Knapp, J.M.: Visual perception of egocentric distance in real and virtual environments. In: Hettinger, L.J., Haas, M.W. (eds.) Virtual and Adaptive Environments, pp. 21–46 (2003)
8. Interrante, V., Ries, B., Anderson, L.: Distance perception in immersive virtual environments, revisited. In: Proc. Virtual Reality Conference, pp. 3–10 (2006)
9. Gruber, H.E.: The relation of perceived size to perceived distance. The American Journal of Psychology 67, 411–426 (1954)
10. Berryhill, M.E., Fendrich, R., Olson, I.R.: Impaired distance perception and size constancy following bilateral occipitoparietal damage. Exp. Brain Res. 194(3), 381–393 (2009)
11. Witmer, B.G., Kline, P.B.: Judging perceived and traversed distance in virtual environments. Presence: Teleoper. Virtual Environ. 7(2), 144–167 (1998)
12. Thompson, W.B., Willemsen, P., Gooch, A.A., Creem-regehr, S.H., Loomis, J.M., Beall, A.C.: Does the quality of the computer graphics matter when judging distances in visually immersive environments (2002)
13. Messing, R., Durgin, F.H.: Distance perception and the visual horizon in head-mounted displays. ACM Trans. Appl. Percept. 2(3), 234–250 (2005)
14. Ball, K., Smith, D., Ellison, A., Schenk, T.: Both egocentric and allocentric cues support spatial priming in visual search. Neuropsychologia 47(6), 1585–1591 (2009)
15. Committeri, G., Galati, G., Paradis, A.L., Pizzamiglio, L., Berthoz, A., LeBihan, D.: Reference frames for spatial cognition: different brain areas are involved in viewer-, object-, and landmark-centered judgments about object location. J. Cogn. Neurosci. 16(9), 1517–1535 (2004)
16. Sakata, H., Taira, M., Kusunoki, M., Murata, A., Tanaka, Y.: The tins lecture the parietal association cortex in depth perception and visual control of hand action. Trends Neurosci. 20(8), 350–357 (1997)
17. Decety, J., Jeannerod, M., Prablanc, C.: The timing of mentally represented actions. Behav. Brain Res. 34(1-2), 35–42 (1989)
18. Mon-Williams, M., Tresilian, J.R.: The size-distance paradox is a cognitive phenomenon. Exp. Brain Res. 126(4), 578–582 (1999)
19. Armbrüster, C., Wolter, M., Kuhlen, T., Spijkers, W., Fimm, B.: Depth perception in virtual reality: distance estimations in peri- and extrapersonal space. Cyberpsychol. Behav. 11(1), 9–15 (2008)
20. Grossman, T., Balakrishnan, R.: An evaluation of depth perception on volumetric displays. In: AVI 2006: Proceedings of the working conference on Advanced visual interfaces, pp. 193–200. ACM, New York (2006)
21. Milner, A.D., Goodale, M.A.: Visual pathways to perception and action. Prog. Brain Res. 95, 317–337 (1993)
22. Knapp, J.M., Loomis, J.M.: Limited field of view of head-mounted displays is not the cause of distance underestimation in virtual environments. Presence 13, 572–577 (2004)

Combination of Annealing Particle Filter and Belief Propagation for 3D Upper Body Tracking

Ilaria Renna[1], Catherine Achard[1], and Ryad Chellali[2]

[1] Institut des Systèmes Intelligents et de Robotique, Université Pierre et Marie Curie/CNRS UMR 7222, BC 173, 4 place Jussieu, F-75005 Paris, France
[2] IIT - Italian Institute of Technology Robotics, Brain and Cognitive sciences Dpt. Intaro Team, Via Morego, 30 - 16012 Genova, Italy
{ilaria.renna,catherine.achard}@upmc.fr,
ryad.chellali@iit.it

Abstract. 3D upper body tracking and modeling is a topic greatly studied by the computer vision society because it is useful in a great number of applications such as human machine interface, companion robots animation or human activity analysis. However there is a challenging problem: the complexity of usual tracking algorithms, that exponentially increases with the dimension of the state vector, becomes too difficult to handle. To tackle this problem, we propose an approach that combines several annealing particle filters defined independently for each limb and belief propagation method to add geometrical constraints between individual filters. Experimental results on a real human gestures sequence will show that this combined approach leads to reliable results.

Keywords: Body tracking, particle filter, belief propagation.

1 Introduction

In the last years, the interaction between man and machines has become an important research topic, as it can generalize the use of robots from conventional robotics aspects to companion machines which interact with humans. The development of a robot companion poses several important challenges because the dynamics of the interactions with humans imply that a robot companion would have to exhibit cognitive capacities. Moreover, the communication between a human and an assistant robot must be as natural as possible. It is an evident fact that gestures are an important part of natural means that humans employ to communicate, especially when speech-based communication is not possible [1] or has to be completed by gesture (designating objects or locations for instance). So, one of the basic task of human-robot interaction deals with an accurate detection and real-time tracking of human body parts.

The previous problem can be stated as tracking articulated motion in images. This involves the localization and identification of a set of linked and articulated subparts in a three-dimensional space. Along with the difficulties of single object tracking, some additional problems appear for the tracking of articulated body due to the sizeable degrees of freedom of the target. In fact, a realistic articulated model of the human body is usually composed by at least 20 DoF, and as computational costs

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 824–833, 2009.

increase exponentially with the number of DoF, the exploration of the configuration space has to be optimized.

One way to solve the tracking problem is to use a probabilistic approach. Indeed, the motion analysis can be expressed as a Bayesian inference problem, and as the body parts are dependant the probability of a given configuration is conditioned by the upper body topology. Among known Bayesian solvers, one is well adapted for our problem, namely the particle filter. The strategies supported by this method allow the representation of the posterior distribution by a set of samples (or particles) with associated weights [2]. This set is updated over time taking into account the measurements (image features in our case), a prior knowledge on the system dynamics, and observation models. Unfortunately, it is well known that

1 – the number of particles required raises exponentially with the dimensionality of the configuration space.

2 – to have an accurate and plausible solution, we need a maximum of particles.

To avoid these antagonist requirements, we considered different developed methods that have been investigated with the aim to reduce the particles number. Some techniques proposed the Annealed Particle Filter (APF), which performs a coarse-to-fine layered search [3]. This modified particle filter uses a continuation principle based on annealing, to introduce the influence of narrow peaks in the fitness function, gradually. This allows to reduce by a factor of 10 the number of particles and, as a consequence, to significantly decrease computation times.

To adapt the previous approach to our problem, we represent 3D human body as a graphical model, where individual limbs are characterized by nodes and relationships between body parts are represented by edges connecting nodes and encoded by conditional probability distributions. Additional edges can also be introduced to manage partial or fully occlusions. This graphical model allows to track each subpart individually, and then to add constraints between adjacent limbs. By doing so, it was possible for us to add the Belief Propagation (BP) inference algorithm [4-9]. In this way, the initial high dimensionality problem is expressed as several problems of lower dimension, and thus the complexity of the search task is linear rather than exponential according to the number of body parts.

This article presents the development of a markerless human motion capture system that works with a standard camera coupled with a PC and does not require additional equipments. The system is based on a 3D articulated upper human body model and combines the advantages of above mentioned approaches to decrease the algorithm complexity induced by the high dimensionality of the problem. Rather than track the whole articulated body, each limb is tracked independently thanks to several particles filters (one for each limb); then, a BP method on factor graphs is used to estimate the current marginal of each limb according to geometrical constraints between limbs. Indeed, since belief propagation messages are represented as sums of weighted samples, the belief of each limb is approximated by a collection of samples. So, the association of belief propagation and particles filters algorithms is quite natural. Rather than a simple particle filter, we propose to use the annealing particle filter in this context. This combination of APF and BP allows to decrease the number of particles required per limb (and thus computation times) without modifying the quality of results.

The outline of the paper is as follows. In section 2, we introduce the graphical model that represents human body, while the APF method is explained in section 3. Section 4 sums up briefly the BP algorithm, introduces the proposed combination of APF and BP and its application to upper body tracking. Experiment results are shown and discussed in section 5, while conclusions are drawn in section 6.

2 The Articulated Body Model

The body is represented by a graphical model [4,10] of nodes, where each node in the graph corresponds to a body part, and each edge represents the spatial constraints between adjacent connected body parts. Each node has an associated configuration vector defining position and orientation of the body part in the 3D space and a corresponding image likelihood function that models the probability of observing image measurements conditioned on the position and orientation of the part. Each edge has an associated conditional distribution that models the probabilistic relationship between parts. Additional edges related to non collision constraints or to the propagation of state variables across the temporal domain are added. A factor graph [10] is then constructed to decompose the join probability as product of factors.

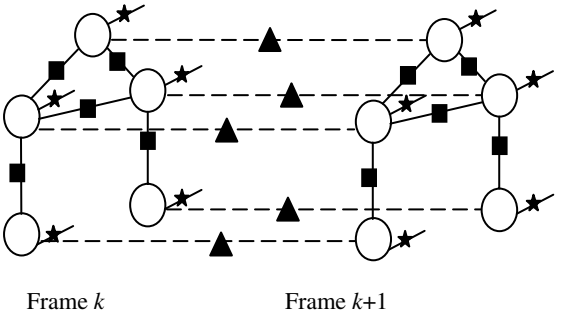


Fig. 1. Factor graph for a five part articulated structure. Circles represent state nodes, squares link factors, triangles time coherence factors and stars image likelihood factors.

The considered factors are of three different types: *link factors* between two nodes at the same time, *image likelihood factors* between all parts and their corresponding observations at the same time and *time coherence factors* that link a part at two consecutive times (figure 1). The individual motion of subparts is left to evolve and be detected independently, so that each subpart may be solved individually, while the full body is assembled by inference over the graphical model.

Let us denote with X_{μ}^{k} ($\mu \in [1, n]$) the state vector associated to limb μ at time k ($k \in [0, K]$), and with Y_{μ}^{k} the corresponding observation. We introduce the following model parameters of each limb μ:

- the interaction potentials (or link factors) $\Psi_{\mu,\nu}(X_\mu^k, X_\nu^k)$, which measure the likelihood between two connected body parts μ and ν;
- the observations probabilities $\Phi_\mu(X_\mu^k, Y_\mu^k)$, which measure the likelihood between the state vector and the corresponding observation (image likelihood factors);
- and the time coherence factors $T(X_\mu^k, X_\mu^{k-1})$, which determine the likelihood for the same limb between two consecutive times.

Then the joint probability at time k can be written as [8]:

$$P(X \mid Y) = \prod_{k=0}^{K} \Phi(X^k, Y^k)\Psi(X^k) \prod_{k=1}^{K} T(X^k, X^{k-1}), \tag{1}$$

where

$$\Phi(X^k, Y^k) = \prod_{\mu=1}^{n} \Phi_\mu(X_\mu^k, Y_\mu^k), \tag{2}$$

$$\Psi(X^k) = \prod_{(\mu,\nu)\in S} \Psi_{\mu,\nu}(X_\mu^k, X_\nu^k), \tag{3}$$

$$T(X^k, X^{k-1}) = \prod_{\mu=1}^{n} T(X_\mu^k, X_\mu^{k-1}), \tag{4}$$

and S is the set of all links between connected body parts.

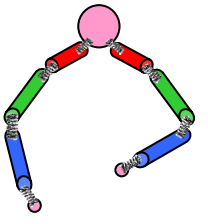


Fig. 2. Soft articulated human model

In our application of upper body human tracking, the graph is composed of nine nodes as we consider the motion of 9 limbs including the head, two clavicles, two arms, two forearms and two hands (figure 2). At each node is associated a five-dimensional space vector (x, y, z, φ, θ) except for the head represented by a four-dimensional one (x, y, z, θ). A 3D point and two angles are enough to localize a limb, such as an arm, modeled by a cylinder because the rotation of the limb around its main axis is not considered. For the head, as we suppose that it is faced to the camera, a 3D point and a single angle are employed. The first step of a tracking iteration consists to track each limb with a particle filter.

3 Particle Filtering

The classical filtering problem consists in estimating an unknown signal from observed measurements. In computer vision the observations are image sequences, and the discrete time steps are given by the frame rate. Particle filters approximate conditional densities as a collection of weighted point samples. These approximations are stochastically updated by using Monte Carlo methods on the set of weighted point samples.

Essentially, tracking with a classical particle filter works by performing three steps: (*i*) re-sampling from the weighted particle set obtained at the previous iteration, (*ii*) predicting stochastic movement and dispersion of particles, (*iii*) measuring and consequently updating the particle set. The main method drawback is that the number of particles needed to approximate the conditional densities grows exponentially as dimensionality increases. So, for a high dimensional problem such as human upper body tracking, the complexity becomes substantial.

The marginal probability of each limb is represented by a sum of *N* weighted particles:

$$P_{\mu}^{k}(X_{\mu}) \propto \sum_{i=1}^{N} \omega_{\mu,i}^{k} \delta(X_{\mu}^{k} - X_{\mu,i}^{k}), \tag{5}$$

where the weights are normalized so that $\sum_{i=1}^{N} \omega_{\mu,i}^{k} = 1$. Each particle $X_{\mu,i}^{k}$ represents a hypothetical position of the limb μ with a corresponding likelihood $\omega_{\mu,i}^{k}$ at time *k*. Then, the marginal probability density function is obtained recursively in a *prediction* and an *update* stage [2]. The stability and robustness of particle filters can often be improved by various methods. Among them, a particle-based stochastic search algorithm, called Annealed Particle Filtering (APF) [3], was developed. It uses a continuation principle, based on annealing, to introduce gradually the influence of narrow peaks in the conditional density function.

In each time-step, a multi-layered search (starting from layer $m = M$ to layer 1) is conducted. A smoothing of the weighting functions $\omega_{\mu,i}^{k,m} = (\omega_{\mu,i}^{k})^{\beta_m}$ is achieved by a set of values $\beta_M < \beta_{M-1} < \ldots < \beta_1$, where $\omega_{\mu,i}^{k}$ is the original weighting function.

A large β_m produces a peaked weighting function ω_m resulting in a high rate of annealing. Small values of β_m will have the opposite effect. At the same time, the amount of diffusion added to each successive annealing layer decreases according to $[P(X_{\mu,i}^{k} / X_{\mu,i}^{k-1})]^{\alpha_m}$, where the series α_m is such as $\alpha_M < \alpha_{M-1} < \ldots < \alpha_1$.

Method efficiency for recovering full articulated body motion depends on the choice of the tracking parameters α_m, β_m and on the particles number *N*. The purpose of the annealing filter is meanly to gain robustness and at the same time to reduce the particles number *N*.

At the end of an annealing iteration, the marginal probability of each limb is represented as a sum of weighted samples as for the standard particle filter. During this first step, the prediction and weighting are performed independently for each limb. Then, before considering the next iterations of annealing or the next frame, the final marginal probability is re-estimated with the belief propagation algorithm on the factor graph to take into account geometrical constraints between limbs.

4 Belief Propagation

In the BP algorithm, human body and relationships between body parts are represented by a graphical model, called factor graph. The factor graph is composed of a variable node for each variable X_μ (body part) and a factor node on each edge connecting variable node μ to variable node v. The algorithm recovers the pose of each body part by considering the relationships between every two adjacent body parts. Additional edges have also been introduced to add non collision constraint between some limbs (this avoids for example that the two hands remain together).

After designing interaction functions and observation functions, BP is used to search for body parts' belief by iteratively updating messages sent from a node to another one. Messages are propagated for all nodes at each frame for a variable number N_{BP} of iterations, and then propagated only once from a frame to the following one [8].

In practice, each message of the BP algorithm is approximated by a set of N weighted samples. The message $m_{v\to\mu,i}$ sent from the node v to the particle i of node μ is:

$$m^k_{v\to\mu,i} = \sum_j \Psi_{\mu,v}(X^k_{\mu,i}, X^k_{v,j})\Phi^k_{v,j}(X^k_{v,j}) \prod_{v'\in S(v)\setminus\mu} m^k_{v'\to v,j}\,, \tag{6}$$

where $S(v)\setminus\mu$ is the set of the neighbors of node v except μ, and $\Phi^k_{v,j}(X^k_{v,j})$ is the local likelihood of the sample j of node v. The belief at the node μ is then estimated by:

$$P^k_\mu(X_\mu) \sim \sum_{i=1}^{N} \hat{\omega}^k_{\mu,i}\delta(X^k_\mu - X^k_{\mu,i})\,, \tag{7}$$

where

$$\hat{\omega}^k_{\mu,i} = \Phi^k_{\mu,i}(X^k_{\mu,i}) \prod_{v\in S(\mu)} m^k_{v\to\mu,i}\,. \tag{8}$$

The resulting combined APF-BP algorithm is described in Table 1.

As we want to be able to track unconstrained human motions, any specific movement model could be used and the prediction is achieved according to Gaussian distributed diffusion. The amount of diffusion for each joint angle j is dependent on the image frames per second (fps). In our experiments σ_j ranges from 10 degrees to about 40 degrees to take in account the different extension field of each limb. For example, variations of rotation angles are smaller for clavicles than for hands.

The images features used to estimate the image compatibility factors $\Phi^k_{\mu,i}(X^k_{\mu,i}, Y^k_\mu)$ have to be strongly discriminant to allow the limb tracking. For each limb, the 3D model corresponding to the particle i of the state vector is projected on the image plane and the likelihood between this projection and image observations is estimated. This measure is based on oriented edge matching, motion energy and background subtraction. For the head and the two hands, factors are also based on a skin color

Table 1. Algorithm resulting from the combination of APF and BP methods

- Initialisation : $X_{\mu,i}^{0} = X_{\mu}^{0}$
- For each time k
  - For $m = M,\ldots, 1$ of the APF
    - Evolution: a new set of particles is drawn according to the system dynamics $[P(X_{\mu,i}^{k} / X_{\mu,i}^{k-1})]^{\alpha_m}$
    - Weight : compute the weight $\Phi_{\mu,i}^{k,m} = [P(X_{\mu,i}^{k}, Y_{\mu}^{k})]^{\beta_m}$
    - N_{BP} iterations of BP

      $m_{\nu\to\mu,i}^{k,m} = \sum_{j} \Psi_{\mu,\nu}(X_{\mu,i}^{k}, X_{\nu,j}^{k}) \Phi_{\nu,j}^{k,m}(X_{\nu,j}^{k}) \prod_{\nu' \in S(\nu)\setminus\mu} m_{\nu'\to\nu,j}^{k,m}$

      and re-estimation of $\omega_{\mu,i}^{k,m}$ with $\omega_{\mu,i}^{k,m} \propto \Phi_{\mu,i}^{k,m} \prod_{\nu\in S(\mu)} m_{\nu\to\mu,i}^{k,m}$
    - Resample the set of particles according to $\omega_{\mu,i}^{k,m}$
  - End for
- End for

probability map [11]. Link factors $\Psi_{\mu,\nu}(X_{\mu,i}^{k}, X_{\nu,j}^{k})$ representing the likelihood between two body parts are expressed as Gaussians of the distance between the reference points for the articulation between two successive limbs. For hands and head non collision constraints are added to avoid occlusions problems.

5 Experiments

This algorithm was tested on a sequence of images with a resolution of 360x288 pixels, acquired with a frequency of 15 frames per second with a PC HP, processor Intel ® Core™2 Duo, CPU 2.00GHz, 1.99GB of RAM. The video sequence shows a person making different everyday movements, like pointing, beckoning or waiting. The performance of our algorithm was evaluated by varying relevant model parameters, that are, the number of particles N, the layer number of APF M and the number of cycles of BP N_{BP}, with the proposal of reducing tracking time without losing robustness. Some sequence frames with the corresponding tracking results are shown in figure 3. These results are obtained with 100 particles per limb, 3 iterations of simulated annealing and 1 iteration of BP. It was very difficult to qualitatively evaluate the results. Furthermore, no ground truth was available for the collected data. Hence, for each image of the sequence, we decided to manually click on the body joints in order to obtain the coordinates of interest. Although these data are not completely accurate (for some body positions it is quite difficult to properly locate the coordinates of a joint), they constituted our 2D ground truth. This allowed to evaluate in a quantitative way, the precision of algorithm results: we computed the distance between the 2D

Fig. 3. Human posture estimations for some usual gestures

point of the ground truth and the projection on the image of the 3D point representing the limb articulation point. The articulation point was obtained making the mean between the values estimated by the algorithm for two consecutive limbs. These distances represent the tracking error for each articulation.

In order to quantify the advantages of the combination of simulated annealing and BP, we have performed several tests with different sets of parameters. Results reported on Table 2 are the mean distance errors for all the limbs (all these results have been obtained by averaging 5 realisations of tracking because of stochastic variables in the algorithm). Moreover, it is worth to precise that the algorithm has not been yet optimised. So only the order of magnitude between the different times has to be considered. In a first time we observe that just one BP cycle is enough to obtain a good cohesion between limbs (as shown in figure 3). Moreover when several iterations of BP are performed to reach the convergence, results are damaged probably because human body doesn't exactly validate the articulated model. Nevertheless, it is important to note that without BP cycles, all limbs are disconnected and the tracking diverges quickly.

Table 2. Error in pixel for different sets of parameters, comparing BP cycles. Times are measured in seconds.

Method	Error	$N_{BP} = 1$	$N_{BP} = 2$	$N_{BP} = 3$	$N_{BP} = 5$	$N_{BP} = 10$
$N = 100$, $M = 3$	Mean	4.4	5.0	5.2	5.1	5.0
	Standard deviation	2.9	3.3	3.5	3.3	3.3
	Computation time	0.75	0.75	0.84	0.82	1.15
$N = 300$ $M = 3$	Mean	4.2	4.5	4.5	4.7	4.6
	Standard deviation	3.1	3.3	3.2	3.4	3.3
	Computation time	3.49	3.49	3.87	4.75	5.89
$N = 300$ $M = 1$	Mean	4.7	5.8	5.9	5.4	5.3
	Standard deviation	2.9	3.3	3.6	3.7	3.3
	Computation time	1.43	1.42	1.52	1.89	2.34

The best results on Table 2 are obtained with 300 particles and 3 iterations of annealing but computation time is considerable (around 3.5 s). To reduce it, one can either decrease the number of particles or the number of annealing iterations. Errors presented on Table 2 show that it is better to preserve the annealing that allows to use fewer particles without tracking failure and with an acceptable computation time. In fact, in our algorithm 100 particles per limb with 3 annealing loops leads to accurate results with a reasonable error and a small standard deviation on this error. As shown in Table 3, the use of 3 APF layers allows to utilise less particles (just 100) keeping robustness with a lower computation time of 0.75s compared to a time of 1.43s when just a layer but more particles are used (results quality is then quite similar). Moreover, it appears that results on shoulder and wrist are more reliable: in fact, shoulders move more lightly than other limbs and wrist tracking is easier thanks to hands skin colour.

Table 3. Error in pixel for different sets of parameters

Method	Error	shoulder	elbow	wrist	Global error	Computation time
$N = 100$ $M = 3$ $N_{BP} = 1$	Mean	3.3	5.2	4.7	4.4	0.75
	Standard deviation	1.9	4.3	2.5	2.9	
$N = 300$ $M = 1$ $N_{BP} = 1$	Mean	3.2	5.4	5.4	4.7	1.43
	Standard deviation	1.9	4.0	2.7	2.9	

6 Conclusion and Future Work

In this paper we proposed a new approach for visual tracking of upper body using an algorithm that combines annealed particle filter and belief propagation methods. This algorithm shows that the use of this combination is effective to reduce the number of particles and, as a consequence, reduces computation time without losing robustness: (*i*) the use of simulated annealing decreases the number of particles used to track each

limb thanks to the introduction of narrow peaks in the fitness function; (*ii*) BP method assures spatial coherence and an independent tracking for each limb decreasing, moreover, the dimensionality of state space.

As perspective, we propose to use several cameras to improve the tracking and to learn gradually a colour model of the target. Another improvement that is possible concerns the implementation of the developed algorithm on a highly parallel architecture. Indeed, our formulation is well adapted to such architecture and execution times should be reduced consequently.

Furthermore, this body tracking algorithm is part of a more global project about humanoids animation that aims to understand gestures, how they are produced and what could be their role in human communication.

References

1. Krauss, R.M., Chen, Y., Chawla, P.: Nonverbal behavior and nonverbal communication: What do conversational hand gestures tell us? In: Zanna, M. (ed.) Advances in experimental social psychology, pp. 389–450. Academic Press, San Diego (1996)
2. Isard, M., Blake, A.: Conditional density propagation for visual tracking. International Journal of Computer Vision 29(1), 5–28 (1998)
3. Deutscher, J., Blake, A., Reid, I.: Articulated body motion capture by annealed particle filtering. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 126–133 (2000)
4. Yedidia, J.S., Freeman, W.T., Weiss, Y.: Constructing free energy approximations and generalized belief propagation algorithms. Technical Report 2004-040, MERL (May 2004)
5. Sigal, L., Isard, M., Sigelman, B.H., Black, M.J.: Attractive people: Assembling loose–limbed models using nonparametric belief propagation. Neural information processing systems (2003)
6. Sudderth, E.B., Ihler, A.T., Freeman, W.T., Willsky, A.S.: Nonparametric belief propagation. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 605–612 (2003)
7. Isard, M.: PAMPAS: Real–valued graphical models for computer vision. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 613–620 (2003)
8. Bernier, O., Cheung-Mon-Chan, P., Bouguet, A.: Fast nonparametric belief propagation for real-time stereo articulated body tracking. Computer Vision and Image Understanding 113, 29–47 (2009)
9. Shen, C., van den Hengel, A., Dick, A., Brooks, M.J.: 2D Articulated Tracking With Dynamic Bayesian Networks. In: Proceedings of the Fourth International Conference on Computer and Information Technology, pp. 130–136 (2004)
10. Kschischang, F.R., Frey, B.J., Loeliger, H.A.: Factor graphs and the sum-product algorithm. IEEE Trans. on Information Theory 47(9), 498–519 (2001)
11. Noriega, P., Bernier, O.: Multicues 3D Monocular Upper Body Tracking using Constrained Belief Propagation. In: British Machine Vision Conf., Warwick, UK, pp. 680–689 (2007)

Stereo Vision Based Floor Plane Extraction and Camera Pose Estimation

Lei Chen[1], Zhongli Wang[1,2], and Yunde Jia[1]

[1] Beijing Laboratory of Intelligent Information Technology, School of Computer Science, Beijing Institute of Technology, Beijing 100081, P.R. China
[2] State Key Lab. of Robotics and System, Harbin Institute of Technology, Harbin, P.R. China
{chenlei,wangzhongli,jiayunde}@bit.edu.cn

Abstract. In this paper, we present a unified approach for floor plane extraction and camera pose estimation. A histogram based method is performed on the reconstructed 3-D points obtained by an onboard video-rate stereo vision system to extract candidate points of the floor plane and determine the pose of vision system simultaneously. The obstacle area is easily localized given the floor plane region. In order to improve reliability and accuracy of camera pose estimation results, the Least Median of Squares (LMedS) based fitting method is applied to estimate the floor plane parameters with the extracted candidate points. The precise pose of the onboard stereo vision system is directly acquired related to the floor plane parameters. Experimental results in real indoor environments are discussed and show the good performance.

1 Introduction

Perception of floor plane is important information for the mobile robot in unknown and unstructured environments. For this purpose, several stereo vision based approaches have been proposed to address the floor plane extraction problem during recent years. According to the representation of the data space, the existing state-of-art stereo vision based approaches can be broadly classified into two categories: the Euclidean space based method and the disparity space based method. The 3-D workspace is explicitly expressed in the first category, and the floor plane is generally acquired by fitting the plane model to the available 3-D points in the Euclidean space. Sappa et al. [1] as well as Burschka et al. [2] apply the Random Sample Consensus (RANSAC) based approach to estimate the plane parameters in the candidate floor region, which is obtained by selecting points according to a probability function that takes into account the density of points at a given depth. Okada et al. [3] as well as Sabe et al. [4] use the randomized Hough transform to vote for the 3-D plane parameters in the transformed space. Each point is associated with the closest plane candidate. The expectation maximization algorithm is employed by Thrun et al. [5] and Triebel et al. [6] to estimate the number of the surfaces and their location simultaneously. In the disparity space based methods, the 3-D workspace is inherently derived from the disparity map. Labayrade et al. [7] extract

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 834–845, 2009.

the road geometry in outdoor environments through the v-disparity image in which ground plane is projected as the ground correlation line. Broggi et al. [8] improve the v-disparity approach by taking into account the local features in the image. The Hough transform is usually adopted for computing the ground correlation line parameters. Trucco et al. [9] provide a close-form expression for the disparity of a plane. For each point in the disparity map, plane parameters are calculated by the local neighborhood points with the least-squares fitting method. Planes are extracted by carrying out clustering in the plane parameter space. Rosselot and Hall [10] present a XH-map algorithm which is likened to a histogram reduction method for removing the floor plane background in the disparity map. An iterative segmentation estimation framework is proposed by Thakoor et al. [11] for planar surfaces segmentation using the maximum likelihood estimation method.

The aforementioned approaches separately process the floor plane point extraction and camera pose estimation, which means before the camera pose estimation, the candidate points belonging to the floor plane must be extracted first. Moreover, if the fitting process just uses the candidate points from a local neighborhood, the results are easily subject to the effects of unavoidable noise data, and it leads to inaccurate and unstable floor plane parameters.

In this paper, we propose a unified approach to extract the floor plane and estimate the pose of the vision system simultaneously. It is performed on the 3-D points in the Euclidean space which are transformed from the dense disparity maps provided by the onboard real-time stereo vision system. The histogram information of the OY axis direction in the reference camera coordinate system is calculated to extract all the candidate points belonging to the floor plane. The areas occupied by obstacles or other objects can be detected by removing the floor plane region. To improve the estimation results further, these points are used to reliably estimate the floor plane parameters by means of the Least Median of Squares (LMedS) based method. The accurate pose of the onboard stereo vision system can be directly determined from the floor plane fitting parameters.

The remainder of this paper is organized as follows. Section 2 briefly describes the real-time stereo vision system used for acquiring dense disparity maps. Section 3 provides the floor plane model and the extraction method of the floor plane is presented in Sect. 4. The approach for reliable and accurate pose estimation of the onboard vision system is described in Sect. 5. Experiments in indoor scenarios are provided in Sect. 6. Section 7 gives the conclusions.

2 Real-Time Stereo Vision System

The dense disparity maps used for the extraction of the floor plane is provided by an onboard stereo vision system [12], which will be installed on a P3-AT mobile robot. As shown in Fig. 1, the system is mainly composed of three sub-modules, a stereo head with three CMOS image sensors, an FPGA-based real-time processing unit, and a communication interface based on IEEE 1394, and can achieve a throughput of more than 30fps with 320×240 pixels dense disparity maps in 32-pixel disparity searching range.

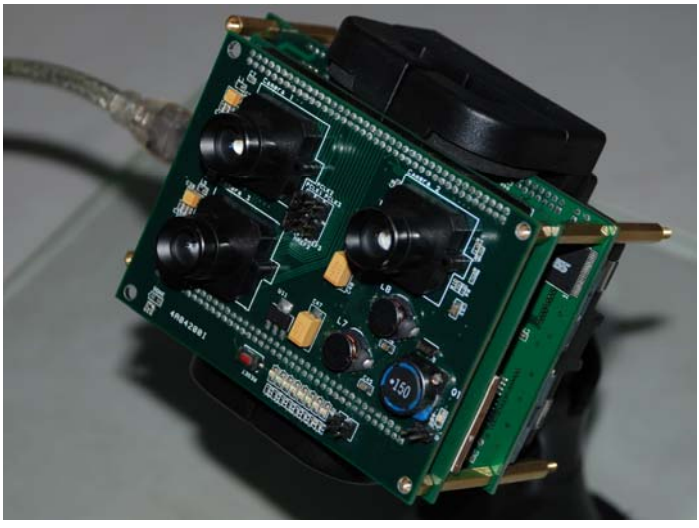


Fig. 1. The real-time stereo vision system

Three image sensors are configured with right-angled triangular position on the stereo head and can synchronously capture 640×480 pixels grayscale or color images at 30fps. The trinocular vision system, with little increasing of hardware cost, can achieve better results than a typical binocular stereo vision system by providing a second pair of image sensors for reducing the ambiguity greatly.

The FPGA-based real-time processing unit mainly consists of a high performance FPGA and tightly coupled local SRAMs. The FPGA serves as the processing core of the whole system and its principal functions can be summarized as follows: (a) supplying external synchronous signal for all image sensors because images for stereo matching must be captured simultaneously; (b) executing stereo matching computation to calculate disparities; (c) managing data access control of SRAMs; (d) providing all the necessary glue logic with the stereo head and the communication interface.

The communication interface mainly includes an FPGA, a DSP and IEEE1394 chipsets. The FPGA controls the data transmission. DSP and IEEE1394 chipsets implement the IEEE 1394a high serial bus protocol for transferring data to the host computer at a high speed of 400Mbps.

3 The Floor Plane Model in the Camera Frame

Disparity map inherently represents the 3-D information of the scene in the reference camera coordinate system ($OX_CY_CZ_C$) of the stereo vision system. Basically, the approach proposed in this paper assumes that the floor plane is flat and it is a predominant geometry in the given workspace.

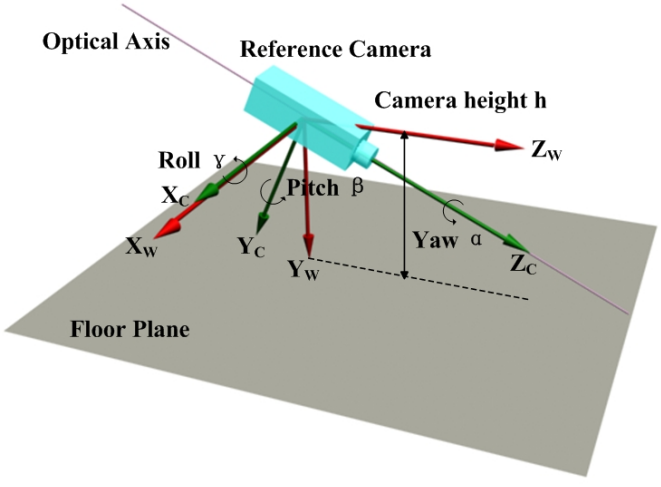


Fig. 2. The geometry between the reference camera and the floor plane. The reference camera frame $OX_CY_CZ_C$ is represented in green and the world frame $OX_WY_WZ_W$ is represented in red.

As illustrated in Fig. 2, the origins of the reference camera coordinate system and the world coordinate system are both set to the optical center of the reference camera, and X_WZ_W plane in the world coordinate system is assumed to be parallel to the floor plane. The expression of the floor plane $ax + by + cz = 1$ is obtained by dividing the general plane equation $ax + by + cz + d = 0$ by $-d$, since we already know that d is not equal to zero. If the angle between the optical axis of the reference camera and the floor plane as well as the height of the reference camera is obtained, then the pose of the reference camera can be determined. The camera's height (h) and orientation (θ), referred to the floor plane parameters (a, b, c), are given by

$$h = \frac{1}{\sqrt{a^2 + b^2 + c^2}}$$
$$\theta = \arctan(\frac{c}{b}) \quad . \tag{1}$$

4 Extraction of the Floor Plane

The extraction process consists of two steps. Initially, 3-D points in the disparity space are mapped onto the Euclidean space. A histogram based method is adopted in the Euclidean space to extract all the candidate points belonging to the floor plane and obtain pose rough estimation of the onboard vision system simultaneously. Then, the Least Median of Squares (LMedS) based fitting method is applied to acquire the floor plane parameters with the extracted candidate points in the first step. Finally, the

precise pose of the onboard stereo vision system is directly determined through the fitted parameters. Figure 3 presents the flowchart of the proposed approach.

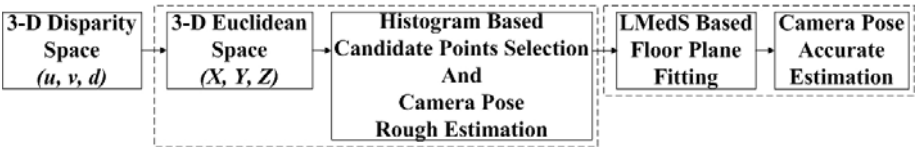


Fig. 3. Algorithm flowchart

Since the points belonging to the floor plane are extracted from the points in the Euclidean space in the reference camera coordinate system, a mapping process from the disparity space onto the Euclidean space is first executed.

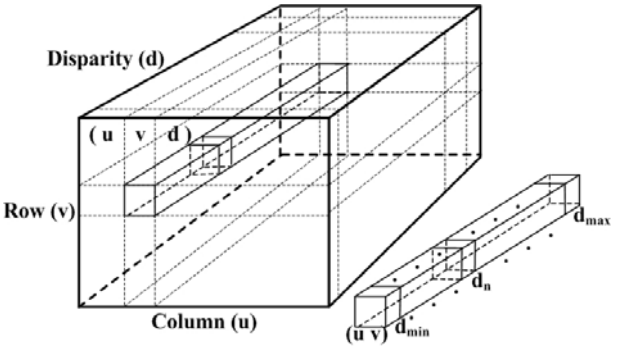


Fig. 4. Disparity space description. d_{min} and d_{max} are the minimum and maximum value of the valid disparity search range of the stereo vision system respectively

As shown in Fig. 4, a point in the disparity space can be expressed with (*u, v, d*), where (*u, v*) are the coordinates in the disparity map and *d* is the corresponding disparity value.

For the stereo vision system used in this paper, the transformation from the disparity space (*u, v, d*) to the Euclidean space (X_C, Y_C, Z_C) related to the reference camera coordinate system can be expressed as

$$\begin{aligned} X_C &= \frac{Z_C(u-U_0)}{f/d_x} \\ Y_C &= \frac{Z_C(v-V_0)}{f/d_y} \\ Z_C &= \frac{B}{d} f/d_x \end{aligned} \quad , \tag{2}$$

where (U_0, V_0) is the principal point[1] of the reference camera, f denotes the focal length of the reference camera, d_x and d_y represent the pixel size in the image column and row direction respectively and B is the stereo baseline. The intrinsic parameters such as f, d_x, d_y of the reference camera are pre-calibrated.

For the extraction of the floor plane candidate points, a pure rotation transformation is performed on the 3-D points (X_C, Y_C, Z_C) in the original camera frame to obtain the corresponding rotated points (X_R, Y_R, Z_R) in the rotated frame, and then all the obtained points (X_R, Y_R, Z_R) are projected to the OY axis in the rotated frame for computing a histogram along the OY axis. The rotation transformation can be given by

$$(X_R \quad Y_R \quad Z_R)^T = R(X_c \quad Y_c \quad Z_c)^T , \tag{3}$$

where R denotes the rotation matrix. According to the relationship between the camera and the floor plane shown in Fig. 2, the rotation matrix R is

$$R = \begin{pmatrix} 1 & 0 & 0 \\ 0 & cos\phi & -sin\phi \\ 0 & sin\phi & cos\phi \end{pmatrix}, \tag{4}$$

Where ϕ is the rotation angle. For each presetting ϕi, the histogram HG_i along the OY axis in the new rotated frame is obtained and the peak value P_i in the corresponding HG_i is picked out. After all the histograms HG_i correspondent to presetting ϕi are calculated, a curve L including all the peak values P_i with each presetting ϕi is plotted out as shown in Fig. 5.

We assume that the floor plane is flat with a predominant geometry in the field of view (FOV). Therefore, the maximum value P_{max} of the curve L should appear at

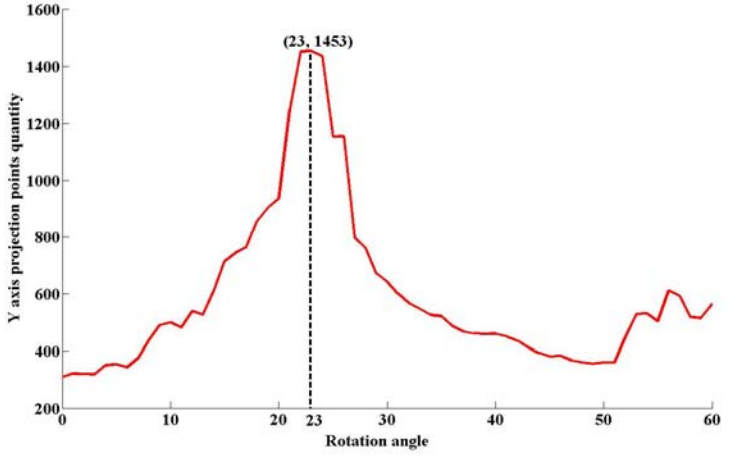


Fig. 5. The histogram peak values related to the rotation angle ϕ

[1] Assumed to be at the image center.

where the rotation angle ϕ is ideally equal to the unknown angle between the optical axis of the reference camera and the floor plane as seen in Fig. 5. Practically, we can experientially preset a range of rotation angle ϕ (e.g., zero degree to sixty degree in Fig. 5) with the prior knowledge. Some fast searching strategies, such as binary search strategy, can be applied to accelerate the searching process.

Once P_{max} is determined, the corresponding histogram *HG* will be used to identify the floor plane candidate points. In the ideal case, the floor plane points are all projected to the same value which is also the peak value in the correspondent histogram. However, because of the noises, the floor plane points will distributed in a range other than the same value.

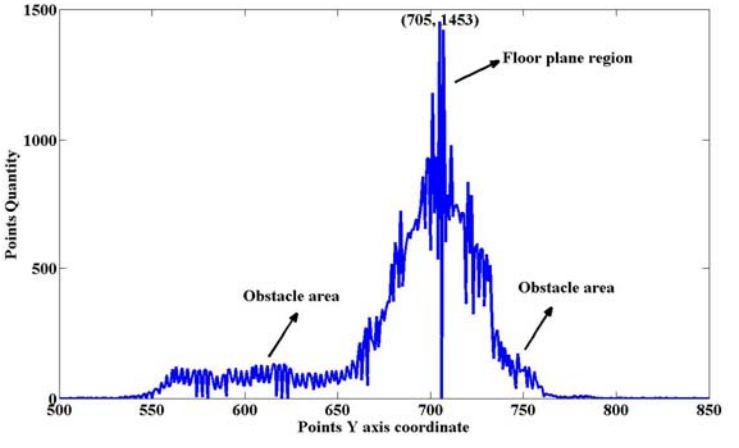


Fig. 6. The histogram correspondent to the rotation angle $\phi = 23°$

Figure 6 shows the histogram in a real condition when the rotation angle ϕ is equal to 23 degree which is determined in Fig. 5. The peak value of the histogram in Fig. 6 is contributed by the points belonging to the floor plane and should correspond to the camera height *h*. As far as noise data in the disparity maps is concerned, the floor plane candidate points are extracted following two rules below:

1. Points belonging to the floor plane are almost mapped within a small range around the peak value in the histogram.
2. In contrast to the obstacle, the peak value of the points belonging to the floor plane is much higher.

In terms of the statistic results shown in Fig. 5 and Fig. 6, all the points belonging to the floor plane are extracted out in the reference camera; meanwhile, we can acquire an estimation of the camera's height and orientation: h = 705 mm and $\theta = 23^{\circ}$. However, because of the mismatching or errors for the textureless regions

in the image, there exists unavoidable outliers among the extracted floor plane candidate points, and thus it leads to inaccurate pose estimation. In the following section, a robust and more accurate camera pose estimation process is employed to improve the estimation results. Once given the position information of the floor plane, the obstacle region in the workspace can be easily determined by removing the part of the floor plane area.

5 Camera Pose Estimation

It is obvious that the more accurate the floor plane parameters are, the better estimation of camera pose achieves. Therefore, a robust plane fitting method which can resist the effects of outliers in fitting plane model should be applied to tackle the camera pose estimation problem. In this paper, the Least Median of Squares (LMedS) method [13] is adopted which holds the advantages of excellent global robustness and high break down points (i.e., 50%).

A two-step method is proposed for the precise and reliable camera pose estimation in this section. In the first step, the LMedS based method is employed to fit the plane model $ax + by + cz = 1$. Then, with the fitted parameters (a, b, c), the pose of the reference camera is directly estimated according to Eqs. (1).

The basic procedures for the LMedS based floor plane model fitting are completed by iteratively executing the following three steps with N times:

1. Draw a random sampling of three candidate points in the Euclidean space.
2. Compute the plane parameters (a, b, c) with the selected three points.
3. Compute the median squared residual. For the plane obtained in the step 2, calculate the squared residuals for all the 3-D floor candidate points, then sort and find the middle one.

The random sampling is repeated N times so that the probability P that at least one of the sample in the N samples is "clean" is almost 1 (e.g., $P = 0.99$). "Clean" means observation is good and not contaminated by the outliers. The sample times N is given by

$$N = \frac{\log(1-P)}{\log[1-(1-\varepsilon)^m]}, \tag{5}$$

where ε denotes the fraction of outliers contained in the whole set of floor candidate points and m is the sample size. m is set to 3 because three points are necessary to define a plane.

After random sampling procedure, the correct fitting will be correspondent to the one with the least value among N median squared residuals. The calculated fitting parameters will be selected as the floor plane parameters.

To effectively and precisely estimate parameters of the fitting plane, the random sampling procedure suffices the following three constraints:

1. Each sampling set is different from others in the whole N sampling sets.
2. The sampling points are selected among points which are close to the camera, since for the task of the floor plane fitting, the larger the disparity is, the more accurate plane fitting parameters are [4]. In practical implementation, we select point SP_i whose disparity value is larger than a predefined threshold $Disparity_T$

$$\forall SP_i \quad i \in [1, 3*N] \quad d(SP_i) > Dispairty_T \quad . \tag{6}$$

3. Three points in each sampling set should not be too close to each other in order to achieve a coverage to the whole floor plane candidate points. In practical implementation, the Euclidean distance between random two points SP_i and SP_j among three points (SP_1, SP_2, SP_3) in one sampling set should exceed a predefined threshold $Distance_T$

$$\forall SP_i, SP_j \in \{SP_1, SP_2, SP_3\} \quad \left|SP_i - SP_j\right| > Distance_T \ (i \neq j) \quad . \tag{7}$$

The constraints are tunable via thresholds. The trade-off is between quality and quantity of the data. If the thresholds are tuned too high, it may not give enough coverage to the whole data set. On the other hand, the procedure will be subject to errors.

6 Experimental Results

The proposed method has been tested in different indoor environments. The resolutions of both disparity and grayscale images are 320×240 and the valid disparity range is between 1 and 32 for the disparity map. In the disparity map, the closer the objects to the camera, the lighter the intensity level is, and black areas represent the unconfident region or points which are rejected.

Figure 7 displays the extraction results of three typical cases. The first row illustrates the experimental results of a floor plane with a stone. The second row demonstrates the experimental results with no objects except the floor plane, and the third row shows the experimental results of the floor plane with a thin and high pole. The first and second columns respectively represent the grayscale images and disparity maps provided by the stereo vision system. The third column shows the results of floor plane extraction, all the floor plane points are removed in the disparity maps but the obstacles.

Because of the unconfident disparities in the areas with little visual texture, some pixels belonging to the floor plane fail to be removed in the disparity map. However, it is obvious that most floor plane pixels are extracted successfully and the obstacles in the environments can be localized by removing the floor plane region. Moreover, this benefits the fitting process of the floor plane and the pose estimation of the vision system.

Table 1 indicates the obtained plane parameters (a, b, c) using the LMedS based plane fitting method with $P = 0.99$, $\varepsilon = 0.4$, $Disparity_T = 22$ and $Distance_T = 300$mm. Table 2 shows results of the onboard vision system pose estimation related to the floor plane parameters in Table1.

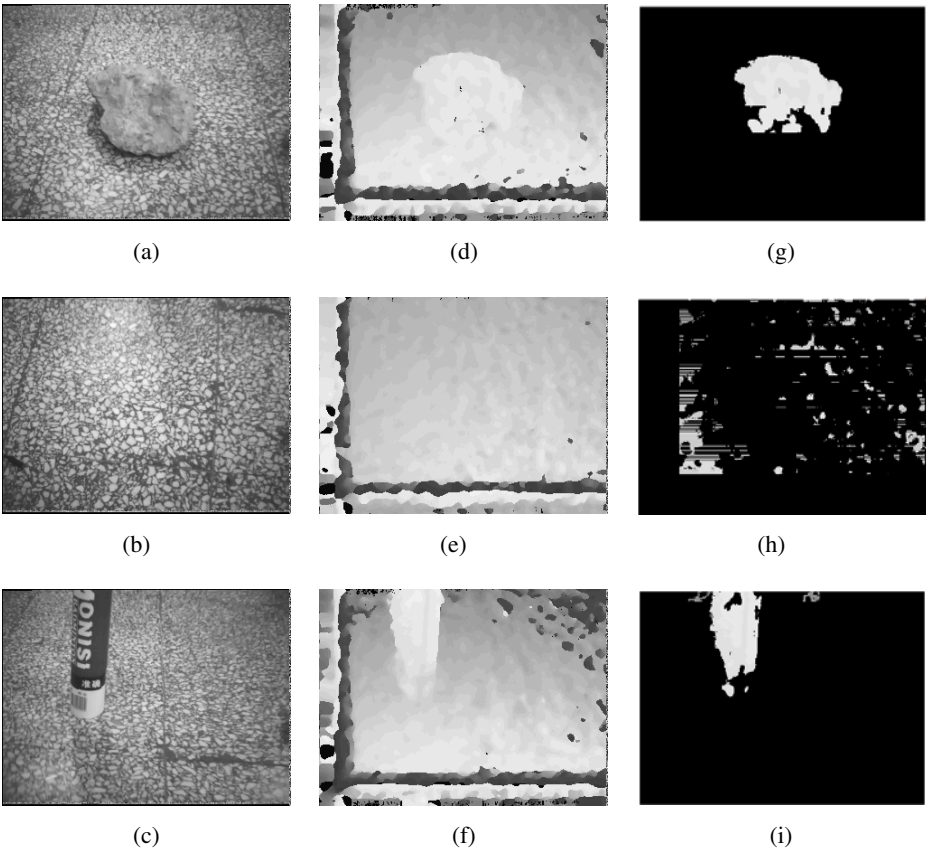


Fig. 7. The results of the floor plane extraction. The extracted points belonging to the floor plane are set to zero (black) in the disparity map.

Table 1. The floor palne fitting parameters

Image	a	b	c
Fig. 6 (a)	-0.001005564	0.001509615	0.000672976
Fig. 6 (b)	-0.000003862	0.000852063	0.000511784
Fig. 6 (c)	-0.001626025	0.001513240	0.000694761

Table 2. Reference camera pose estimation

Image	height h (mm)	orientation θ (°)
Fig. 6 (a)	516.9	24.0
Fig. 6 (b)	1006.2	31.0
Fig. 6 (c)	429.7	24.7

7 Conclusions

We have proposed a unified approach for floor plane extraction and camera pose estimation in indoor environments simultaneously. To extract the floor plane points, a histogram based method is applied to 3-D points in the Euclidean space, which are transformed from the disparity maps. The fitting parameters of the floor plane in the reference camera coordinate system are estimated first by the Least Median of Squares (LMedS) method, which is based on the extracted 3-D candidate points belonging to the floor plane. Then the parameters of the onboard stereo vision system pose, including the orientation and height, are accurately estimated. The performance of proposed approach has been validated in the indoor environments.

Acknowledgements

This work was partially supported by the Natural Science Foundation of China (60675021), the Chinese High-Tech Program (2009AA01Z323), and Beijing key discipline program, and partially supported by State Key Laboratory of Robotics and System, Harbin Institute of Technology (HIT).

References

1. Sappa, A.D., Dornaika, F., Ponsa, D., Gerónimo, D., López, A.: An Efficient Approach to Onboard Stereo Vision System Pose Estimation. IEEE Transactions on Intelligent Transportation Systems 9(3), 476–490 (2008)
2. Burschka, D., Hager, G.: Scene Classification from Dense Disparity Maps in Indoor Environments. In: International Conference on Pattern Recognition, pp. 708–712 (2002)
3. Okada, K., Kagami, S., Inaba, M., Inoue, H.: Plane Segment Finder: Algorithm, Implementation and Applications. In: IEEE International Conference on Robotics and Automation, pp. 2120–2125 (2001)
4. Sabe, K., Fukuchi, M., Gutmann, J., Ohashi, T., Kawamoto, K., Yoshigahara, T.: Obstacle Avoidance and Path Planning for Humanoid Robots using Stereo Vision. In: IEEE International Conference on Robotics and Automation, pp. 592–597 (2004)
5. Thrun, S., Martin, C., Liu, Y., Hähnel, D., Emery-Montemerlo, R., Chakrabarti, D., Burgard, W.: A Real-Time Expectation Maximization Algorithm for Acquiring Multiplanar Maps of Indoor Environments with Mobile Robots. IEEE Transactions on Robotics and Automation 20(3), 433–443 (2004)
6. Triebel, R., Burgard, W., Dellaert, F.: Using Hierarchical EM to Extract Planes from 3D Range Scans. In: IEEE International Conference on Robotics and Automation, pp. 4437–4442 (2005)
7. Labayrade, R., Aubert, D., Tarel, J.: Real Time Obstacle Detection in Stereovision on Non Flat Road Geometry Through "V-disparity" Representation. In: IEEE Intelligent Vehicle Symposium, pp. 646–651 (2002)
8. Broggi, A., Caraffi, C., Fedriga, R.I., Grisleri, P.: Obstacle Detection with Stereo Vision for Off-Road Vehicle Navigation. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 65–72 (2005)

9. Trucco, E., Isgrò, F., Bracchi, F.: Plane Detection in Disparity Space. In: International Conference on Visual Information Engineering, pp. 73–76 (2003)
10. Rosselot, D., Hall, E.L.: The XH-Map Algorithm: A Method to Process Stereo Video to Produce a Real-Time Obstacle Map. In: SPIE, vol. 6006(1), pp. 1–11 (2005)
11. Thakoor, N., Jung, S., Gao, J.: Real-time Planar Surface Segmentation in Disparity Space. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2007)
12. Chen, L., Jia, Y.: A Parallel Reconfigurable Architecture for Real-Time Stereo Vision. In: International Conference on Embedded Software and Systems, pp. 32–39 (2009)
13. Rousseeuw, P.J.: Least Median of Squares Regression. Journal of the American Statistical Association 79(388), 871–880 (1984)

Autonomous Mapping Using a Flexible Region Map for Novelty Detection

Muhammad Fahmi Miskon and Andrew R. Russell

Intelligent Robotic Research Center (IRRC), Monash University, Australia
{muhammad.miskon,andy.russell}@eng.monash.edu.au

Abstract. This paper presents an autonomous method for a robot to map the normal condition of its environment using a flexible region map. The map is used as a reference to allow a mobile robot to perform novelty detection. The map has a flexible structure which could accommodate to the distribution of different types of entity in the environment. However, updating information in the map for autonomous mapping is not a trivial task since it requires changing the structure of the map. The contribution of this paper is twofold. First, the method for reshaping a flexible region map is discussed. Then, an algorithm that is inspired by the habituation principal for performing autonomous update is presented. Experimental results show that autonomous update was achieved by using the habituation principal and by allowing the flexible region to reshape itself to accommodate to changes in the environment.

Keywords: Novelty detection, mobile robot, map referencing, autonomous mapping.

1 Introduction

Novelty detection is a mechanism to highlight unusual conditions in an otherwise normal environment [1]. Recently, novelty detection has been used by mobile robots to do many task like collision detection [2], to assist learning [3], to focus attention [4] and inspection [5-7]. Brooks in [8] suggested that, in order to appear intelligent, a robot needs to display a number of different behaviors that operate in parallel, with behaviors subsuming each other as required by the application and environment. Placed at level 5 of his suggested task achieving behaviors is monitoring changes, of which novelty detection is an important part. In the future all domestic/office robots should be generally aware of their environment and although they may be cleaning/delivering/etc. robots, they should also be able to monitor changes and report anomalies just as we would expect a human doing the same task to do.

In order to perform novelty detection in an extended environment, a robot needs to refer to a thematic map which associates locations in the environment with their normal condition. The benefit of autonomous mapping is that any new robot or a robot which is equipped with a new sensor could easily learn its new environment or adapt to changes in the environment.

In this project, a flexible region map from our previous work [9] is employed to map the normal condition in the environment. The reason is that a flexible region map

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 846–855, 2009.

is data driven and has a flexible structure. This means that reference is only made when and where the data is available, unlike grid maps [10, 11] or quadtrees [12, 13]. In addition to that, the size of its cells can change to suit the distribution of the data, as opposed to any normal perception based map like in [7]. As a result, the size of storage for the map can be reduced. This is of benefit to mobile robots which have a relatively small processing capability and limited power supply from batteries. These restrictions are especially true for small mobile robots or for a novelty detection mechanism which runs on a standalone subsystem. However, the advantages come with a price. When the structure of the regions varies, extra effort is required to update information associated with them.

Motivated by the benefits of autonomous mapping using a flexible region map, a method to autonomously update a flexible region map and learn the normal representation of an environment has been developed.

The rest of this paper is organized as follows. An overview of the autonomous mapping is given in Section 2. This is followed by a detail description of the map representation and the autonomous mapping algorithm in Section 3 and Section 4. The experimental results are given in Section 5. The paper is concluded in Section 6.

2 System Overview

The main sensor used for this project is a Hokuyo URG 400 laser range finder. It has 0.35° bearing resolution and a scanning range of 270° with 1 mm range resolution. The sensor is mounted on a Pioneer 3 mobile robot manufactured by ActivMedia Robotics. The robot also carries other types of sensors which are used for monitoring the state of the environment.

The robot navigates by wall following and localizes itself using a particle filter taking information from laser measurements and odometry. In order to maintain approximately the same heading, the robot aligns itself parallel to the wall every time it performs novelty detection. The robot employs the Habituating Self Organizing Map [14] as its novelty filter. There are many different novelty detection approaches described in the literature. Many are application specific. It is was found that the HSOM [14] has many characteristics which make it suitable for the system described here. First, it has been used for a mobile robot application before. Second, it is highly adaptable to a new environment. Third, being based on a neural network architecture, no model is required, which is of benefit when using a multi sensory system like the one described here. Each sensor has its own neural network.

A laser range finder is used as the main sensor for the experiment described in this paper because being sensitive to robot pose; it requires more design considerations. The laser measurements are divided into 8 sectors which are then averaged to provide the input to the neural network. By combining laser range measurements into a reduced number of sectors, it appears like the system does not take full advantage of having such high resolution information. However, there are several reasons why the laser measurements are down sampled. First, the sensitivity of the laser needs to be reduced to tolerate the difference between measurements especially due to robot's inability to achieve a target pose. Second, as the motivation of the work in this project

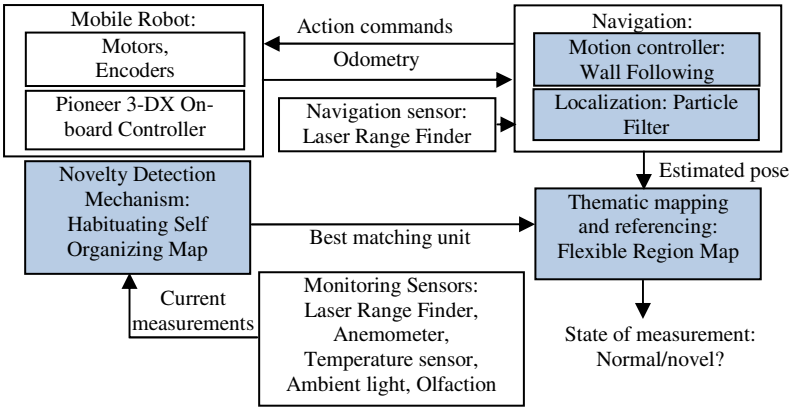


Fig. 1. System overview

is to reduced storage, this would certainly reduce the amount of storage required by the system. Third, the down sampling demonstrates that the system will still work with other cheaper but less accurate alternatives like sonar and infra red sensors.

3 Flexible Region Map

A detailed description of our earlier work regarding the flexible region map is presented in [9]. In this project, a new process i.e. separation is introduced.

3.1 Shape and Description of a Region

The Minimum Bounding Rectangle (MBR) is known to be the computationally simplest of all the linear bounding containers [15]. In this work, the MBR is described using its width, w and length, l extending from a reference point (x,y). A region is associated with two pieces of information; robot pose and normal data perceived at the pose. The normal data is represented using neurons from the Habituating Self Organizing Map (HSOM) architecture.

3.2 Region Tolerance, R_T

Region tolerance is the maximum allowable difference between perceptions at the target pose on a planned path and other poses in a region. The minimum value of R_T is affected by the robot positioning accuracy which is how close the robot can move to its target pose as well as localization error. The higher the value of R_T, the smaller is the number of regions needed to represent an environment but the novelty filter will be less sensitive.

3.3 Creation and Deletion

A new region (R_i) is initiated by establishing a w_{init} mm wide and l_{init} mm long rectangle with the centre being the position of the sensor/robot (see Fig. 2). The values of the

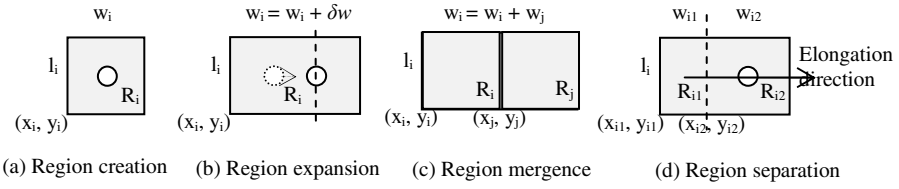


Fig. 2. Processes involved in updating a flexible region map.

initial width, w_{init} and the initial length, l_{init} depend on the minimum value of the region tolerance. A region will be deleted if it is separated and if the information it is associated with is no longer represent the normal condition of the environment.

3.4 Expansion

A region should be expanded if the difference between a measurement taken immediately outside of a region and the normal measurement associated with the region is less than the region tolerance, R_T.

3.5 Mergence

Two regions are merged by combining their width or length. Two regions could be merged if similar information is associated with them and if both are not overlapped with other regions.

3.6 Separation

A region is separated to update part of its area. Separation occurs if a new region coincides with an existing/old region or if a robot deviates from or returns to a point on its path where an old region exists. In order for separation to occur, first the elongation direction of the region is determined (usually the direction of travel of the robot). Then the intersection between the elongation line which is projected from the region reference point (x,y) and the separation line which is projected from the separation point is determined. The intersection is used as the point of separation. In this project, a region could be elongated in the x and y directions only. From the example

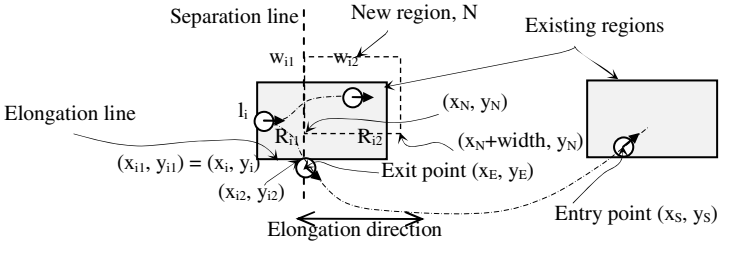


Fig. 3. Separation process

in Fig. 3, the separation point is the robot exit position (x_E, y_E) or the new region reference point (x_N, y_N). As a result, in the example the parameters of the resulting regions will then be (x_{i1}, y_{i1}) = (x_i, y_i) and (x_{i2}, y_{i2}) = (x_N OR x_E, y_i). $w_{i1} = |x_{i2} - x_{i1}|$ and $w_{i2} = w_i - w_{i1}$.

The following algorithm summarizes the separation process:

```
Separate(){
  Find_separation line,lS and all affected region()
  Find_region line,lR()
  Find_intersection between lS and lR()
  Determine_xywl of separated regions()
  Delete affected region()
}
```

4 Autonomous Update

The habituation mechanism is employed to quantify the measure of novelty of measurements associated with a region. A decaying function as given by (1) is used to model habituation where novelty measure, *n* is the measure of novelty of a region, *o* is the number of occurrences where the contents of a region matches the perception in the environment and τ is a constant that controls the rate of decay. The use of the maximum threshold, $h_{T\max}$ and minimum threshold, $h_{T\min}$ depicted in Fig. 4 allows the habituation process to exhibit hysteresis. The hysteresis introduces a delay when changing states from normal to novel and vice versa. When a new region is created, the novelty measure is set to be equal to h_{Tmax}. The region remains novel until the value decreases below h_{Tmin}. After that, the region will remain normal until its novelty measure rise above h_{Tmax}. The state of the region when its novelty measure is between h_{Tmin} and h_{Tmax} depends on the last threshold that it crossed. This effect is vital to avoid switching states (from novel to normal or vice versa) due to noisy inputs. A region will be deleted only if the novelty measure reaches 1.

$$n(o) = e^{\frac{-o}{\tau}} \tag{1}$$

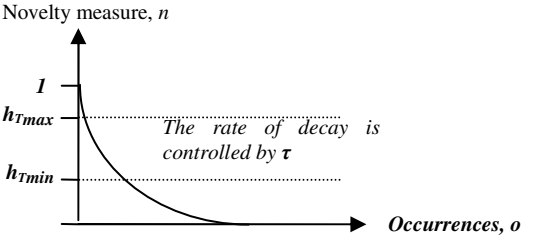


Fig. 4. Decaying novelty function with thresholds

One added modification for the Habituating Self Organizing Map (HSOM) is that new neurons should be created if the available neurons in the neural network cannot represent the changes in the environment. In practice, other more sophisticated

methods which can add or remove neurons when required such as the Grow When Required (GWR) network [16] could also be employed to replace the HSOM.

For autonomous mapping, the expansion process is restricted to newly created regions and is not allowed for established regions. The reason is that there will be a conflict in determining the value of the novelty measure if a region is already established and the extended area is new. Separation of an old region should only be done after a new region is fully expanded. For the mergence process, two neighboring regions with the same content could be merged only if both are habituated (where their novelty measures are below h_{Tmin}) and they are not coinciding with any other regions. The novelty measure of the merged region is the average of the novelty measures of the regions before they are merged.

Mapping involves two main process; restructuring and updating. Restructuring consists of changing the structure of the flexible region map to accommodate to changes in the environment. Updating is done once at the end of every epoch. This requires the robot to run through a segment (a pre-defined start/end position) of its environment before it can update. The pseudo code of the algorithm is given below:

```
Notation:
BMU – best matching neuron
S_BMU – distance between BMU and sensor measurement
S_T – similarity threshold

Initialization:
Train HSOM using perception during the first few runs
in a new environment
Set V_i = 0

Main(){
  Move_within_a_segment()
  While_in_segment()
    Get_pose()
    Get_sensor_measurement()
    Find BMU()
    If(S_BMU > S_T)
      newHSOMneuron = current perception
    EndIf
    Create()
    Expand()
    Separate()
    Mark_visited_region()
  EndWhile
}
Update_region_status(){
  For all regions
    If marked
      Increase its occurrences
    Else
      Decrease its occurrences
    EndIf
  EndFor
```

```
    Reset visited region marking
    If (n==1)
      Delete region
    Else if (n>=h_Tmax)
      State of region i = novel
    Else if (h_Tmin <n<h_Tmax)
      State of region i = current state
    Else
      State of region i = normal
    EndIf
  }
  Delete()
  Merge()
```

5 Experimental Results

The objectives of the experiments are to observe the results of region separation, to observe the value of novelty measure of new/separated regions and to test the performance of novelty detection when using the map. The experiments were conducted using a Pioneer 3 mobile robot in the L shaped room shown in Fig. 5. The robot was made to follow the wall on its left over a certain length of the room while taking laser measurements at every 100mm interval. This was repeated 5 times when the room was empty and then a further 10 times when an 100(w)x800(l)x600(h)mm^3 object was positioned at the center of the room. The region tolerance was set to a high value of 0.2 so that a small number of regions was created for easier visualization of the results. The novelty status threshold, h_{Tmax} was set to 0.9 and h_{Tmin} was set to 0.4.

Fig. 5. The mobile robot with its sensors (left) and the environmental setup (right). The dimensions of the L-shaped environment are given in Fig. 6 (in mm).

The results of the mapping are shown in Fig. 6. During the first five epochs, the environment was represented using region 1. When the object was introduced during epoch 6, new regions were created and region 1 was separated into several regions. The result of novelty detection during epoch 15 when the object was removed from the room are highlighted by using the using red dots to indicated the location of the

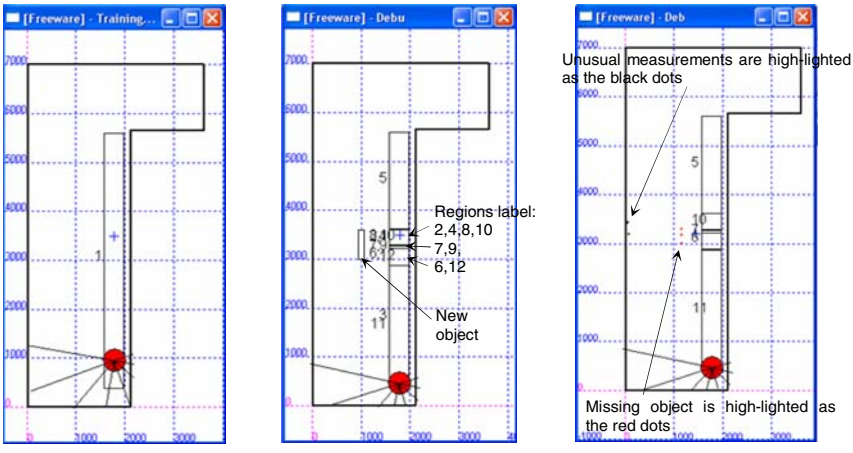


Fig. 6. An region (left frame) is restructured when a new object is introduced during epoch 6 (middle frame). The robot successfully highlighted the missing object (right frame) during epoch 15 by using the repetitive observation strategy [17].

missing object. This was done using a repetitive observation strategy [17] by comparing the normally seen laser measurements at several successive positions on the route with the current measurements.

The novelty measures of each of the regions over the 15 epochs are depicted in Fig. 7. We can see that during epoch 6 when the change was made to the room, some of the novelty measures are discontinued as they were deleted after being separated. Regions 2, 3 and 4 appear to have no novelty measure as they were separated (then deleted) into other regions during the same epoch. We can see that eventually after several epochs, some of the regions like regions 5, 6, 10 and 11 were habituated and established as normal whereas some like regions 8 and 12 were dis-habituated and deleted.

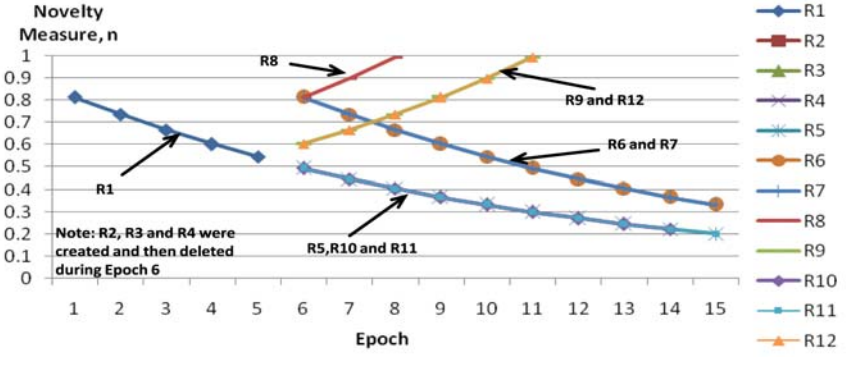


Fig. 7. Values of novelty measures over 15 learning epochs

6 Conclusion

An autonomous mapping method of normal data for novelty detection purposes using a mobile robot has been presented. The results show that the robot can autonomously adapt to its environment. One of the problem that arose from using a flexible region map was solved by allowing restructuring of the shape of the regions through creation, deletion, expansion, mergence and separation processes. The normality status of the perception in each region was successfully updated by the robot by using the habituation concept. This capability allows a robot to be continuously aware of any changes in its environment and then to adapt to the changes autonomously. The implementation of the mapping using a flexible region map also means that it could be used by mobile robots or standalone novelty filters which have data storage constrains due to their size, processing capability or limited power supply. One possible application for autonomous mapping is that the pattern of changes of a region's novelty measure could be use to study or filter dynamic changes at specific locations in the environment.

References

1. Markou, M., Singh, S.: Novelty detection: A review - Part 1: Statistical approaches. Signal Processing 83, 2481–2497 (2003)
2. Fox, D., Burgard, W., Thrun, S., Cremers, A.: A hybrid collision avoidance method for mobile robots. In: Proceedings of the IEEE International Conference on Robotics and Automation (1998)
3. Denham, M.J., McCabe, S.L.: Biological temporal sequence processing and its application in robot control. In: UKACC International Conference on Control 1996, vol. 25, pp. 1266–1271 (1996)
4. Marsland, S., Nehmzow, U., Saphiro, J.: Novelty detection for robot neotaxis. In: Proceedings of 2nd International ICSC Symposium on Neural Computation, Berlin, pp. 554–559 (2000)
5. Crook, P., Hayes, G.: A robot implementation of a biologically inspired method for novelty detection. In: Proc. Towards Intelligent Mobile Robot Conference, Manchester (2001)
6. Neto, H.V.: Real-time Automated Visual Inspection using Mobile Robots. Journal of Intelligent and Robotic Systems 49, 293–307 (2007)
7. Chakravarty, P., Zhang, A.M., Jarvis, R., Kleeman, L.: Anomaly Detection and Tracking for a Patrolling Robot. In: Australasian Conference on Robotics and Automation (ACRA), Brisbane, Australia (2007)
8. Brooks, R.: A robust layered control system for a mobile robot. IEEE Journal of Robotics and Automation 2, 14–23 (1986)
9. Miskon, M.F., Russell, R.A.: Mapping Normal Sensor Measurement Using Regions. In: IEEE International Conference of Industrial Technology, Melbourne, Australia (2009)
10. Elfes, A.: Using occupancy grids for mobile robot perception and navigation. Computer 22, 46–57 (1989)
11. Moravec, H.P.: Sensor fusion in certainty grids for mobile robots. AI Magazine 9, 61–74 (1988)
12. Samet, H.: An Overview of Quadtrees, Octrees, and Related Hierarchical Data Structures. NATO ASI Series, vol. F40 (1988)

13. Samet, H.: Neighbor Finding Techniques for Images Represented by Quadtrees. Computer Graphics and Image Processing 18, 37–57 (1982)
14. Marsland, S., Nehmzow, U., Saphiro, J.: A model of habituation applied to mobile robot. In: Proceedings of Toward Intelligent Mobile Robots, Bristol (1999)
15. Caldwell, D.R.: Unlocking the Mysteries of the Bounding Box. Coordinates. Online Journal of the Map and Geography Round Table Ser. A (2005)
16. Marsland, S., Nehmzow, U., Saphiro, J.: On-line novelty detection for autonomous mobile robots. Journal of Robotics and Autonomous Systems 51, 191–206 (2005)
17. Miskon, M.F., Russell, R.A.: A Repetitive Observation Strategy for Recognizing a True Anomaly and Estimating its Position. In: Austalasia Conference on Robotics and Automation, Canberra (2008)

A Fast Connected-Component Labeling Algorithm for Robot Vision Based on Prior Knowledge

Jun Liu, Guang Lu, Binbin Tao, Fang Chen, Haitao Gao, and Zhisheng Zhang

School of Mechanical Engineering, Southeast University
, o b

Abstract. Machine vision is now a major technique for intelligent robot system to sense the outside world. Connected-component labeling is a simple and e - cient way to help robot identify a specific region of interest (ROI). In this paper, the improvement of a two-scan algorithm based on prior knowledge is presented: (1) the rule of label assignment in the mask is improved and ROI orientation is introduced; (2) two cases of label equivalence in the mask were extracted to optimize the strategy of scanning; (3) two fast-connect ways were proposed to reduce the times of scanning. After the algorithm was implemented, parameters of each connected component are calculated to identify the ROI. In addition, this algorithm was also implemented in DSP platform on a service robot to identify a water cup. The experiment results demonstrated the e ciency of the algorithm is enhanced using the above strategies.

Keywords: Connected-component labeling, robot vision, ROI orientation, prior knowledge.

1 Introduction

Image segmentation is one of the most important parts of robot vision, it helps the robot to identify and locate the ROI. Labeling connected components in a binary image is one of the most fundamental operations in image segmentation and pattern recognition, which has so wide applications. A connected component in the binary image is consisted of connected pixels which may indicate the ROI and the labeling algorithm is to identify each connected component. When labeling is finished, each connected-component is assigned a unique label, and then the characteristic parameters of a component such as area, center position, perimeter and roundness can be calculated to identify the ROI. Generally, the component connectivity can be divided into 4-connectivity and 8-connectivity. (See Fig. 1). In this paper, only 8-connectivity is discussed.

Many algorithms have been proposed for addressing this issue, since many real-time applications such as tra c-jam detection, automated surveillance, and target tracking [1] are crucial to the performance of the labeling algorithm. In this paper, connected-component labeling is the major method for the service robot to identify ROI because of its easy implementation and high performance. Characteristic parameters of a component such as area, aspect ratio can be calculated based on the result of labeling. The real-time performance is important to the robot because: (1).it needs to identify the ROI and calculate its position repeatedly while moving. (2). When the robot tries to grasp

Corresponding author.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 856–865, 2009.

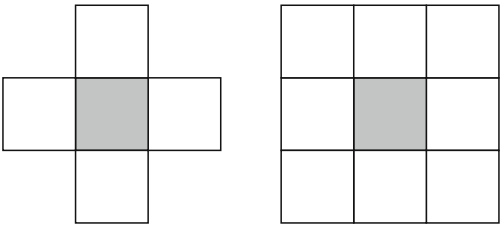


Fig. 1. Two definitions of connectivity in binary image. (a) 4-connectivity, (b) 8-connectivity.

the object, the real-time position of the robot arm and the cup is estimated constantly as a feedback to control the robot arm. For all the labeling algorithms, they can be divided into four main classes: (1) Multi-scan algorithms. In algorithms [2, 3], an image is scanned in forward and backward raster directions with two different masks alternately to resolve label equivalences. (2) Two-scan algorithms. In algorithms [1, 4], labeling is completed in two scans: in the first scan, provisional labels are assigned to object pixels, label equivalences are also recorded, and label equivalences operation is proceeded. In the second scan, label equivalences operation is completely finished which means each connected component will have a same label. (3) Algorithm based on run-length coding [5]: this algorithm introduces run-length encoding into DSP real-time image processing to reduce memory occupation and the quantity of objects. It can memorize all label equivalences in one scanning by changing the assignment of label connection table. (4) Algorithm based on contour-tracing [6]: resolving of label equivalences is avoided in this algorithm by tracing the contours of object pixels.

This paper presents an improved two-scan algorithm for labeling connected components in binary images based on prior knowledge extracted from the procedure of label equivalence and the scanning. For resolving label equivalence, the method of using equivalent label sets and the representative label table [4] is used in our algorithm because of its brevity and e ciency. With experiment carried out, we demonstrated the accuracy and e ciency of our algorithm.

2 The Improvements of Two-Scan Algorithm

2.1 Outline of Conventional Two-Scan Algorithms

Generally, two-scan algorithms complete the labeling of binary image by scanning an image in the raster scan direction twice and the following mask is typically used [7]. See Fig. 2.

The first scan is described as follows [2]:

if $(b(x\ y)\quad V_{\mathrm{B}})$ no operation

else if $(b_{min}(x\ y)\quad V_{\mathrm{B}})$ $(b(x\ y)\quad m\ m\quad m\quad 1)$

else $b(x\ y)\quad b_{min}(x\ y)$

$b_{min}(x\ y)\ \ \min[b(i\ \ j)\quad b(i\ \ j)\ \ M_{\mathrm{S}}]$,

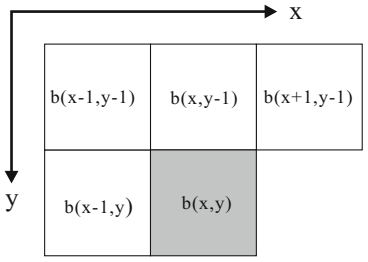


Fig. 2. Mask for eight-connected connectivity

where b(x,y) denotes the pixel value at (x, y) in the image, V_O for the pixel value for object pixels and V_B for that for the background pixels. m (which is initialized to 1) is a provisional label, min( ) is an operator for calculating the minimum value, and M_S is the region of the mask.

Equivalent label sets and a representative label table used for in two-scan labeling algorithm [4] have been demonstrated to be simple and e cient. So it is also used in our algorithm which will be introduced later. The Label equivalence is completely resolved in two scans, and the detail is described in [4].

2.2 Improvement of Label Assignment in the Mask

In the algorithm proposed by L.F He[4], during every scanning in the mask, it needs to compare the labels to obtain the minimum label of object pixel in the mask, which is time consuming. As all labels in an equivalent label set are equivalent, we can just assign the label of first-scanned object pixel and record the relation of equivalence without calculation of b_{min}(x,y). In addition, the label assignment of the current pixel is finished once the first object pixel is found, which help reduce the times of scanning in the mask. At last, a drawback that lost the total number of different components will be overcome by one time sorting which is very simple and fast.

The sequence of scanning in the mask ($b(x\ \ 1\ y\ \ 1)\quad b(x\ y\ \ 1)\quad b(x\ \ 1\ y\ \ 1)\quad b(x\ \ 1\ y)$) (See Fig.3a) is used in almost every two-scan algorithm, but there is no explanation why this sequence is used, maybe it is just inertia of mind because raster scan direction is applied. Actually, the prior knowledge can be extracted to determine the scanning sequence for a type of images. For example, the ROI in a type of images usually consists of horizons lines or vertical lines, then scanning sequence of ($b(x\ 1\ y)\quad b(x\ \ 1\ y\ \ 1)\quad b(x\ y\ \ 1)\quad b(x\ \ 1\ y\ \ 1)$) and ($b(x\ y\ \ 1)\quad b(x\ \ 1\ y\ \ 1)\quad b(x\ \ 1\ y\ \ 1)\quad b(x\ \ 1\ y)$) can affect the labeling performance greatly. The two scanning sequences are respectively optimal because the first check in the mask is most probably effective because it will find an object pixel in one time of check. Similarly, if the ROI is prone to 45-degree or 135-degree, the optimal scanning sequence should start from $b(x\ \ 1\ y\ \ 1)$ or $b(x\ \ 1\ y\ \ 1)$. The tendency of the ROI orientation could be obtained through statistical method. The experiment results are showed later. In our algorithm, the scanning sequence ($b(x\ \ 1\ y)\quad b(x\ \ 1\ y\ \ 1)\quad b(x\ y\ \ 1)\quad b(x\ \ 1\ y\ \ 1)$) (See Fig.3b) is used.

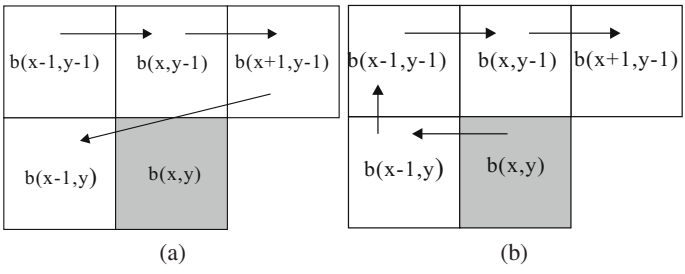


Fig. 3. Conventional scanning sequence and ours in the mask

2.3 Two Cases of Label Equivalence in the Mask

In two-raster-scan algorithm, the label equivalence is realized through the current pixel which connects pixels with different labels in the mask. A detailed analysis is carried out to extract the possible regular patterns in the operation of label equivalence, which will help make pointed operation of equivalence to reduce the times of scanning and improve the e ciency. In the four-pixel mask presented before, any three or four different pixels are not possible to have three or four different labels. So, the connection will happen only when two pixels with different labels in the mask is connected through the current object pixel. There are six different cases of the current object pixel with two object pixels in the mask. See Fig.4.

In the six cases, we can infer that only two cases which may need resolving of label equivalence in the mask. In case (b) and (f), the two object pixels in the mask are not directly connected, thus they may have different labels. We can check the labels and the representative labels of the two pixels to determine whether the resolving of label equivalence is need. To do this, the following operations are performed (illustrated as case (b)).

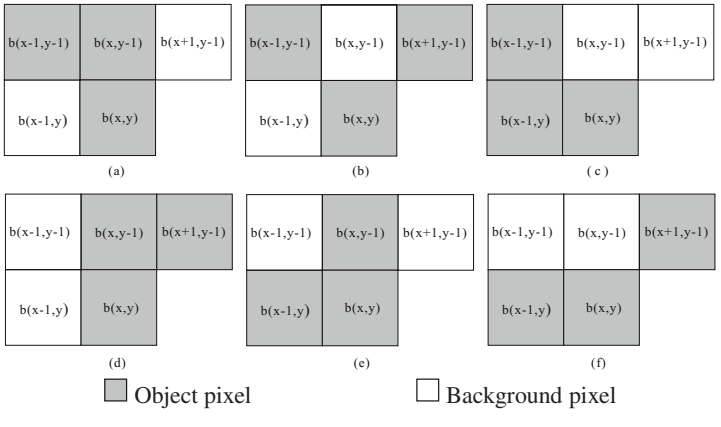


Fig. 4. Six cases of the current object pixel with two object pixels in the mask

if ($b(x \quad 1\ y \quad 1) \quad b(x \quad 1\ y \quad 1) \quad V_O$ &
$l(x \quad 1\ y \quad 1) \quad l(x \quad 1\ y \quad 1)$ & $r(x \quad 1\ y \quad 1) \quad r(x \quad 1\ y \quad 1)$)
resolve $l(x \quad 1\ y \quad 1)\ l(x \quad 1\ y \quad 1)$
else *scan continued*

In other four cases, the two object pixels in the mask either already have the same label or label equivalence has been set up. So it doesn't need to do label equivalence in these cases. The check strategy will be:

(1) if ($b(x \quad 1\ y) \quad V_O$) check $b(x \quad 1\ y \quad 1)$,
(2) if ($b(x \quad 1\ y \quad 1) \quad V_O$) check $b(x \quad 1\ y \quad 1)$,

In (1), it can avoid the check of $b(x \quad 1\ y \quad 1)$ and $b(x\ y \quad 1)$. In (2), the check of $b(x\ y \quad 1)$ can be avoided. So the scanning strategy is optimized using the prior knowledge of the two types of connection.

2.4 Two Ways of Fast Scanning

Two ways of fast scanning is proposed after further analysis of the two-scan algorithm. One of the reasons that the e ciency of the algorithm is limited is repeated checking of pixel values in the scanning. See Fig.5.

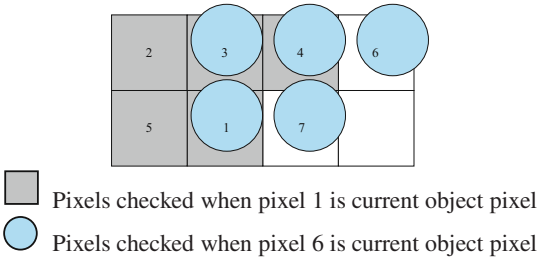


Fig. 5. Two successive series of scanning

In the scanning sequence of pixel 2 3 4 5, at most three times of repeated checking (pixel 1, 3, 4) will proceeded when pixel 1 and pixel 6 are current object pixels.

The two ways of fast scanning is applicable when the first object pixel in the mask is pixel 3 and pixel 4. See Fig.6.(The current object pixel is pixel 1).

When the first object pixel in the mask is pixel 3, the label of current object pixel 1 is copied from pixel 3. Because it doesn't need to resolve equivalent sets in this situation

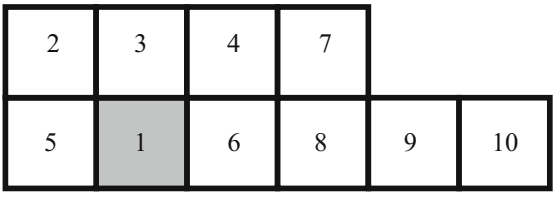


Fig. 6. Two ways of fast scanning

according to the two cases of label equivalence stated above. This cycle of scanning seems finished, it should turn to the next current pixel 6, but we can make this kind of scanning faster by performing the following:

when ($p1$ $p3$ V_O)

l1 l3; check p6
if ($p6$ V_O)
$l6$ $l3$; check p8
if ($p8$ V_O) $l8$ $l3$; original scanning begin from p9
else original scanning begin from p8

, Pn denotes value of pixel n, ln denotes value of label of pixel n.

In this situation, the scanning cycle that the pixel 6 as a current pixel can be omitted and pixel 8 as a current pixel may be omitted.

When the first object pixel in the mask is pixel 4, the following scanning strategy is performed:

when ($p1$ $p4$ V_O)

l1 l4; check p6
if ($p6$ V_O)
$l6$ $l4$; check p8
if ($p8$ V_O) $l8$ $l4$; check p9
if ($p9$ V_O) $l9$ $l4$
else original scanning begin from p10
else original scanning begin from p9
else check p8
if ($p8$ V_O) $l8$ $l4$; check p9
if ($p9$ V_O) $l9$ $l4$
else original scanning begin from p10
else original scanning begin from p9

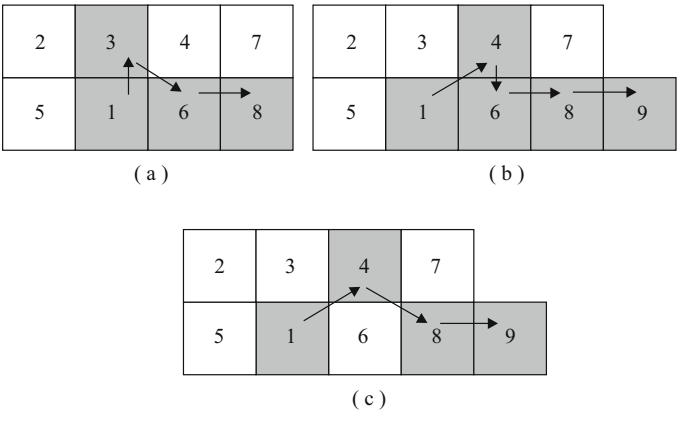


Fig. 7. Fast scanning strategy

These two ways of fast scanning is typical of 8-connectivity. It actually scans the image in the following sequence in Fig.7.

3 Experiment and Results

3.1 Introduction of the Robot Vision System

The algorithm presented above is applied in the vision system of an ”Accompany & Nursing” robot, which aims to help the elderly and the disabled perform tasks such as fetch a medicine bottle or a water cup. The architecture of visual guide system is showed in Fig.8.

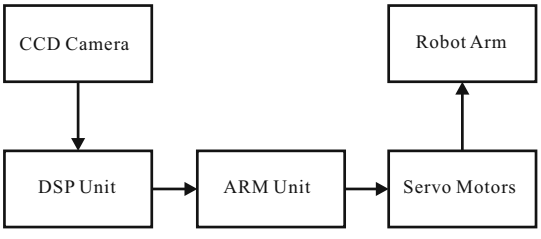


Fig. 8. Architecture of visual guide system

The robot arm is actually a 4-DOF manipulator equipped with four servo motors at the joints. The images are captured by a CCD camera. After image processing and analysis are completed on DSP platform, the information of ROI’s position is delivered to an ARM CPU in which the forward kinematics is calculated, then the servo motors at the joints are controlled by the PWM signal sent from the ARM CPU to perform specific tasks.

3.2 Performance Evaluation and Comparison

The presented algorithm is experimented in C language on a desktop PC (Intel Core 2 Duo E6300@2.10GHz), and the final version is implemented in DSP (Ti TMS32 0DM642 600MHz).

Ten images of Julia-sets (See Fig.9.) from a website [8] were tested. Each of the ten images is actually composed of one or several complex connected components. The resolutions of the ten images are all modified to 800 800 for uniformity and convenience because the resolutions of original images are a little different. The experiment results are presented by listing the execution time for 5000 runs (L. He’s improved algorithm [1] is compared) in Fig.10.

From the results, we can conclude our algorithm is faster than L. He’s. With time limited, we have not tested all other algorithms, the details could be found in [1, 4].

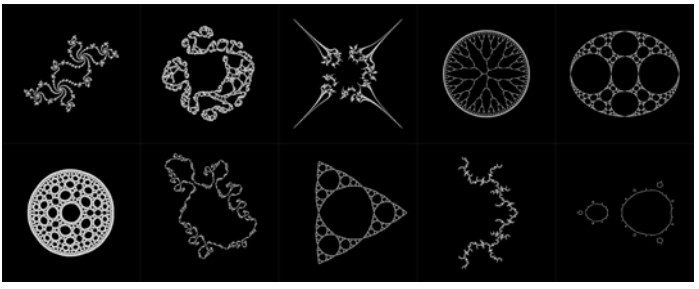


Fig. 9. Ten binary images Julia-sets for test

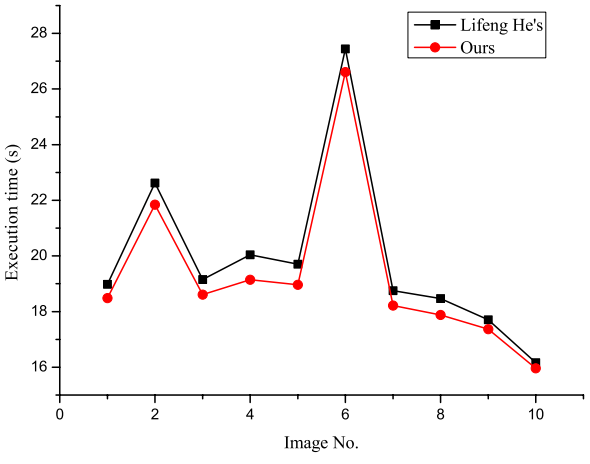


Fig. 10. Comparison of L.F. He's algorithm with ours

3.3 The Experiment about Tendency of ROI Orientation

The information of ROI orientation is extracted from the statistical analysis which is performed by respectively calculating the quantity of the object pixels in the four directions (see Fig. 11) in the mask.

Orientation 1 and orientation 3 is a pair to determine the diagonal dominance; Orientation 2 and orientation 4 is a pair to determine the horizontal (or vertical) dominance.

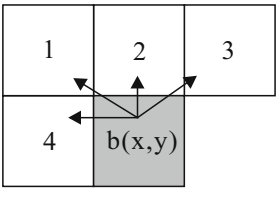


Fig. 11. Four directions in the mask

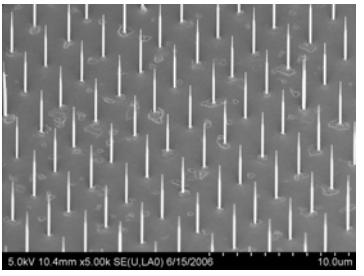
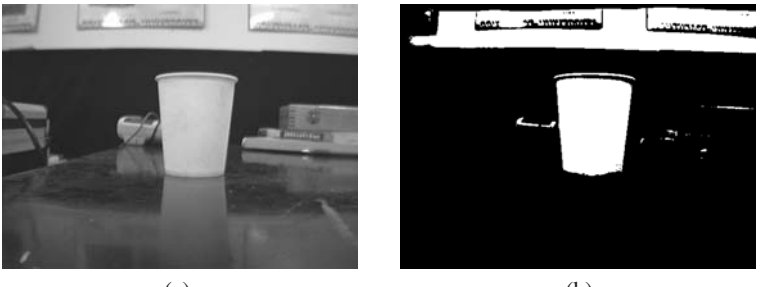


Fig. 12. A typical picture of Nano-Wires

Table 1. The statistical results of ROI orientation

Image	Orientation 1	Orientation 2	Orientation 3	Orientation 4
1	41299	50208	41377	41544



(a) (b)

Fig. 13. Grayscale and binary image with ROI

The two pairs are independently analyzed to determine the dominance in orientation, while the shape of ROI can be inferred when they are combined.

Counting of Nano-Wires is good example of utilizing the information of ROI orientation. A typical picture (See Fig.12) of Nano-Wires [9] was tested and the statistical results are showed in Table 1.

In the image, the orientation 2 is dominant while the other three orientations are almost the same. In this situation, the conclusion that the ROI is mainly in vertical orientation can be made, so the first-check pixel in the mask should be pixel $b(x\ y$ 1). Experiment results with and without the prior knowledge are31.438s and 34.249s, which demonstrate the effectiveness of ROI orientation method.

Fig.13(a) shows a scene with ROI (a water cup) in it, and Fig.13(b) is the binary image after thresholding. Labeling of connected components is completed using our algorithm on DSP platform. We calculate the area, aspect ratio and center of each component to help identify and locate the ROI. Under the criteria that (area 10000)

& (0.65 aspect ratio 0.85), the water cup is identified successfully and the center is (307.27, 191.32) which is delivered to control system.

4 Conclusion

In this paper, a fast two-scan labeling algorithm of connected components was presented. Essentially, all techniques to improve the e ciency of our algorithm are aiming at reduce repeated times of scanning (checking). Prior knowledge extracted from detailed analysis of the conventional two-scan algorithm is crucial to the improvement. The ROI orientation could be a type of prior knowledge which can be used if the images captured have roughly a same orientation. The experimental results demonstrated that our algorithm is superior to conventional two-scan algorithm.

For future work, a recursive scanning strategy is being conceived, which is to further reduce the repeated times scanning (checking) in row direction. Because in many applications, the connected-component is solid (without a hole in it) which means most pixels in a component is surrounded by other eight object pixels. A recursive scanning strategy is able to scan an object row without the mask.

Acknowledgments. This work was supported by National High Technology Research and Development Program of China (No. 2006AA040202), National Nature Science Foundation of China (No. 50805023), and Open Fund Program of the Key State Laboratory (2008-DMET-KF-014).

References

[1] He, L., Chao, Y., Suzuki, K., Wu, K.: Fast connected-component labeling. Pattern Recognition 42, 1977–1987 (2009)
[2] Haralick, R.M.: Some Neighborhood Operations. Real Time/Parallel Computing Image Analysis, 11–35 (1981)
[3] Hashizume, A., Suzuki, K., Yokouchi, H., et al.: An Algorithm of Automated RBC Classification Anditsevaluation. BioMed. Eng. 28(1), 25–32 (1990)
[4] Lifeng, H., Yuyan, C., Suzuki, K.: A Linear-Time Two-Scan Labeling Algorithm. In: IEEE International Conference on Image Processing, 2007. ICIP 2007, pp. 241–244. IEEE Press, San Antonio (2007)
[5] Shijie, C., Qiang, Y.: Optimization and Application of Connected Component Labeling Algorithm Based on Run-Length Encoding (in Chinese). Computer Applications 28, 3150–3153 (2008)
[6] Chang, F., Chen, C.J., Lu, C.J.: A Linear Time Component-Labeling Algorithm Using Contour Tracing Technique. Computer Vision and Image Understanding 93, 206–220 (2004)
[7] Rosenfeld, A., Kak, A.C.: Digital Picture Processing. Morgan Kaufmann, San Francisco (1982)
[8] tt www jo t j o j t t
[9] tt www L t q t L o o

Simultaneous Visual Object Recognition and Position Estimation Using SIFT

Rigas Kouskouridas, Efthimios Badekas, and Antonios Gasteratos

Democritus University of Thrace,
Department of Production and Management Engineering,
Vas. Sofias 12, 67100 Xanthi, Greece
{rkouskou,agaster}@pme.duth.gr, badekas@anadelta.com

Abstract. In the last decade, pattern recognition tasks have flourished and become one of the most popular tasks in computer vision. A wealth of research focused on building vision systems capable of recognizing objects in cluttered environments. Moreover industries address all their efforts to developing new frameworks for assisting people in everyday life. The need of robots working closely to human beings in domestic workplaces, makes a necessity the usage of intelligent sensorial systems that are able to find patterns and provide their location in the working space. In this paper a novel method able to recognize objects in a scene and provide their spatial information is presented. Furthermore, we investigate how SIFT could expand for the purposes of location assignment of an object in a scene.

Keywords: Object Recognition, Position Estimation, SIFT, depth estimation, robot vision.

1 Introduction

Recognizing objects in a scene is one of the oldest tasks in computer vision field and still remains one of the most challenging. Every pattern recognition technique is directly related with the decryption of information contained in the natural environment. During the last decade, remarkable efforts were made to build new vision systems capable of recognizing objects in cluttered environments. Moreover, emphasis was given to recognition systems based on appearance features with local estate [1], [2]. Insensitivity against rotation, illumination and viewpoint changes, constitute the attributes of algorithms extracting features with local interest. Nowadays, vision systems are equipped with such attributes and, as a result, more and more techniques are included in industrial products.

In the last few years, a tendency to introduce autonomous robots into domestic environments is discerned. Industries address all their efforts to developing machines capable of assisting people to everyday life. To this end, a compulsory requirement is that robots must avoid obstacles in both static and dynamic environments. As a result, the sense of depth constitutes the essential attribute of such frameworks. In turn, over the past decade, significant research efforts were

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 866–875, 2009.

devoted to the development of vision systems, capable of providing a sense of location and direction to robots. In this field, techniques for visual Simultaneous Localization and Mapping (vSLAM) [3] [4], are included. Particularly, SLAM methods that are utilized in autonomous robotics, concentrate at building up maps, in unknown environments, whilst keeping track of their own position.

Furthermore, from time to time, researchers emphasized in introducing computer vision techniques into demanding robotics applications. As a result, challenging automatic manipulation tasks can be adequately accomplished by utilizing object recognition techniques based on local appearance. The most favored method of this field is the Scale Invariant Feature Transform (SIFT) that was presented in [5]. Furthermore, SIFT is adopted in modern robotics applications due to the fact that it performs exceptional repeatability and invariance against possible illumination, scale, rotation and viewpoint changes. For instance in [6], a remote-interactive mode for a museum guide robot is presented, where SIFT is used during the object recognition process. In turn, mobile robots' navigation using landmarks can be adequately fulfilled by using SIFT features as it was shown in [7], where each landmark is initially located in the image coordinates by using SIFT features for the recognition and the RANSAC(Random Sample Consensus) for the matching procedure.

In this paper, we describe a novel method to develop a simultaneous visual object recognition and position estimation system at the same time. We investigate how, a very efficient object recognition scheme, can be expanded for the needs of position estimation. Specifically, SIFT was selected among a set of high-level algorithms to describe patterns and objects. We prove that information derived from SIFT, allows the estimation of the distance between camera and objects found in a scene and we describe the respective algorithm. The remainder of this paper is structured as follows: In Section 2, a short introduction to object recognition based on local features is presented and a detailed description of the methods involved in building our recognition and position estimation algorithm, is given. In Section 2.1 we present the parts of SIFT that are used in the later stages of the algorithm. The proposed object recognition and position estimation framework is analytically presented in Section 3, where details about training and implementation of our system are apposed. In Section 4, we experimentally evaluate the proposed method. The work concludes with some final notes and an outlook to future work in Section 5.

2 Local Appearance-Based Recognition

During the past decade, several techniques that enforce the essential role of local features in object recognition tasks [8] were presented. The special visual distinctiveness of an object in a scene is ensured by locally sampled descriptions. The vital issue underlying object recognition based on local features is maintaining this distinctive regional-based information. Detectors and descriptors of areas of interest constitute the sub-mechanisms in every local-based recognition approach. In addition, they provide special attributes such as, insensitivity

against rotation, illumination and viewpoint changes. In [9], local detectors and descriptors are evaluated for object recognition purposes.

The main idea behind interest location detectors is the pursuit of points or regions with unique information in a scene. These spots or areas contain data that distinguish them from others in their local neighborhood. Needless to say that, detector's efficiency relies on its ability to locate, as many distinguishable areas as possible, in an iterative process. One of the most efficient detectors, the Maximal Stable Extremal Regions Detector, was proposed in [10]. In short, regions darker or brighter than their surroundings are detected. The efficiency of the algorithm relies on the relationship between pixels' intensity value and local neighborhood.

Generally speaking, a descriptor organizes the information collected from the detector in a discriminating manner. Thus, locally sampled feature descriptions are transformed into high dimensional feature vectors. In other words, parts of an object located in a scene are represented by descriptors. Putting these descriptors in logical coherence fulfills the final object representation. In all the recent proposed methods, databases containing descriptors from multiple objects are constructed. These databases, that play essential role in all the recently proposed object recognition techniques, are structured in a vocabulary-tree format. Furthermore, vocabulary trees, which are tree-like data structures based on k-means clustering, were proposed in [1].

Speeded Up Robust Features (SURF) that was introduced in [11], implements both a detector and a descriptor. The first is constructed by using a so-called Fast Hessian Matrix that, is based on an approximation of the Hessian Matrix for a given image point. Afterwards, rectangular -9x9-pixels filters are used for the approximation of the second derivative of the Gauss function. The descriptor is produced based on the responses of all the interest points extracted from the detector. Currently, the most widely adopted approach that produces efficient detector and descriptor, is SIFT [5]. As it is mentioned before, the proposed method is based on this approach. For this reason an extended description of SIFT follows.

2.1 Description of SIFT

Initially, the image is convolved with the variable-scale Gaussian for the production of a scale-space image. Afterwards, stable keypoint locations are detected by using scale-space extrema in the difference-of-Gaussian (DoG) function convolved with the image. In [12] it was shown that, DoG function provides a close approximation to the scale-normalized Laplacian of Gaussian. Thus, the scale invariance of the detector is ensured. SIFT's descriptor is produced by using stacked gradient histograms over 4x4 sample regions. Firstly, the gradient magnitude and orientation at each point in a region around the keypoint location are computed. Afterwards, these samples are gathered into orientation histograms collecting the contents over 4x4 sub-regions. Since 8 orientation bins are used, the descriptor, finally, constitutes of 128-element feature vector. In the most the cases, matching between descriptors relies on comparing them one by one. The

matching process in SIFT involves the organization of descriptors from trained images into a vocabulary kd-tree. Moreover, with this way the approximate nearest neighbors to the descriptors are found. SIFT's detector and descriptor are insensitive to possible image scale, rotation, change in 3D viewpoint, addition of noise and change in illumination. These exceptional attributes justify the fact that SIFT is adopted from, almost any, new object recognition approach [13] [14]. To sum up, SIFT's unique properties can be used not only for recognition but also for position estimation of an object in a scene. We are examine how SIFT could be altered for the purposes of such a challenging task.

3 Algorithm Description

Initially, we present fragmentarily the proposed method by showing its main stages. The main idea behind the proposed method is to maintain SIFT's properties whilst exploiting them in order to particularly estimate objects' distance from the camera. Thus, we have constructed a large database containing images from several objects. With a view to database's enrichment, these objects were photographed from different viewpoints and distances from the camera. Moreover, we used SIFT's matching sub-procedure to build an on-line scene search engine. Estimations derived from this engine are taken into account for the position estimation task. The main stages of the proposed algorithm are as follows:

Stage I *Apply SIFT to the scene's and object's image, in order to estimate the features position in each of them.*

Stage II *Obtain the N features that match in the two images by applying the matching sub-procedure of SIFT. Define as* $(X_{S_i}, Y_{S_i}), i = 1, \ldots, N$ *the positions of the N features in the scene image and* $(X_{0_i}, Y_{0_i}), i = 1, \ldots, N$ *the positions of the N features in the object image.*

Stage III *Define as* (X_{S_c}, Y_{S_c}) *and* (X_{0_c}, Y_{0_c}) *the features' centers of mass for both images. This is accomplished by estimating the mean values of the features positions in the two images:*

$$X_{S_c} = \frac{1}{N}\sum_{i=1}^{N} X_{S_i} \quad \textit{and} \quad Y_{S_c} = \frac{1}{N}\sum_{i=1}^{N} Y_{S_i}$$

$$X_{0_c} = \frac{1}{N}\sum_{i=1}^{N} X_{0_i} \quad \textit{and} \quad Y_{0_c} = \frac{1}{N}\sum_{i=1}^{N} Y_{0_i}$$

Stage IV *Calculate the mean Euclidian distance (in pixels) of each feature from the corresponding center of mass that is extracted in the previous stage. Set as* E_S *and* E_0 *the*

mean Euclidian distances in the scene and object image, respectively. The following relations are used:

$$E_S = \frac{1}{N}\sum_{i=1}^{N}\sqrt{(X_{S_i} - X_{S_c})^2 + (Y_{S_i} - Y_{S_c})^2}$$

$$E_0 = \frac{1}{N}\sum_{i=1}^{N}\sqrt{(X_{0_i} - X_{0_c})^2 + (Y_{0_i} - Y_{0_c})^2}$$

Stage V *Estimate d_S which corresponds to the ratio of the two mean distances E_S and E_O. Furthermore, we introduce the pre-computed depth d_O, which is obtained, during the training session and while the object is captured alone.*

$$d_S = \frac{E_0}{E_S}$$

Stage 1 could be apprehended as the training session of our algorithm. In this phase, for each image in the database keypoint features are extracted using SIFT. Each object is photographed at different distances from the camera and the pre-computed depth d_O is stored for further exploitation. This process is performed while the system is offline, thus, executable time is not taken into account. The results are stored for further use at the next phases. In Stage 2, the matching sub-procedure of SIFT is performed. Especially, descriptors that are common in both images (scene and object) are extracted. It is apparent that, one image representing the scene is compared with several others, representing the object from different viewpoints. Furthermore, the locations of the common features are stored for further exploitation.

In Stage 3, the position estimation sub-procedure takes place till the end of the algorithm. Moreover, at this phase, the features' centers of mass in both images are calculated. The last is obtained by estimating the mean values of features locations in both representations. In Stage 4, the distance of each keypoint from the center of mass is calculated. This is measured in pixels with the use of Euclidian Distance. By the end of this sub-routine, we are able to collect significant spatial information of an object in a scene. This is accomplished by simply estimating the distribution of trained features around their center of mass. Finally, in Stage 5, the object's distance from the camera is computed. The pre-computed depth d_O measured during the training session (Stage 1), is taken into account. The ratio d_S is used to measure the proportion of object's features to those found in the scene.

After the necessary training session and the database construction at the initial stages of the method, an on-line search engine follows. This is responsible for querying in the scene for objects contained in the trained database. When an object is found, the scene's image is compared to this object, which provides the majority of common matches. Finally, features' information from both images is interpolated with a view to object's position allocation.

4 Experimental Results

In this section, we assess the properties of the proposed method in detail. The tests were executed on a typical PC with a core2duo 2.2 GHz processor, 2 GB RAM and Windows XP operating system. Furthermore, the camera used (Grasshopper by Point Grey Research) is able to capture images up to 1280x960 pixels resolution and is connected to the PC via a firewire port. The data transmission is accomplished by using IEEE 1394b transfer protocol. The training of the object recognition system is done in MATLAB. The last is preferred from other programming tools, due to the fact that, it offers users-friendly environment and convenient image processing functions.

The proposed method is evaluated through exhaustive tests containing several scenes and objects. In Figure 1(a) we present scene A, which contains three different objects (e.g. a book, a modem's box and a motherboard's box). With a view to reader's better understanding, objects found in the scene are referred as book, modem and box, respectively.

(a) (b)

Fig. 1. (a) Upper: Scene A containing three objects Lower: Each object is captured from different viewpoint. (b) The extracted SIFT features for scene A.

In this experiment, the scene was captured at 100 cm distance from the camera. Objects were captured separately from different viewpoints under varying illumination and geometrical conditions. This procedure, being essential for the system's training was done while the system was offline. Objects are stored as images into the database for further exploitation. According to the first stage of our algorithm the SIFT features of all images in the database should be extracted. In Table 1(a), the results of this phase for the scene A are presented. Apparently, more features are extracted from the most textured objects. The ones containing high amount of local extend favor during the SIFT feature extraction procedure.

In Figure 1(b) the extracted SIFT features from the scene A and the contained objects are depicted. The second stage of our algorithm involves the matching sub-procedure of SIFT, where the features of the scene under investigation and

Table 1. a) SIFT features for scene A and its contents. b) The results of the position estimation procedure for scene A.

(a)

Image	SIFT features
Scene A	208
Book	150
Box	349
Modem	108

(b)

Object	Matches	Z	d_0	d_S	Z^*
Book	18	100	70	1.4992	104.93
Modem	20	120	70	1.6098	112.68
Box	44	140	70	2.005	140.65

of each separate object are compared to obtain the N matches. The scene is collated with all possible objects' views. The total amount of these comparisons depends on the quantity of different views of the same object. During the database construction we altered objects' viewpoint and distance from the camera. It is apparent that, proposed method's efficiency is directly related to the size of the constructed database. A large database maximizes the possibility to perform adequately recognition and more accurate position estimation tasks. The results of these comparisons are shown in the second column of Table 1(b). As it was expected, the object "box" provided more matches. This is due to the fact that it outperformed during the feature extraction process. The most essential issue that one should keep in mind is that, there is a direct relationship between the amount of SIFT features extracted and the final amount of matches. The more the SIFT features are the more the object's matches with the scene and as a result, the higher the possibility the object to be found in the scene.

The next step includes the spatial data estimation. In the third column of Table 1(b) distances in cm for every object in scene A are shown. By applying the last stages of our method we are able to estimate d_S. In particular, we use the pre-computed depth d_0 which is object's distance from the camera when the later is captured alone during the training session. Pre-computed depth d_0 is definitely different from object's distance from the camera (Z) as it is shown in Table 1(b). In order to confirm proposed method's accuracy we needed a ground truth. For this purpose when the scene is captured, we measure and store objects' distances from the camera (Z) with a laser distance measuring device. The goal of the proposed method is to approximate the later distances with as high accuracy as possible. In particular, by executing the final stages of the algorithm we are able to estimate objects' distance (Z^*) from the camera. In the last column of Table 1(b) the results of the position estimation process are illustrated. Figure 2(a) depicts the results of the proposed method. More specifically, in white boxes the number of N matches between the scene and the object is shown. Moreover, in the same figure, the mean Euclidean distances E_S and E_0 extracted in Stage IV, are illustrated. In addition, ratio d_S for every object in a scene is extracted. The above are illustrated as a modified MATLAB figure title that is constructed through the proposed algorithm.

In order to assess proposed method's robustness under 3D rotation and illumination changes we took measurements on the scene shown in Figure 2(b). The

(a) (b)

Fig. 2. (a) Proposed method's results for every object in scene A. (b) Scene B containing rotated objects.

new scene contains the same objects as the previous one but in different alignment. Especially, objects were rotated whilst keeping the same distance (Z) from the camera and altering the illumination conditions. In Table 2(a), the results of the first stage of the proposed method for scene B and its contents are presented. In addition, in Table 2(b) the results of the proposed position estimation process are illustrated. Although scene B differs significantly from A, the efficiency of the proposed object recognition and position estimation method, remains high. In addition to the above results, we introduce an efficiency ratio or an accuracy percentage with a view to further analysis. Thus, we estimate:

$$\alpha = (1 - \frac{\| Z - Z^* \|}{Z}) \times 100.$$

Table 2. a) SIFT features for scene B and its contents. b) The results of the position estimation procedure for scene B.

(a)

Image	SIFT features
Scene B	285
Book	150
Box	349
Modem	108

(b)

Object	Matches	Z	d_0	d_S	Z^*
Book	26	100	70	1.410	98.756
Modem	22	120	70	1.698	118.86
Box	42	140	70	1.957	137.00

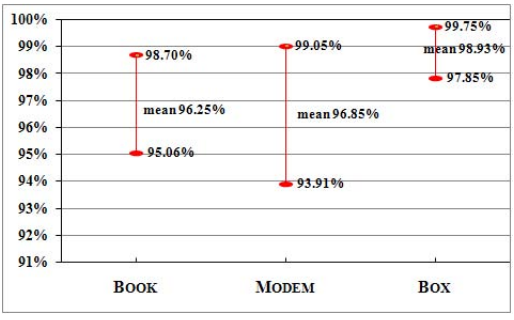


Fig. 3. Accuracy of the proposed algorithm

The proposed algorithm was evaluated through 30 tests and the results acquired are illustrated in Figure 3. In these experiments 3D rotation (up to 30 degrees) and illumination changes were applied. In Figure 3 minimum, maximum and mean accuracy values are respectively shown. Moreover, the most interesting issue that derives is the fact that the extracted accuracy never drops below 93%. Furthermore, as expected larger objects are found with a higher accuracy.

5 Conclusions

A technique for object recognition based on SIFT has been presented in this paper. The novelty here is that we have enriched it to perform depth estimation and, consequently, object localization in an arbitrary scene. The key idea is that the tracked SIFT features are located on given geometric positions, thus they can be considered as the corners of a polyhedron, the center of gravity of which is computed and it is associated to the actual center of mass of the sought object. Once the features' center of mass is known and the object is recognized, the distance of the object from the camera is trivial, given at least one recorded position of the object. The application of such a method in robotic manipulation tasks is apparent, as it permits the identification of the object to be handled, as well as its distance from the robot's eye. Experimental results, where different illumination conditions and objects' viewpoints were sampled, prove that larger objects are easier to be found, whilst, their position is estimated with higher accuracy. To sum up, with an outlook to future work, we will enrich and modify the algorithm in order to perform challenging pose estimation tasks. The overall purpose of the new framework will be the extraction of objects' roll, pitch yaw angles and T_X, T_Y and T_Z translation matrices. Thus, several tasks such as automatic manipulation of objects or assisting autonomous vehicles avoiding obstacles would be completed at high execution times with significant efficiency.

Acknowledgements. This work is supported by the E.C. under the FP6 research project for Autonomous Collaborative Robots to Swing and Work in Everyday EnviRonment ACROBOTER, FP6-IST-2006-045530.
http://www.acroboter-project.org

References

1. Nister, D., Stewenius, H.: Scalable recognition with a vocabulary tree. In: CVPR 2006: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, pp. 2161–2168. IEEE Computer Society, Los Alamitos (2006)
2. Sivic, J., Zisserman, A.: Video google: A text retrieval approach to object matching in videos. In: ICCV 2003: Proceedings of the Ninth IEEE International Conference on Computer Vision, Washington, DC, USA. IEEE Computer Society, Los Alamitos (2003)
3. Schleicher, D., Bergasa, L., Barea, R., Lopez, E., Ocana, M., Nuevo, J.: Real-time wide-angle stereo visual slam on large environments using sift features correction, October 29 - November 2, pp. 3878–3883 (2007)
4. Davison, A.J., Molton, N.D.: Monoslam: Real-time single camera slam. IEEE Trans. Pattern Anal. Mach. Intell. 29(6), 1052–1067 (2007); Member-Reid, I.D., Member-Stasse, O.
5. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision 60(2), 91–110 (2004)
6. Hoshi, Y., Kobayashi, Y., Kasuya, T., Fueki, M., Kuno, Y.: Interactively instructing a guide robot through a network. In: International Conference on Control, Automation and Systems, 2008. ICCAS 2008, pp. 1841–1845 (2008)
7. Zhao, L., Li, R., Zang, T., Sun, L., Fan, X.: A Method of Landmark Visual Tracking for Mobile Robot. In: Xiong, C.-H., Liu, H., Huang, Y., Xiong, Y.L. (eds.) ICIRA 2008. LNCS (LNAI), vol. 5314, pp. 901–910. Springer, Heidelberg (2008)
8. Liao, M., Wei, L., Chen, W.: A novel affine invariant feature extraction for optical recognition, vol. 3, pp. 1769–1773 (August 2007)
9. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. Mach. Intell. 27(10), 1615–1630 (2005)
10. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide-baseline stereo from maximally stable extremal regions. Image and Vision Computing 22(10), 761–767 (2004)
11. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (surf). Comput. Vis. Image Underst. 110(3), 346–359 (2008)
12. Lindeberg, T.: Scale-space theory: A basic tool for analysing structures at different scales. Journal of applied statistics 21(2), 414–431 (1994)
13. Meger, D., Forssen, P., Lai, K., Helmer, S., McCann, S., Southey, T., Baumann, M., Little, J., Lowe, D., Dow, B.: Curious george: An attentive semantic robot. Robotics and Autonomous Systems 56(6), 503–511 (2008)
14. Forssen, P.E., Meger, D., Lai, K., Helmer, S., Little, J.J., Lowe, D.G.: Informed visual search: Combining attention and object recognition, pp. 935–942 (2008)

Active Contour Method with Separate Global Translation and Local Deformation

Linlin Zhu[1,2], Baojie Fan[1,2], and Yandong Tang[1]

[1] Robotics Lab, Shenyang Institute of Automation, Chinese Academy of Science, 110016 Shenyang, China
[2] Graduate School of the Chinese Academy of Science, 100039 Beijing, China
{zhulinlin,fanbaojie,ytang}@sia.cn

Abstract. Active Contour can describe targets accurately and has been widely used in image segmentation and target tracking. Its main drawback is huge computation that is still not well resolved. In this paper, by analyzing curve gradient flow, the evolution of active contour is divided into two steps: global translation and local deformation. When the curve is far away from the object, the curve just does the translation motion. This method can optimize the curve evolving path and efficiency, and then the computation cost is largely reduced. Our experiments show that our method can segment and track object effectively.

Keywords: Active contour, global translation, local deformation, gradient flow.

1 Introduction

Active contour model, also known as snake model, was pioneered in 1987 by Kass et al. in [1] for image segmentation via driving an initial contour toward a desired object edge with a PDE deduced by minimizing an energy functional. Active Contour can describe targets accurately, it is widely used in image segmentation and target tracking. But it has many limitations. For example, it is hard to undergo topological changes, sensitive to initial contour placement, and dependent on curve parameterization. In 1993, geodesic active contour [2], formulated as a weighting Euclidean arc length using an edge-stopped potential functional, was proposed by Casselles et al. in a level set framework [3]. It is independent of curve parameterization and can easily handle curve topological changes. Those models mentioned above are edge-based [2-5]. In most cases, they are less robust for image segmentation than region-based active contour models [6-9] because the latter models utilize certain image global region statistical information which partitions a given image into statistically distinct regions. In order to obtain desirable segmentation results, one important strategy for active contours research is combining certain prior knowledge with image information (gradient and region information etc.) to deal with images with insufficient information. The shape and topology of objects to be segmented are important prior information. Some paradigms using prior shape and topology information can be found in [10-14]. For more recent developments about active contours, we refer the readers to [15, 16].

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 876–884, 2009.

In order to make contour evolution to have certain desirable features to escape irrelevant local minimums, prior information on the deformation field of evolving contours can also be used. Clearly, the opposite of gradient flow driving contour evolution derived by minimizing its corresponding energy functional is directly relative to the prior information and affects the evolve path of curves. However, before the appearances of [17, 18], much of the literatures on active contours focused on energy functional building, while ignored the effect of gradient flows in curve evolving. Due to the arbitrary evolution of active contour which results in many undesirable segmentation and tracking results. Sundaramoorthi and Charpiat primarily proposed to design new inner products to yield desirable gradient flows for object tracking and rigidification simultaneously. Furthermore, Charpiat showed that the gradient of a given energy can also be considered as the result of a minimization problem from a new point and extended the definition of gradient toward more general priors [17].

In section 2, we introduce the relation of average gradient flow and the motion of active contour. Then we propose the global translation active contour in section 3. Our new active contour called two step active contour, which possesses the properties of global transformation and local smooth velocity fields, is proposed in section 4. Some experiments are presented in section 5, which is followed by some conclusions in section 6.

2 The Average Gradient Flow and the Motion of the Active Contour

Let M denote the set of smooth embedded curves in R^n, which is a differentiable manifold. For $C \in M$, the tangent space of M at C is denoted by $T_C M$, which can be seen as the deformation space. Given an energy function $E(C)$, for all admissible deformation fields v defined on C, its Gateaux derivative $\delta E(C, v)$ can be expressed as :

$$\delta E(C, v) \overset{def}{=} \lim_{\varepsilon \to 0} \frac{E(C + \varepsilon v) - E(C)}{\varepsilon} = dE(C) \bullet v \tag{1}$$

In order to get the gradient of $E(C)$, we must model the deformation space $T_C M$ as an inner product space, most of the time we use the $H0$ (L^2- type) inner product. And in $H0$ inner product, we get the following relation:

$$dE(C) \bullet v = \frac{1}{L} \int_C \nabla E_{H0} \bullet v \, ds = \langle \nabla E_{H0}, v \rangle_{H0} \Rightarrow \frac{\partial C}{\partial t} = -\nabla E_{H0} \tag{2}$$

Where L is the length of C, and the gradient flow ∇E_{H0} can be computed via the variational active contour method. The opposite of the $H0$ gradient flow is denoted by:

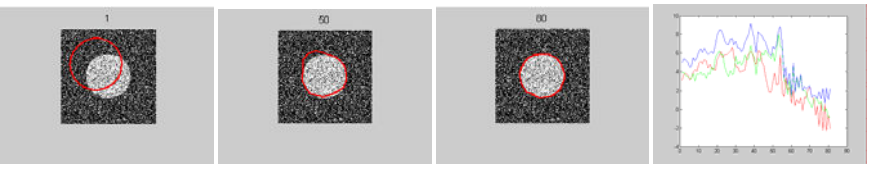
$$v0 = -\nabla E_{H0} \tag{3}$$

Let:

$$\overline{v0} = \frac{1}{L} \int_C v0 \, ds \tag{4}$$

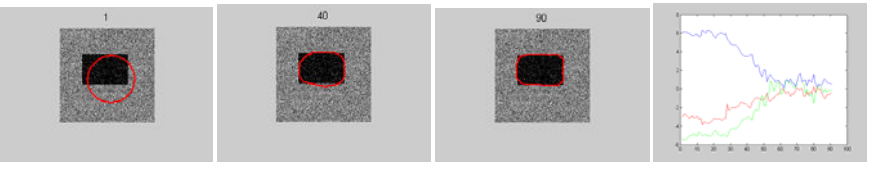
We call $\overline{v0}$ the average gradient flow. Though $\overline{v0}$ is the opposite of the real gradient flow, it still can reflect the motion state of the active contour.

The fig. 1(d) and fig. 2(d) shows the $\overline{v0}$'s change with the evolution of the curve, the red one is the Vx : the x component of $\overline{v0}$, the green one is the Vy: the y component of $\overline{v0}$ and the blue one is V : the magnitude of $\overline{v0}$, $V = \sqrt{Vx^2 + Vy^2}$. The abscissa is the iteration number and the ordinate is value of $\overline{v0}$ ' components.



(a) (b) (c) (d)

Fig. 1. The evolution and the $\overline{v0}$ distribution of a noise circle



(a) (b) (c) (d)

Fig. 2. The evolution and the $\overline{v0}$ distribution of a noise rectangle

Analyzing those figures, we learn that at the beginning of evolution, when the curve is far away from the object, the average evolution speed is high; as the curve moving nearby the object, the average evolution speed will be slowed down. In other words, when the average gradient flow's magnitude is over a certain threshold, the active contour is in global translation; otherwise, the active contour is in local deformation.

3 Global Translation Active Contour

To make the evolution in global translation more effective, we propose a global translation active contour model, in which a global deformation Vg is used to represent the global translation of the curve. Supposing that this method does not change the Gateaux derivative of C, we can get:

$$dE(C) \bullet v = dE(C) \bullet Vg \tag{5}$$

that means:

$$\langle \nabla E_{H0}, v \rangle_{H0} = \frac{1}{L}\int_C \nabla E_{H0} \bullet v ds = \frac{1}{L}\int_C \nabla E_{H0} \bullet Vg ds = \frac{Vg}{L}\int_C \nabla E_{H0} ds = \langle \nabla E_{H0}, Vg \rangle_{H0} \quad (6)$$

Vg satisfies Eq. (7) ,

$$\int_C v0 \bullet v0 ds = Vg \bullet \int_C v0 ds \quad (7)$$

Which is deduced from Eq.(3), (4) and (6). Eq.(7) implies that Vg with the smallest magnitude has the same direction with average gradient flow $\overline{v0}$.

In the global translation active contour, the curve evolves as Eq. (8);

$$\frac{\partial C}{\partial t} = Vg \quad (8)$$

All parts in the curve move in the same direction and speed. The translation evolving is more effective than that with speed $v0$.

4 The Two Step Active Contour

One of the important advantages of active contour is the excellent ability to describe the local deformation of objects exactly. It is necessary to replace the translation active contour by the active contour which allows the shape deformation of curves (traditional active contours) when the active contour close to the object. According to Fig.1,Fig.2 and the analysis in section 2, the average gradient flow $\overline{v0}$ will be slowed down when curve evolves nearby the object, so the magnitude of $\overline{v0}$ can be used to decide the motion of contour. This evolving process can be represented by Eq. (9) and Fig. 3.

$$\frac{\partial C}{\partial t} = \begin{cases} Vg & when \quad \| \overline{v0} \| > threshold \\ v0 & otherwise \end{cases} \quad (9)$$

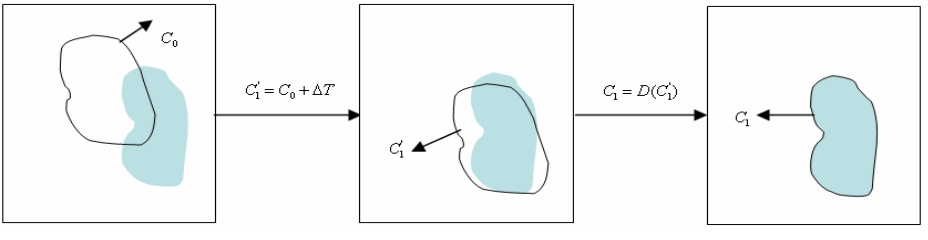


Fig. 3. The evolving course of proposed method. Fig. 3(a) is the target and the initialized contour. In Fig.3(b), the contour translates to the location position of target. Then the contour evolves to the right shape as shown in 3(c).

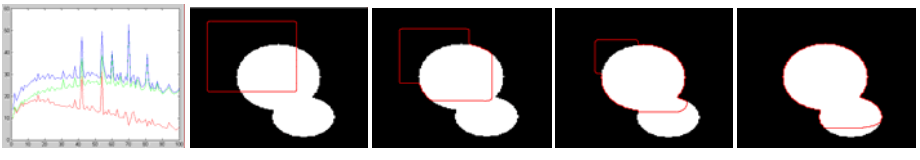
5 Experiments

We implement the proposed method for object segmenting and tracking. The energy models we used are the C-V active contour model (a region based model) and the Geodesic active contour model (a edge based model) which are proposed in references [7] and [2] respectively.

5.1 Segmenting Results

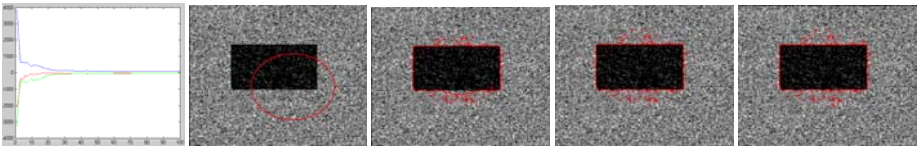
The segmenting results on two synthetic images are presented in fig 4 and fig 5. Fig 4(a~e) and fig 5(a~e) show the $\overline{v0}$ distribution with the curve evolving and the evolving results by the propose method. Meanwhile, fig 4(f~j) and fig5(f~j) are the $\overline{v0}$ distribute and evolving results by traditional method with the same initial curve, energy functional and parameters.

(a) $\overline{v0}$ distribution (b) initial contour (c)20 iterations (d) 60 iterations (e) 800 iterations

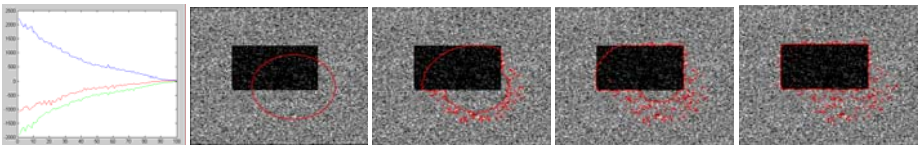


(f) $\overline{v0}$ distribution (g) initial contour (h) 60 iterations (i) 300 iterations (j) 1000 iterations

Fig. 4. The evolve results by proposed method and traditional method for two circles



(a) $\overline{v0}$ distribution (b) initial contour (c)100 iterations (d) 200 iterations (e) 800 iterations



(f) $\overline{v0}$ distribution (g) initial contour (h)100 iterations (I)400 iterations (j) 800 iterations

Fig. 5. The evolve results by proposed method and traditional method for a rectangle in noise

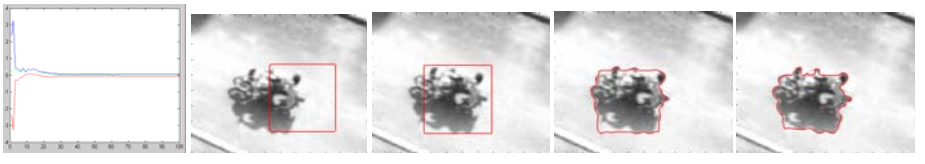
The convergence of $\overline{v0}$ by proposed method is faster compared to that of the traditional method, as shown in the following distribution maps, i.e., the curves in our method can move to the object faster and get balance in fewer iterations. It is validated by the evolving results of active contour in Fig. 4 and Fig. 5. Noise in images does not change the convergence character of our method.

The segment results for two real images are presented in fig 6 and fig 7. The $\overline{v0}$ distribution with the curve evolving and the evolving results by our proposed method are shows in the Fig. 6. (a~e) and Fig. 7. (a~e). Fig. 6. (f~j) and Fig. 7. (f~j) are the $\overline{v0}$ distribute and evolving results by a traditional method. The initial curve, energy functional and parameters are same in these methods. The results show that our proposed method reduces the iteration times and brings the better segmentation results . In the all experiments for segmentation, the C-V model is used.

(a) $\overline{v0}$ distribution (b)initial contour (c)10 iterations (d) 60 iterations (e) 200 iterations

(f) $\overline{v0}$ distribution (g) initial contour (h) 100 iterations (i) 300 iterations (j) 500 iterations

Fig. 6. The evolve results by proposed method and traditional method for a car



(a) $\overline{v0}$ distribution (b) initial contour (c)20 iterations (d) 60 iterations (e) 100 iterations

(f) $\overline{v0}$ distribution (g) initial contour (h) 100 iterations (i) 300 iterations (j) 500 iterations

Fig. 7. The evolve results by proposed method and traditional method for a motor with persons

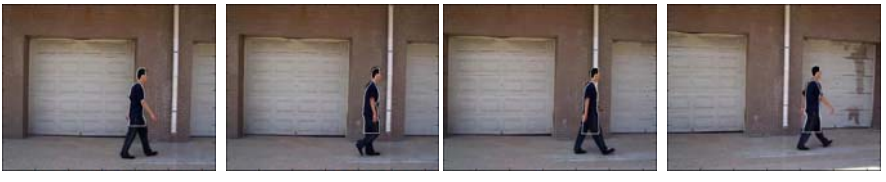
5.2 The Tracking Results

The proposed active contour method is also implemented in frame sequence for target tracking. Some results are shown below.

frame 20 frame 40 frame 60 frame 100

frame 120 frame 160 frame 190 frame 220

Fig. 8. The tracking results for a car



frame 2 frame 6 frame 10 frame 16

frame 20 frame 26 frame 32 frame 42

Fig. 9. The tracking results for a walking person

The tracking results are shown in figure 8~9. The car video sequence (figure 3) has 220 frames with the pixel resolution 320 * 240. Because of the complex background, we use the Geodesic active contour method to track the car in these frames. We implement our method with Matlab, and it takes 0.094s to handle each frame in average. The average iteration number for translation is 1.64. For the frames of walking person, the C-V active contour is used to get the silhouette of the object. It takes 0.8383s

to handle each frame in average; the average iteration number for translation is 5.12. From the results, we can see that even in dynamic and complex background, our method can track object accurately.

6 Conclusions

In this paper, we proposed a two-step active contour that divides the evolution of active contour into global translation and local deformation. The average gradient flow is analyzed to decide the motion of the curve. The global deformation variable Vg can be computed by a global active contour and represents the global translation of the curve when the evolving curve is far away from object. After the curve is close to the object, the local deformation replaces the global translation; then silhouette of the object will be described. The global translation active contour optimizes the evolving path and leaves a good curve for the deformation evolving. Because of the good characters of global translation active contour, the two-step active contour can reduce the iterative times and get the better evolving results.

Acknowledgments. This work is funded by the Natural Science Foundation of China (Grant No. 60871078 and 50876110). We would like to thank Fan Huijie, Sun Jing, Dong Wei and Du Yingkui for their advice and support.

References

1. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. International Journal of Computer Vision 1, 321–331 (1987)
2. Caselles, V., Catte, F., Coll, T., Dibos, F.: A geometric model for active contours in image processing. Numer. Math. 66, 1–31 (1993)
3. Osher, S., Sethian, J.A.: Fronts propagating with curvature dependent speed: algorithms based on Hamilton-Jacobi formulations. J. Comp. Phys. 79, 12–49 (1988)
4. Kichenassamy, S., Kumar, A., Olver, P.J., Tannenbaum, A., Yezzi, A.: Gradient flows and geometric active contour models. In: International Conference on Computer Vision, pp. 810–815 (1995)
5. Xu, C., Prince, J.L.: Snakes, Shapes, and Gradient Vector Flow. IEEE Transactions on Image Processing 7(3), 359–369 (1998)
6. Zhu, S.C., Lee, T.S., Yuille, A.L.: Region competition: Unifying snakes, region growing, energy /Bayes/MDL for multi-band image segmentation. International Conference on Computer Vision, 416–423 (1995)
7. Chan, T.F., Vese, L.A.: Active contours without edges. IEEE Trans. Image Processing 10(2), 266–277 (2001)
8. Yezzi, A., Tsai, A., Willsky, A.: A fully global approach to image segmentation via coupled curve evolution equations. Journal of Visual Communication and Image Representation 13, 195–216 (2002)
9. Kim, J., Fisher, J.W., Yezzi, A., Cetin, M., Willsky, A.S.: A nonparametric statistical method for image segmentation using information theory and curve evolution. IEEE Transactions on Image Processing 14, 1486–1502 (2005)

10. Leventon, M.E., Grimson, W.E.L., Faugeras, O.: Statistical shape influence in geodesic active contours. In: Proc. IEEE Comput. Soc. Conf. CVPR, vol. 1, pp. 316–323 (2000)
11. Chen, H.D., Tagare, S.T., Huang, F., Wilson, D., Gopinath, K.S., Briggs, R.W., Geiser, E.A.: Using prior shapes in geometric active contours in a variational framework. International Journal of Computer Vision 50(3), 315–328 (2002)
12. Cremers, D.: Dynamical statistical shape priors for level set based tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence 28, 1262–1273 (2006)
13. Han, X., Xu, C., Prince, J.L.: A Topology Preserving Level Set Method for Geometric Deformable Models. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(6), 755–768 (2003)
14. Sundaramoorthi, G., Yezzi, A.: More Than Topology Preserving Flows for Active Contours and Polygons. In: Tenth IEEE International Conference on Computer Vision, vol. 2, pp. 1276–1283 (2005)
15. Cremers, D., Rousson, M., Deriche, R.: A Review of Statistical Approaches to Level Set Segmentation: Integrating Color, Texture, Motion and Shape. International Journal of Computer Vision 72(2), 195–215 (2007)
16. Chan, T., Sandberg, B., Moelich, M.: Some Recent Developments in Variational Image Segmentation, University of California, Los Angeles: Los Angeles, CAM06-52, p. 36 (2006)
17. Charpiat, G., Maurel, P., Pons, J.P., Keriven, R., Faugeras, O.: Generalized Gradients: Priors on Minimization Flows. International Journal of Computer Vision 73(3), 325–344 (2007)
18. Sundaramoorthi, G., Yezzi, A.J., Mennucci, A.: Sobolev Active Contours. International Journal of Computer Vision 73(3), 109–120 (2007)

Automatic Soldering System Based on Computer Vision

Zhenhua Xiong, Xinjue Zhou, Yulin Wang, and Han Ding

School of Mechanical Engineering, Shanghai Jiao Tong University
800 Dong Chuan Road, Shanghai 200240, P.R. China
{mexiong,hit_zou,wyl_sjtu,hding}@sjtu.edu.cn

Abstract. Due to the fine pitch of today's micro-chips, accurate alignment is vital to acquire high quality joints with less detrimental effects on other components. We developed an integrated soldering system which combines an X-Y stage, a computer vision and a laser head. Based on the analysis of experimental data, we find that lens distortions and assembly angular errors of the X-Y stage play a distinct role on the positioning errors. To improve the positioning accuracy of this system, we proposed an algorithm based on straight line method to reduce image distortions, and a compensation algorithm on the non-perpendicular X-Y positioning stage. Finally, experimental results and comparative data are presented to evaluate the performance of these algorithms.

Keywords: Computer vision, system calibration, image distortion.

1 Introduction

With the increasing requirement on high integration of semiconductor packaging, fine pitch chips are desired to further decrease the manufacturing cost. Accordingly, an accurate and precise integration system is indispensable. Since laser soldering has the advantages of precise work areas, high-energy efficiency, reliability and ease for automation, we developed an integrated system to automatically sense the soldering pads, fulfill the alignment of laser and pads, and finish the soldering work. However, the precision of the system is determined not only by the work areas of laser beam but also the alignment of the part to be soldered. Therefore, improving the positioning accuracy of the automatic soldering system would be necessary for acquiring high quality joints with less detrimental effects on other components of the chips.

The positioning errors can be due to camera lens distortions, machine assembly errors, and improper parameters of control mechanism. Among these error sources, lens distortions and machine assembly errors play a dominant role in the positioning precision of the system. All of these errors can be generally classified under two main categories: 1) systematic errors which are completely repeatable and reproducible; and 2) apparently random errors which vary under similar operating conditions [1].

This paper is mainly focusing on lens distortions and mechanical assembly errors, especially the angular error of the X-Y positioning stage. To reduce lens distortions of the system, an algorithm counteracting image distortions would be proposed. The distortion parameters are usually estimated as intrinsic parameters[2-4]. However, in this case, a certain calibration model is needed, which requires the model's space

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 885–896, 2009.

coordinate and pixel coordinate of corresponding image. Moreover, solving non-linear distortion parameters requires processes of camera intrinsic and extrinsic solution simultaneously. Due to such limitations, we adopted the algorithm based on straight line method which is independent on the limitations mentioned above.

To compensate for the mechanical positioning errors for the system, the main issues would relate to mechanical assembly errors of the X-Y positioning stage, in particular its non-vertical error. Early compensation methods mainly utilized mechanical correctors, in the form of leadscrew correctors, cams, reference straightedges etc [1,5]. However, mechanical correction inevitably increases the complexity and cost of the system. Software-based error compensation schemes, due to its ease of realization and low cost, can be more suitable to our system which works as flexible packaging unit or chip reworks.

This paper is organized as follows. Section 2 generally describes the developed automatic soldering system. Section 3 presents main error sources in developed system. Correcting lens distortion and compensating mechanical errors are presented in Section 4 and 5, repectively. Experimental results are shown in Section 6, followed by our conclusion in Section 7.

2 Soldering System Description

A block diagram of the automatic soldering system is shown in Fig. 1. The system is composed of a laser module including a laser diode generator, fiber-optic bundle, focusing lens; and a computer vision module including a camera, imaging lens and an illuminator; and the motion module, the X-Y positioning stage.

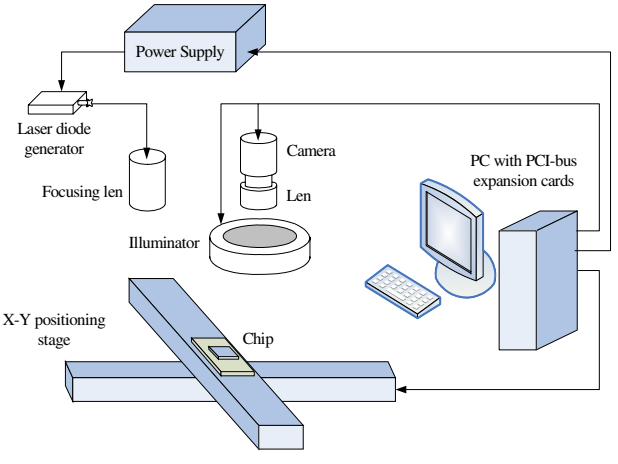


Fig. 1. Block diagram of the developed automatic soldering system

The X-Y stage drives the chip to the camera's vision area, allowing the camera to view the target area; and to the laser's work area, allowing the target to be operated by laser beam. A PCI-bus card in the PC controls the X-Y stage. Illumination of the

target for image processing is provided by a red LED light source beneath the camera. The camera is connected by a PCI-bus frame grabber in the PC. Finally, the output of laser, driven by the current up to 2800mA from a specific laser power supply, is directed through focusing lens down toward the target area.

According to the graphical programming language that uses icons instead of lines of text to create applications, LabVIEW offers the advantages of efficient programming, ease of user interface design, and convenient integrated environments interfacing with real-world signals. This automatic soldering system's interface is constructed by LabVIEW. The user interface and the system are shown in Fig. 2. The main window is mainly composed of three parts – window for video and position display, states display field, and operation fields including laser module, computer vision module and motion control module. The communication interface consists of three data sockets including interface between PC and motion control card, frame grabber, and laser diode driver, respectively.

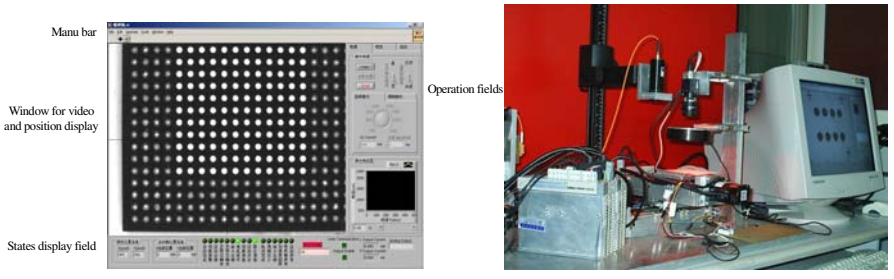


Fig. 2. The user interface and the developed system

3 Error Sources in the Developed System

Since the X-axis and the Y-axis are both of low cost, initial experiment data shown that the linear positioning and accumulated errors of the two axes are within 20μm and 30μm, respectively, under the point-to-point motion manner. Fig. 3 shows the stage and single axis positioning errors of both X-axis and Y-axis. However, actual positioning precision is unsatisfied when localizing a specified target. The positioning deviation may be up to 100 to 200μm. Excluding the linear positioning error of X-Y stage, we speculate that such errors may involve with lens distortions or assembly errors of the X-Y stage.

With some experiments, we find that the ideal positions and the actual positions have regular deviation, which indicates that the X-Y stage might introduce angular errors. In Fig. 4, we show an experiment justifying the existence of the regular deviation and its role in causing inaccurate positioning. In this experiment we drive the stage to move step to step through the direction of X-axis under a constant step while fixing the position of Y-axis. A sequence of position line was obtained from the experiment. Under ideal condition, such line should be perpendicular to Y-axis as the colorful triangles shown in Fig. 4. However, actual position line shows that it has an angle error with X-axis, though it has good linear characteristics. We find that the travel of 20mm in X-direction will introduce around 0.5mm deviation in Y-direction.

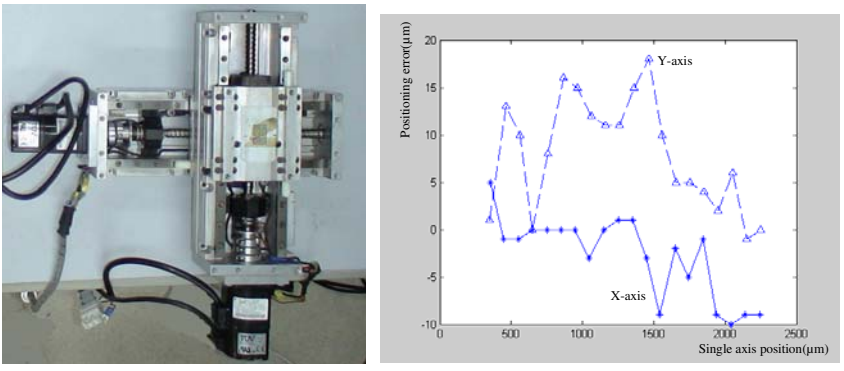


Fig. 3. The stage and single axis linear positioning precision

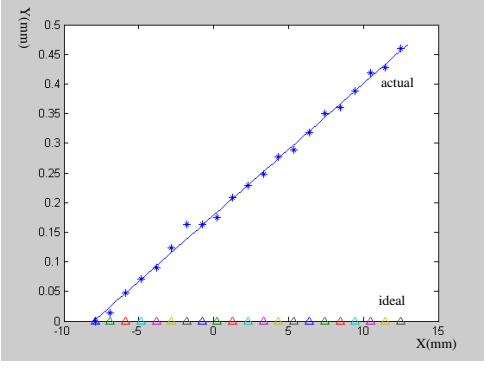


Fig. 4. Experiment of measuring positioning precision on the X-Y stage

Furthermore, the other key factor influencing the target localization can attribute to the lens distortion. When the distortions are large compared to other sources of error, the application requires prior calibration. Image taken with inexpensive lens may diverge substantially from a perspective camera model.

4 Compensation for Image Distortion

We adopted the algorithm based on straight line method which is independent on the limitations mentioned in the introduction section.

4.1 Algorithm Principle

Straight line method can be applied to solve the coefficient of lens distortion, as long as images have line characteristic objects.

1. Extracting characteristic figures

In order to analysis the slope of each line in image, we first acquire several pixels in image, group similar pixels, and then link corresponding pixels as line segment.

2. Solving distortion coefficient using optimization method

First we suppose the straight line is:

$$x_u \sin\theta + y_u \cos\theta = \rho , \tag{1}$$

where ρ and θ represent the distance between (x, y) and image center, and inclination between this line and the X-axis repectively.

The discrepancy objective function is defined as

$$F = \sum_{m=1}^{M} \sum_{k=1}^{N_m} (\bar{x}_k \sin\theta_m + \bar{y}_k \cos\theta_m - \rho_m)^2 , \tag{2}$$

where F is the distortion parameter, M represents the quantity of characteristic lines, and N_m is the number of sampling points on the m-th line.

3. Correcting non-linear distortion of image

Generally, correcting extracted characteristic pixels and whole pixels are two main methods of correction. We proposed an improved algorithm based on straight line in this paper.

Considering the line equation of (1), assuming (x_u^i, y_u^i) belongs to line l (i=1,…,N). So the slope of the line l is:

$$k^i = \tan\theta = \frac{y_u^i - y_u^{i-1}}{x_u^i - x_u^{i-1}} . \tag{3}$$

Since each point of the line coordinate is relate to a point in distorted image coordinate, the line l can be transformed to line l' in image coordinate. So we have:

$$f(u^i, v^i) = \frac{y_u(u^i, v^i) - y_u(u^{i-1}, v^{i-1})}{x_u(u^i, v^i) - x_u(u^{i-1}, v^{i-1})} = k^i . \tag{4}$$

In order to get distortion parameters, optimization function is set as:

$$E = \frac{1}{N-2} \sum_{i=2}^{N-1} (f(u^i, v^i) - f(u^{i+1}, v^{i+1}))^2 , \tag{5}$$

where (x_d^i, y_d^i) is the point of line l with the sequence of i (i=1,…,N) and N is the quantity of sampling points of line l. It is an optimization problem which can be solved using Least-Squares Fit or other linear optimized methods.

Equation (4) can be further transformed as:

$$y_u(u^i, v^i) - y_u(u^{i-1}, v^{i-1}) - k^i (x_u(u^i, v^i) - x_u(u^{i-1}, v^{i-1})) = 0 , \tag{6}$$

which is a linear function and can be transformed to matrix form

$$AX = B , \tag{7}$$

where A is a $a \times b$ matrix, the a represents the quantity of sampling points, and the b represents the quantity of unknown parameters. Therefore, replaced with equation (6), the optimization function (5) would be

$$E = \frac{1}{N-2}\sum_{i=2}^{N-1}(y_u(u^i, v^i) - y_u(u^{i-1}, v^{i-1}) - k^i(x_u(u^i, v^i) - x_u(u^{i-1}, v^{i-1})))^2 . \quad (8)$$

Specifying the number of sampling lines to be M , thus the optimization function is

$$E = \frac{1}{M(N-2)}\sum_{j=1}^{M}\sum_{i=2}^{N-1}(y_u^j(u^i, v^i) - y_u^j(u^{i-1}, v^{i-1}) - k_{ji}(x_u^j(u^i, v^i) - x_u^j(u^{i-1}, v^{i-1})))^2 . \quad (9)$$

The algorithm described above only needs to solve one linear function, and thus improve the efficiency of correcting process.

4.2 Numerical Simulation and Real Example

Since high order items have little influence on whole distortion and could be ignored, we just consider the parameter K_1 which influences the distortion greatly. The simplified distortion model is

$$\begin{aligned} x_u &= u + \bar{u}K_1 r_d^2 , \\ y_u &= v + \bar{v}K_1 r_d^2 , \end{aligned} \quad (10)$$

where $(X_u \quad Y_u)$ is ideal coordinate $(u \quad v)$ is real coordinate, $\bar{u} = u - u_0 \quad \bar{v} = v - v_0$ are the pixcel distance from image center, $r_d^2 = \bar{u}^2 + \bar{v}^2 \quad K_1$ is the parameter for radial distortion.

Supposing the $K_1 = 2\times10^{-5}$, and $s_x = s_y = 1$ pixel/mm. A 320*240 image consisting of 2 lines, $l_1 : y = 100$ and $l_2 : x + y = 50$, is used as the test image. Using known distortion parameters, the line points were distorted. Table 1 shows the ideal sampling points and their corresponding distorted points.

Table 1. Numerical simulation on image distortions

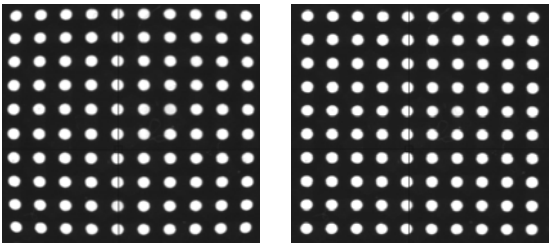
l_1		l_2	
(x_u, y_u)	(x_d, y_d)	(x_u, y_u)	(x_d, y_d)
(100,100)	(79.728,79.728)	(50,100)	(42.385,84.770)
(50,100)	(42.385,84.770)	(0,50)	(0,47.814)
(0,100)	(0,86.883)	(-50,0)	(-47.814,0)
(-50,100)	(-42.385,84.770)	(-100,50)	(-84.770,42.385)

The slope of line l_1 and l_2 can be estimated by least-square linear regression and have answers of 0 and 1, respectively. Solved linear functions about K_1 under the distortion model of

$$y_u = v + (v - v_0) \cdot \left((u - u_0)^2 + (v - v_0)^2 \right) \cdot K_1 . \tag{10}$$

Then we have $K_1 = 2.001 \times 10^{-5}$.

Then the average error between ideal position and actual position of corrected image is $\overline{e} = 0.007$ pixel.



Source Corrected

Fig. 5. Experimetal images before and after compensation

Fig. 5 shows the contrastive images of source and corrected one when applying this algorithm on a 10 by 10 image template. The centers of the white dots, which are testing pads on the PCB board, are used to form line segements in the computation.

5 Compensation on the X-Y Positioning Stage

5.1 Measuring the Assembly Errors

There are two methods of measuring the assembly angular errors: fixing one of the axes and then driving the other one, or driving both of them simultaneous. The first one has the advantage of more accurate results than the second one, but more laborious since the same measurement should be repeated to acquire the errors of X-direction and Y-direction, respectively. For better experimental accuracy, we decided to apply the first tesing method.

First of all, we suppose that the X-axis of X-Y stage is parallel to the X-axis of image and specify the step of motion as 1mm. Secondly, the X-axis is driven as constant steps with a fixed Y-axis. Thirdly, the pad's position of every step is acquired through image processing and stored. Finally, Least-Squares Fit method is introduced to calculate the error between the mechanical coordinate and pixel coordinate. Fig. 6 shows the principle of the measurement.

Final results calculated by MATLAB show the slope of the fitted line:

$$a_1 = 0.023 .$$

Therefore, the Y-axis's angular error between mechanical coordinate and pixel coordinate is

$$\theta_1 = \tan^{-1} a_1 = 1.2775^\circ .$$

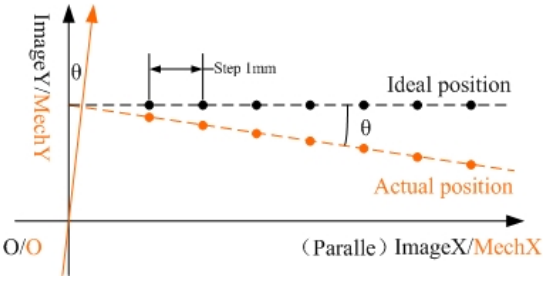
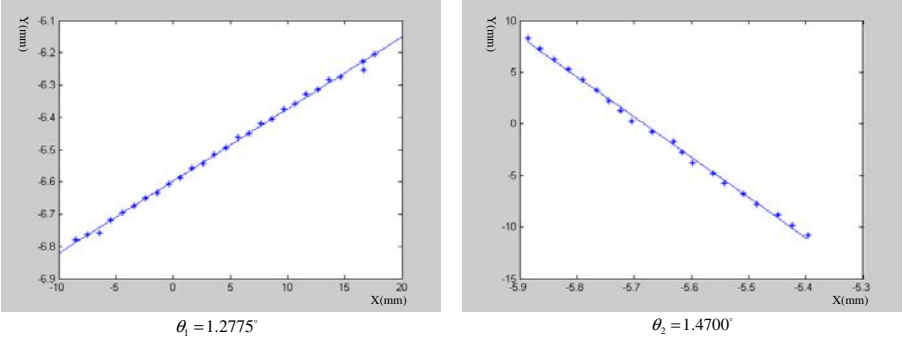


Fig. 6. Measuring the mechanical angular errors of X-Y positionning stage

Similarly, the X-axis's angular error between mechanical coordinate and pixel coordinate is

$$\theta_2 = \tan^{-1} a_2 = 1.4700^{\circ}.$$

Fig. 7 shows the line fitted by method of Least-Squares Fit.



$\theta_1 = 1.2775^{\circ}$ $\theta_2 = 1.4700^{\circ}$

Fig. 7. Least-squares fitting of the experimental data

5.2 Compensation Algorithm

In order to compensate the assembly errors, we show the correlation of angular errors, pixel coordinate and mechanical coordinatein Fig. 8.

In Fig. 8, θ_1 and θ_2 represent the errors of X-direction and Y-direction, respectively. X_a is the ideal target X-position's projection on X-axis of mechanical coordinate and Y_a represents the Y-position's projection on Y-axis of mechanical coordinate.

From the figure, we can easily have a conclusion that the real position of target in mechanical coordinate should be the orange star points shown in Fig. 8, if we want target's position to match the specified position under the condition of the existence of assembly errors. In other words, due to the existence of assembly angular errors,

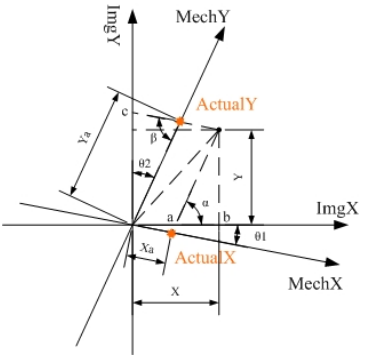


Fig. 8. Principle of compensation

we need to drive the target to the position of orange star points rather than the ideal position we assumed before. For example, assuming that the ideal position is (5,5). In this case, the position should be modified to (4,6) to match the specification.

According to the relationship shown in Fig. 8, we have:

$$\frac{X_a}{\sin\alpha} = \frac{oa}{\sin(\frac{\pi}{2} - \theta_1 + \theta_2)}, \tag{11}$$

where $\alpha = \frac{\pi}{2} - \theta_2$, $oa = X - Y\tan\theta_2$.

Therefore, we have the actual position of X is

$$X_{mech} = X_a = \frac{X - Y\tan\theta_2}{\sin(\frac{\pi}{2} - \theta_1 + \theta_2)} \cdot \sin(\frac{\pi}{2} - \theta_2). \tag{12}$$

Similarly, we have

$$\frac{oc}{\sin\beta} = \frac{Y_a}{\sin(\pi - \beta - \theta_2)}, \tag{13}$$

where $\beta = \frac{\pi}{2} + \theta_1 - \theta_2$, $oc = Y + X\tan\theta_1$. And the actual position of Y is

$$Y_{mech} = Y_a = \frac{Y + X\tan\theta_1}{\sin(\frac{\pi}{2} + \theta_1 - \theta_2)} \cdot \sin(\frac{\pi}{2} - \theta_1). \tag{14}$$

Equation (12) and (14) are the final results, with which we can command our motion control unit to align the laser beam and the soldering pads accurately.

6 Experimental Results

Plenty of target positioning experiments have been attempted on the developed soldering system based on the proposed compensation method. In this section, we show the experimental data and compare them with the counterpart of un-compensated system.

Experiments mentioned in section 5.1 are repeated in this section. The results show that angular errors introduced by assembly process were decreased greatly. Table 2 addresses the results before compensation and after compensation.

Table 2. Angular errors of the X-Y positioning stage

	Before compensation	After compensation
Error in X-direction	1.2775°	0.0306°
Error in Y-direction	1.4700°	0.0416°

Moreover, the experiment to measure the target positioning precision is attempted. During the experiment, the center of the pixel coordinate is defined as the center of the image and all of the positioning motion is confined in the scope of the camera, which means that under the ideal condition, the X-Y stage should drive the target pad to locate on the center of the image when the target position is acquired. First of all, we specified the target pad. Secondly, computer vision captured the target pad and then mapped the position of the target pad in the pixel coordinate to corresponding mechanical coordinate. Thirdly, the X-Y stage was driven to the target position under the map of computer vision. Finally, target positioning error was obtained when computer vision capturing the target position in image again, since the target pad was located on the image center, re-mapping its position would indicate the discrepency between ideal and actual position.

The target positioning precision of the system is presented in Table 3.

Table 3. Positioning precision of X-Y stage before and after compensation

	Before compensation		After compensation	
	X-axis	Y-axis	X-axis	Y-axis
Mean of the positioning error (mm)	0.02736	0.02349	0.00012	-0.00109
σ (mm)	0.05511	0.07433	0.007032	0.005812

From Table 2 and Table 3, it is clear that the system's positioning precision is improved greatly after compensation. The results of this experiment show that compensated system has the precision of 1μm±7μm(1σ) and satisfies the positioning precision of the soldering job with bump's diameter less than 0.3mm.

Given the high precision of the positioning system, we can get good alignment of the laser beam, the BGA ball and the corresponding sodering pad. Moreover, the computer vision system can be used to inspect the quality of reflowed BGA balls from their appearance.

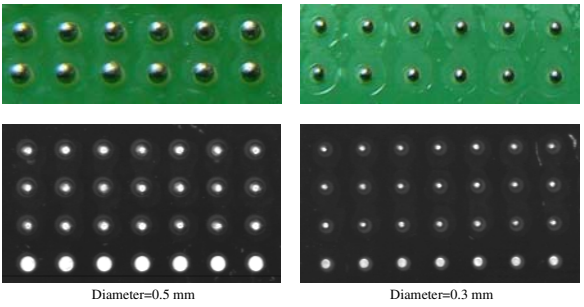


Fig. 9. Soldered balls by the automatic soldering system

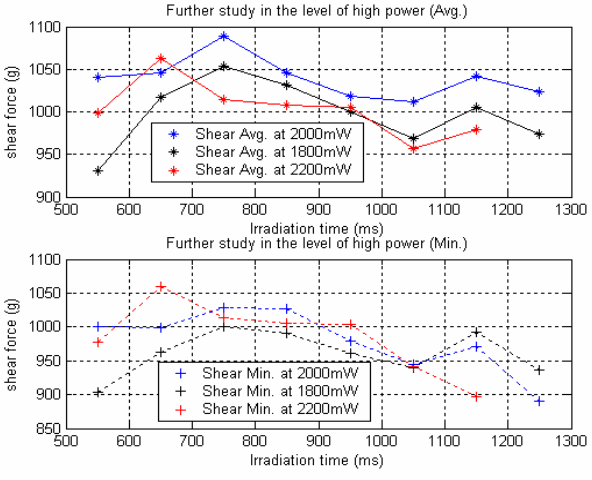


Fig. 10. Comparison of soldering quality vs different input powers

Furthermore, a variety of soldering tasks have been attempted by the soldering system just described. Fig. 9 shows the samples of solded BGA with this system. Fig. 10 shows the shear force test results under different laser inout powers. From these results, we can find the optimal parameters of the system.

7 Conclusion

Developing a low cost and more flexible automatic soldering system with high positioning accuracy should greatly increase the attractiveness on packaging applications.

A prototype automatic soldering system has been constructed and introduced in this paper. The key factors on inaccurate target positioning, such as image distortion and assembly errors of positioning stage, have been identified. Consequently, the algorithms for calibrating image distortion and compensating mechanical assembly

errors are discussed. The experimental data evaluated the performance of the developed system after applying the proposed approaches. The target positioning satisfies the precision specification of the application on micro-BGA.

Acknowledgments

This work is supported in part by the National High-tech Research and Development Program (863 Program) under grant No. 2007AA04Z316, the Science & Technology Commission of Shanghai Municipality under grant No. 07QA14032 and 08JC1411600.

References

1. Tan, K.K., Huang, S.N., Seet, H.L.: Geometrical Error Compensation of Precision Motion Systems Using Radial Basis Function. IEEE Transcriptions on Instrumentation and Measurement 49(5), 984–991 (2000)
2. Shah, S., Aggarwal, J.K.: Intrinsic Parameters Calibration Procedure for A (Hign-Distortion) Fish-Eye Lens Camera With Distortion Model and Accuracy Estimation. IEEE Trans-PAMI 27(11), 346–351 (1996)
3. Yan, G., Chen, I.-M.: A Novel Kinematic Calibration Algorithm for Reconfigurable Robotic System. In: Proc of 1997 IEEE International Conf. on Robotics and Automation, Albuquerque, New Mexico, April 1997, pp. 3197–3202 (1997)
4. Stevenson, D.E., Fleck, M.M.: Nonparametric Correction of distortion. In: Proc. 3rd IEEE Work Shop on Applications of Computer Vision, Sarasota, FL, December 1996, pp. 214–219 (1996)
5. Choi, H.-S., Han, C.-S., Kim, S.-S., Kim, E.-Z., Choi, T.-H., Na, K.-H.: A Study on the Development of A Micro-Positioning Device for Forming Milli-structure. In: SICE 2003 Annual Conference, Fukui, Japan, August 2003, pp. 1644–1649 (2003)

PLG-Based Visual Tracing for Eye-in-Hand Puma 560 Robot

Wen-qiang Hu[1], Kai Tian[2], and Rong-Lei Sun[2,*]

[1] School of Software Engineering
[2] School of Mechanical Science and Engineering,
Huazhong University of Science and Technology, Wuhan 430074, China
ronglei@mail.hust.edu.cn

Abstract. In this paper, one algorithm used to control eye-in-hand robot motion is studied and simulated. First, the fundament matrix of visual system is computed based on epipolar geometry; then the matrix is segmented to get the instantaneous poses and moving trajectories of end-effector based on IBVS. Finally one simulating experiment is completed and the experimental results demonstrate the efficiency of the proposed mothod.

Keywords: IBVS, Epipolar Geometry, Visual Servo Control.

1 Introduction

Visual servo control refers to use computer vision data to control the motion of a robot. Now it is one main way to control n-joints robot in $\boldsymbol{m}$ dimensions working spaces [1][2]. And the keys are recovering relative pose of the target with respect to initialization and compute the motion parameters. So estimating the relative poses and motion parameters from image information is becoming studied hotspots.

Roughly speaking, there are two main visual servo control approaches. One is position-based visual servoing(PBVS), the other is image-based visual servoing (IBVS) [3]. In PBVS, the relative pose of the target with respect to initialization must be estimated. And the robot's moving trajectory in the cartesian space can be computed using path tracking techniques. But in this method, the 3D reconstruction from image information must be computed accurately in real-time constraints. And, to a PBVS system, the stability is affected easily by calibration errors. In IBVS, the relative pose of the target with respect to initialization can be resolved based on jacobian matrix. And the reconstruction process is not in need, so IBVS is compatible with the real time constraint. However, IBVS requires an a priori knowledge on the 3D geometry of the target.

In this paper, IBVS is chosen as the basis control method because of the time demand. But 3D geometry of the target can not former known in most. In order to resolve the question, epipolar geometry and image technology are fused to extract and match feature points. Then, these feature points are dealt with using the epipolar geometry and IBVS.

* Corresponding author.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 897–905, 2009.

2 Basis Principles

In order to analysis and design the control model of robot, The control error function is as follow in the equation(1).

$$e(t) = s(m(t), a) - s^{*} \tag{1}$$

Vector $\boldsymbol{m}(t)$ is image measurement data. $\boldsymbol{a}$ represents potential additional knowledge about the system. Vector $\boldsymbol{s}$ contains desired target. In IBVS, $\boldsymbol{s}$ is one group data about these extracted feature points.

In single visual system, robustness of IBVS is affected by demarcated error and image sounds. And the feature points must lay in the vision field during camera moving.

Given a two-view epipolar geometry consisting of 3D point p and two focular cores $\boldsymbol{C}_{\mathrm{l}}$, $\boldsymbol{C}_{\mathrm{r}}$ is as shown in Fig. 1 [4]. According to the epipolar constrain poses of one camera in different time can be obtained from image information. And further more the moving trajectory of the camera can be computed.

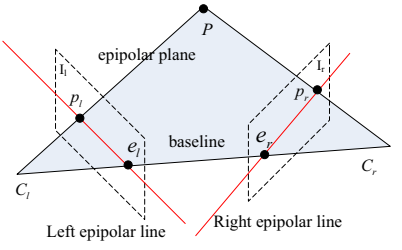


Fig. 1. Epipolar geometry

Points $\boldsymbol{p}_l$ and $\boldsymbol{p}_r$ are the corresponding points of $\boldsymbol{p}$ in two images $\boldsymbol{I}_{\mathrm{l}}$ and $\boldsymbol{I}_{\mathrm{r}}$. Base-line $\overline{C_l C_r}$ is the conjunctive line of facular cores of two cameras. $\overline{e_l p_l}$ and $\overline{e_r p_r}$ are the intersection between epipolar plane $\boldsymbol{C_l p C_r}$ and image planes $\boldsymbol{I_l}$ and $\boldsymbol{I_r}$. Intersection of all epipolar lines in one image is the pole. If $\boldsymbol{p}_{\mathrm{l}}$ lays in $\boldsymbol{I}_{\mathrm{r}}$, $\boldsymbol{p}_r$ must lay in epipolar line in $\boldsymbol{I}_{\mathrm{r}}$, and the equation (2) can be gotten.

$$p_r^T \cdot F \cdot p_l = 0 \tag{2}$$

F is the fundamental matrix. If the coordinate of $\boldsymbol{p}_l$ is known, one linear eauqtion about $\mathrm{p_r}$ is constructed.

Once structure parameters ($\boldsymbol{R}$ and $\boldsymbol{t}$), and inner parameter matrixs (A_l and A_{r}) of two cameras are known, the fundamental matrix is expressed as in equation (3). S is the antisymmetric matrix of the parallel moving vector $\boldsymbol{t}$.

$$F = A_r^{-T} S R A_l^{-1} \tag{3}$$

When A_l and A_r are known, the unitary image coordinate can be computed. $\hat{p}_l = A_l^{-1} p_l$, $\hat{p}_r = A_r^{-1} p_r$, $\hat{p}_l$ and $\hat{p}_r$ are the unitary coordinate of $\boldsymbol{p}_l$ and $\boldsymbol{p}_r$. So the equation (9) can be expressed using the equation (4).

$$\hat{p}_r^T RS\hat{p}_l = 0 \tag{4}$$

F is used to analysis the geometrical relation of these images of camera with different poses. Once F is computed, the pose of camera is obtained.

In this paper, one visual servo control system based the epipolar geometry and IBVS is studied and simulated. And we chose Eye-in-Hand Puma 560 Robot as the controlled object, thus control with time t can be computed according to moving trajectories of camera.

3 Algorithm Outline

Here, pose of the end-effector is adjusted according to the feedback image information based on epipolar geometry and IBVS. First, feature points are extracted and matched according to pixel grey and epipolar constrain; second, rotating trajectory of camera from the initial to target pose is gotten using epipolar geometry policy; third, the translational trajectory of camera can be computed using IBVS. Through these steps, the robot can move easily to the target.

There are many methods that can be used to extract and match feature points, such as Harris corner point algorithm. The basic idea of Harris algorithm is to extract corner point based on the gray value. Thus Harris algorithm is not affected by the poses of camera and light. And the equation of Harris corner algorithm is as shown in (5):

$$\begin{aligned} R(x,y) &= \det\left[M(x,y)\right] - k \cdot trace^2\left[M(x,y)\right] \\ M(x,y) &= \begin{bmatrix} \left[\dfrac{\partial I}{\partial x}\right]^2 & \left[\dfrac{\partial I}{\partial x}\right]\left[\dfrac{\partial I}{\partial y}\right] \\ \left[\dfrac{\partial I}{\partial x}\right]\left[\dfrac{\partial I}{\partial y}\right] & \left[\dfrac{\partial I}{\partial y}\right]^2 \end{bmatrix} \end{aligned} \tag{5}$$

$I(x \quad y)$ expresses gray value of the pixel. Once the Harris operator R is bigger than threshold k designed, the pixel is one corner. Since one corner is found, the corresponding corner can be found as follows. First, one search field of matching point in image I_l is defined, then every corner in the field is found out, and compute its correlation degree with every corner in image I_r. If one pair corners in I_l and I_r have the biggest pertinence, the two corners are looked as matched pair.

For point $p_i \in I_r$, $\hat{p}_i(X) = [u_i(X) \quad v_i(X) \quad 1]^T$, the epipolar line is $[a_i \quad b_i \quad c_i]^T = E \cdot \hat{p}_i$. And the error function can be constructed as follow:

$$e_i(X) = \hat{p}_i^T E \hat{p}_l = [u_i(X) \quad v_i(X) \quad 1] \cdot \begin{bmatrix} a_i \\ b_i \\ c_i \end{bmatrix} \\ = a_i \cdot u_i(X) + b_i \cdot v_i(X) + c_i \tag{6}$$

The equation (12) is derivated, thus the equation (7) can be gotten.

$$e_1'(X) = a_i \cdot u_i'(X) + b_i \cdot v_i'(X) = [a_i \quad b_i] \cdot \begin{bmatrix} u_i'(X) \\ v_i'(X) \end{bmatrix} \tag{7}$$

The Jacobean matrix is expressed using the equation (8) and is added into the equation(7), thus the equation (9)can be constructed.

$$\begin{bmatrix} \dot{u} \\ \dot{v} \end{bmatrix} = J_{img} \cdot \begin{bmatrix} v_x \\ v_y \\ v_z \\ \omega_x \\ \omega_y \\ \omega_z \end{bmatrix}, J_{img} = \begin{bmatrix} \frac{f}{z} & 0 & -\frac{u}{z} & -\frac{uv}{f} & \frac{f^2+v^2}{f} & -v \\ 0 & \frac{f}{z} & -\frac{v}{z} & \frac{-f^2-v^2}{f} & \frac{uv}{f} & u \end{bmatrix} \tag{8}$$

$$e_1'(X) = [a_i \quad b_i] \cdot J_{img} \cdot \begin{bmatrix} \mathrm{v_c} \\ \omega_c \end{bmatrix} \tag{9}$$

In this paper, the controlling input is velocity $U = [\mathrm{v_c} \quad \omega_c]^T$, and error function is exponential convergence $\dot{e} = -\lambda \cdot e \quad \lambda > 0$. From the equation (9), $U = -\lambda \cdot J_{img}^+ \cdot e$, and J_{img}^+ is the pseudo-inverse matrix of J_{img}. Though above step, the opposite feature point must lay in the epipolar line, and the error function is equal to 0, thus the pose of the camera is adjusted to target. So the task in next step is moving the camera to the object location based on the polar construction. And the second task function is constructed, as is shown in follow:

$$e_2 = (x_i - x_d)^2 + (y_i - y_d)^2 + (z_i - z_d)^2 \tag{10}$$

Thus the whole task function is constructed:

$$e = W^+ e_1 + \beta (I_6 - W^+ W) \frac{\partial e_2}{\partial X} \tag{11}$$

For hand-in-eye system, once the relative pose of the target with respect to initialization is computed. The pose of the edd-effector can be computed. And all variable parameters of joints of robot are gotten using inverted kinematics. The flow chart in detail is shown in Fig. 2.

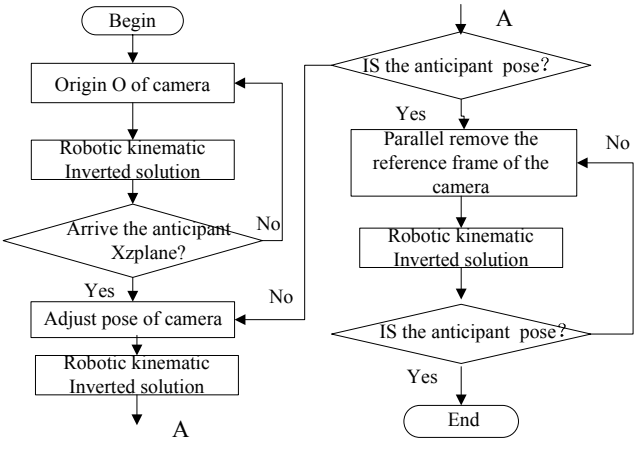


Fig. 2. Flow chart of control algorithm

4 Simulation Experiment

In this paper, the PUMA560 robot is chosen for experiments. PUMA560 has 6 degrees of freedom. The former three degrees confirm the coordinate of reference point of the end-effector, and the latter three ones define the pose of the end-effector [5]. The reference frames of connecting rods of robot are as shown in Fig. 3, and the DH parameter is shown in Table 1.

The equation (12) expresses the relative pose of the connecting rods in reference frame corresponding to the basis reference frame {0}.The ${}^{i-1}_{i}T(\theta_i)$ is the transformation

Table 1. The DH parameters of PUMA560

i	a_{i-1}	α_{i-1}	d_i	θ_i	variable range
1	0	0°	0	θ_1	-160° 160°
2	0	-90°	149.09	θ_2	-225° 45°
3	431.8	0°	0	θ_3	-45° 225°
4	20.32	-90°	433.07	θ_4	-110° 170°
5	0	90°	0	θ_5	-100° 100°
6	0	-90°	0	θ_6	-266° 266°

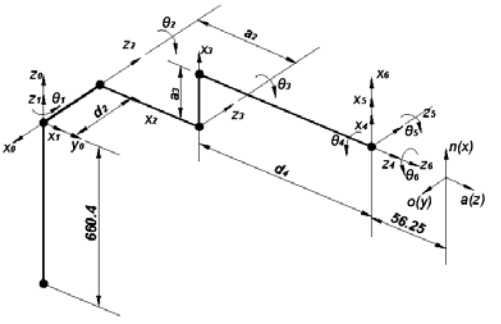


Fig. 3. The coordinate frames of all links of PUMA560

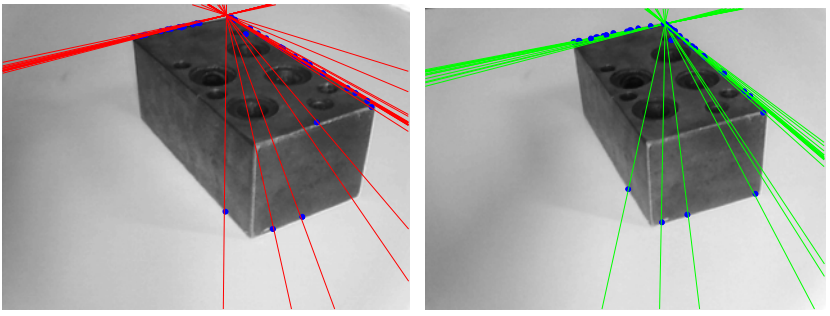
matrix of connecting rods. θ_i is joint variable parameter of the joint *i*, and n is the normal vector of the paw, o is the azimuth vector. and a is the approximating vector, p is the position vector.

Kinematic inverse solution of the PUMA560 is computed using inverse transform method. If the pose of the end-effector is known, the variable parameters of the joints θ_i can be computed.

$$ {}^0_6T = {}^0_1T(\theta_1)\,{}^1_2T(\theta_2)\,{}^2_3T(\theta_3)\,{}^3_4T(\theta_4)\,{}^4_5T(\theta_5)\,{}^5_6T(\theta_6) = \begin{bmatrix} n_x & o_x & a_x & p_x \\ n_y & o_y & a_y & p_y \\ n_z & o_z & a_z & p_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{12} $$

$$ \left.\begin{aligned} n_x &= c_1[c_{23}(c_4c_5c_6 - s_4s_6) - s_{23}s_5c_6] + s_1(s_4c_5c_6 + c_4s_6), \\ n_y &= s_1[c_{23}(c_4c_5c_6 - s_4s_6) - s_{23}s_5c_6] - c_1(s_4c_5c_6 + c_4s_6), \\ n_z &= -s_{23}(c_4c_5c_6 - s_4s_6) - c_{23}s_5c_6, \\ o_x &= c_1[c_{23}(-c_4c_5s_6 - s_4c_6) + s_{23}s_5s_6] + s_1(c_4c_6 - s_4c_5s_6), \\ o_y &= s_1[c_{23}(-c_4c_5s_6 - s_4c_6) + s_{23}s_5s_6] - c_1(c_4c_6 - s_4c_5s_6), \\ o_z &= -s_{23}(-c_4c_5c_6 - s_4s_6) + c_{23}s_5s_6, \\ a_x &= -c_1(c_{23}c_4s_5 + s_{23}c_5) - s_1s_4s_5, \\ a_y &= -s_1(c_{23}c_4s_5 + s_{23}c_5) + c_1s_4s_5, \\ a_z &= s_{23}c_4s_5 - c_{23}c_5, \\ p_x &= c_1[a_2c_2 + a_3c_{23} - d_4s_{23}] - d_2s_1, \\ p_y &= s_1[a_2c_2 + a_3c_{23} - d_4s_{23}] + d_2c_1, \\ p_z &= -a_3s_{23} - a_2s_2 - d_4c_{23}, \end{aligned}\right\} $$

In this paper, Matlab and Robotics Toolbox are chosen as the simulating tools. The Robotics Toolbox supports the D-H method, and is based on the kinematics and dynamics of the robot.



(a) (b)

Fig. 4. Polar lines and pole

If Harris threshold is set to 500, searching field is equal to 11, and the biggest distance is set to 50, so the result corners are extracted from two images using image processing toolbox. There are 186 corners in the left Fig., and 156 corners in the right Fig., then these points are matched, at last 69 matched pairs are gotten, as is shown in Fig.4.

In this paper, the fundamental matrix is estimated through computing the 69 pair matched points. Points are segmented into inner and outer ones. Here, the threshold used to judge the outer points is 0.002, and the fundamental matrix is estimated using image processing toolbox. And the result is shown as follow:

$$F_{\max} = \begin{bmatrix} 0 & 0 & -6.0729e-004 \\ 0 & 0 & -3.4289e-003 \\ 4.2054e-004 & 3.23473e-003 & 2.1948e-001 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & -0.0006 \\ 0 & 0 & -0.0034 \\ 0.0004 & 0.0032 & -0.2195 \end{bmatrix}$$

According to the equation(2), the epipolar lines in which corresponding points lay are computed, and coordinates of the pixels in two images are continuously computed, so the result in table 2 is gotten.

In order to improve the computing efficiency, the reference frame of the end-effector coincides with frame of camera. The initial coordinate is set to (-0.5 -1 0.5), and the

Table 2. The polar coordinates of the estimated points

polar coordinate		RANSAC
Fig.2(a)	u	882.2636029159
	v	41.96440254536
Fig.2(b)	u	962.7893031281
	v	60.777841952522

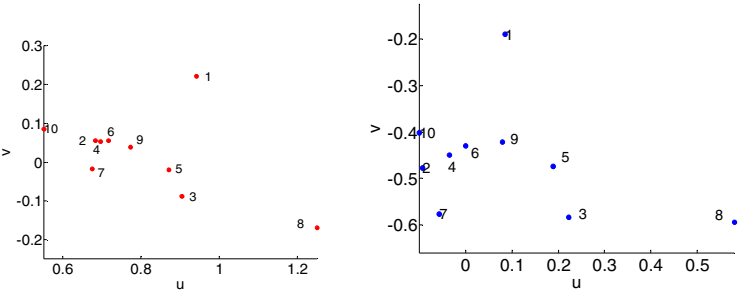


Fig. 5. (a) Feature points in initialization (b) Feature points in target

initial pose is set as (20 -15 15). So the homogeneous coordinates of the initial and objecttive pose is as shown in follow:

$$T_0^c = \begin{bmatrix} 0.933 & -0.3287 & -0.1464 & -0.5 \\ 0.2588 & 0.3304 & 0.9077 & -1 \\ -0.25 & -0.8848 & 0.3933 & 0.5 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad T_0^d = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Feature points are found in the vision scene and lay in the field of vision of the camera. Here, ten points gotten randomly and is shown in Fig. 5. The left camera expresses the initial state and the right camera expresses the target pose.

Here, some inner parameters of the camera are known, such as f=1 u_0=0 v_0=0. Now the camera is looked as perfect. The systemic gain is set 0.1.

In Fig.6(a), all epipolar lines intersect one point. And in Fig.6(b), the feature points move to the target and moving trajectories are shown. All feature points first move to epipolar line by adjusting pose of camera, once arrive the epipolar line, the camera moves parallel along the polar line to the target step by step. And error is very small.

The Fig. 7 describes change of the moving parameter during camera motion. The camera first adjusts the pose. And the changes of angle velocities in three directions only are very small after the former 50 times iteration. Then the camera moves

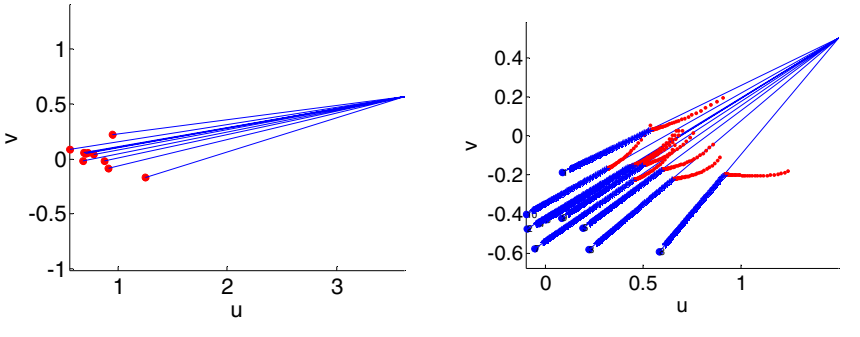


Fig. 6. (a) Feature points and epipolar lines in initialization (b) Moving routine of feature points

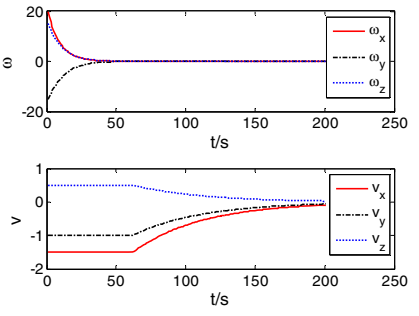


Fig. 7. Change curve of moving parameter

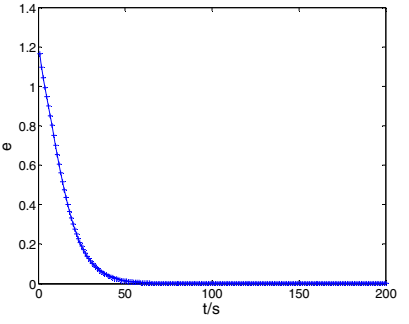


Fig. 8. Curve of error function

parallelly, and the angle velocities in three directions is almost equation to zero, and changing curve of the velocity is in evidence. The Fig.8 is the curve about error function. e expresses the geometrical average value of the coordinate error of ten feature points. From the Fig. 8, we know that the feature points lay all along in the polar lines when the camera moves parallel and the error equal approximately to 0.

5 Conclusions

Aiming to the complexity of the visual servo control problem, one visual servo algorithm based on epipolar geometry and IBVS is presented. Epipolar geometry depicts corresponding points constrain in images and is not affected by outside environment. IBVS can estimate the relative pose of the target with respect to initialization and compute the moving trajectories of end-effector in real-time constraint. In this paper, one visual servo control system for Puma560 is constructed and the simulating experiment is conducted. And the experimental results demonstrate efficiency of the proposed method.

Acknowledgments. The work is supported by the National Nature Science Foundation of China (Grant No 50875100), the National Key Technology R&D Program of China (Grant No 2008BAI50B04) and '863' Hi-Tech Research and Development Program of China (Grant No 2007AA04Z204).

References

1. Hutchinson, S.: Visual servo control, Part II: Advanced approaches. IEEE Robotics and Automation Magazine 14(1), 109–118 (2007)
2. Chaumette, F., Hutchinson, S.: Visual servo control, Part I: Basic approaches. IEEE Robotics and Automation Magazine 13(4), 82–90 (2006)
3. Paulo, A., Paris, C.: Uncalibrated Visual Servoing in 3D workspace. In: Campilho, A., Kamel, M.S. (eds.) ICIAR 2006. LNCS, vol. 4142, pp. 225–236. Springer, Heidelberg (2006)
4. Benhimane, S., Malis, E.: Homography-based 2d visual servoing. In: IEEE ICRA 2006, Orlando, Fl (May 2006)
5. Xiong, Y.: Robotics technology (wuhan). publishing house of Huazhong University of Science and Technology (2004)

Obtaining Reliable Depth Maps for Robotic Applications from a Quad-Camera System

Lazaros Nalpantidis, Dimitrios Chrysostomou, and Antonios Gasteratos

Laboratory of Robotics and Automation
Department of Production and Management Engineering
Democritus University of Thrace
University Campus Kimmeria
GR-671 00, Xanthi, Greece
{lanalpa,dchrisos,agaster}@pme.duth.gr
http://robotics.pme.duth.gr

Abstract. Autonomous navigation behaviors in robotics often require reliable depth maps. The use of vision sensors is the most popular choice in such tasks. On the other hand, accurate vision-based depth computing methods suffer from long execution times. This paper proposes a novel quad-camera based system able to calculate fast and accurately a single depth map of a scenery. The four cameras are placed on the corners of a square. Thus, three, differently oriented, stereo pairs result when considering a single reference image (namely an horizontal, a vertical and a diagonal pair). The proposed system utilizes a custom tailored, simple, rapidly executed stereo correspondence algorithm applied to each stereo pair. This way, the computational load is kept within reasonable limits. A reliability measure is used in order to validate each point of the resulting disparity maps. Finally, the three disparity maps are fused together according to their reliabilities. The maximum reliability is chosen for every pixel. The final output of the proposed system is a highly reliable depth map which can be used for higher level robotic behaviors.

Keywords: Stereo vision, quad-camera system, disparity maps fusion.

1 Introduction

Reliable depth estimation is commonly needed in robotics in order to obtain numerous autonomous behaviors. Autonomous navigation [1], localization and mapping are just a few of them [2,3]. Vision-based solutions are becoming more and more attractive due to their decreasing cost as well as their inherent coherence with human imposed mechanisms. However, robotic applications place strict requirements on the demanded speed and accuracy of vision depth-computing algorithms. Depth estimation using stereo vision, comprises the stereo correspondence problem. Stereo correspondence is known to be very computational demanding. The computation of dense and accurate depth images, i.e. disparity maps, in frame rates suitable for robotic applications is an open problem for

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 906–916, 2009.

the scientific community. Most of the attempts to confront the demand for accuracy focus on the development of sophisticated stereo correspondence algorithms, which usually increase the computational load exponentially. On the other hand, the need for real-time frame rates, inevitably, imposes compromises concerning the quality of the results. However, results' reliability is of crucial importance for autonomous robotic applications. This work proposes a quad-camera, depth estimation system able to produce reliable results while maintaining the computational load within acceptable limits.

1.1 Related Work

Autonomous robots behavior greatly depends on the accuracy of their decision making algorithms. In the case of stereo vision-based navigation, the accuracy and the refresh rate of the computed disparity maps are the cornerstone of its success [4]. Dense local stereo correspondence methods calculate depth for almost every pixel of the scenery, talking into consideration only a small neighborhood of pixels each time [5]. On the other hand, global methods are significantly more accurate but at the same time more computationally demanding, as they account for the whole image [6]. However, since the most urgent constraint in autonomous robotics is the real-time operation, such applications usually utilize local algorithms.

Muhlmann et al. in [7] describe a method that uses the sum of absolute differences (SAD) correlation measure for RGB color images. Applying a left to right consistency check, the uniqueness constraint and a median filter, it can achieve 20 fps for 160x120 pixel images. Another fast SAD based algorithm is presented in [8]. It is based on the uniqueness constraint and rejects previous matches as soon as better ones are detected. It achieves 39.59 fps speed for 320x240 pixel images with 16 disparity levels and the root mean square error for the standard Tsukuba pair is 5.77 . The algorithm reported in [9] achieves almost real-time performance. It is once more based on SAD but the correlation window size is adaptively chosen for each region of the picture. Apart from that, a left to right consistency check and a median filter are utilized. The algorithm is able to compute 7 fps for 320x240 pixel images with 32 disparity levels. A detailed taxonomy and presentation of dense stereo correspondence algorithms can be found in [5]. Additionally, the recent advances in the field as well as the aspect of hardware implementable stereo algorithms are covered in [10].

Early previous work focused on developing stereo algorithms mostly for binocular camera configurations. More recently, however, due to significant boost of the available computational power, vision systems using multiple cameras are becoming increasingly feasible and practical. The transition from binocular to multi-ocular systems has the advantage of potentially increasing the stability and accuracy of depth calculations. The continuous price-reduction of vision sensors allowed the development of multiple camera arrays ready for use in many applications. For instance, Yang et al. [11] used a five-camera system for real-time rendering using modern graphics hardware, while Schirmacher et al. [12] increased the number of cameras and built up a six-camera system for on-the-fly

processing of generalized Lumigraphs. Moreover, developers of camera arrays have expanded their systems so as to use tens of cameras, such as the MIT distributed light field camera [13] and the Stanford multi-camera array [14]. These systems are using 64, and 128 cameras respectively.

Most of the aforementioned camera arrays are utilized for real-time image rendering. On the other hand, a research area that could also be benefited by the use of multiple camera arrays is the so called cooperative stereo vision; i.e., multiple stereo pairs being considered to improve the overall depth estimation results. To this end, Zitnick [15] proposed an algorithm for binocular occlusion detection and Mingxiang [16] expanded it to trinocular stereo. The proposed system goes a step further using, for the first time, four cameras and exploiting the concept of the diagonal disparity.

2 Proposed System

The proposed system is a combination of sensory hardware and a custom-tailored software algorithm. The hardware configuration, i.e. the four cameras' formation, produce three stereo image pairs. Each pair is submitted to the simple and rapid stereo correspondence algorithm, resulting, thus, in a disparity map. For each disparity map a certainty map is calculated, indicating each pixel's reliability. Finally, the three disparity maps are fused, according to their certainties for each pixel. The outcome is a single disparity map which incorporates the best parts of its producing disparity maps. The combined hardware and software system is able to produce accurate dense depth maps in frame rate suitable for autonomous robotic applications.

2.1 Sensory Hardware Requirements

The sensory configuration of the proposed system consists of four identical cameras. The four cameras are placed so as their optical axes to have parallel orientation and their principal points to be co-planar, residing on the corners of the same square, as shown in Fig. 1(a). The images captured by the upper-left camera are considered as the reference images of each tetrad. Each one of the other three cameras produces images to be corresponded to the reference images. Thus, for each tetrad of images three, differently oriented, stereo pairs result, i.e. an horizontal, a vertical and a diagonal one. The concept, as well as the result of such a group of cameras are presented in Fig. 1(b).

2.2 Software Architecture

The proposed algorithm consists of two processing steps. The first one is the stereo correspondence algorithm that is applied to each image pair. Then, during a fusion step the results for all the stereo pairs are merged.

(a) (b)

Fig. 1. (a) The Quad-camera Configuration and (b) the Results (Up-Left) and Scene Capturing (Right) Using a Quad-camera Configuration

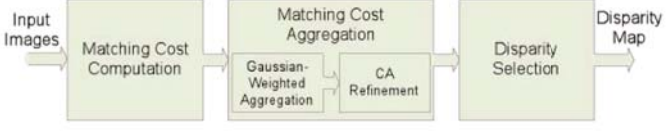


Fig. 2. Block Diagram of the Utilized Stereo Correspondence Algorithm

Stereo Correspondence Algorithm. The proposed system utilizes a custom tailored, simple, rapidly executed stereo correspondence algorithm applied to each stereo pair. Stereo disparity is computed using a three-stage local stereo correspondence algorithm. The algorithm utilized is an up-to-date version of the algorithm presented in [17]. It combines low computational complexity with sophisticated data processing. Consequently, it is able to produce dense disparity maps of good quality in frame rates suitable for robotic applications. The structural elements of this algorithm are presented in Fig. 2. The main attribute that differentiates this algorithm from the majority of the other ones is that the matching cost aggregation step consists of two sophisticated sub-steps rather than one simple summation. In addition, the disparity selection process is a non-iterative one and for speed reasons the final refinement step is absent. The results' refinement step is moved inside the aggregation, rather than being an additional final procedure. Instead of refining the results that were chosen through a strict selection process, the proposed algorithm performs a refinement procedure to all the available data. Such a procedure enhances the quality of the results. Thus, the disparity selection step can remain a simple winner-takes-all (WTA) choice. The absence of an iteratively updated selection process significantly reduces the computational payload of this step.

The matching cost function utilized is the absolute differences (AD). AD is inherently the simplest metric of all, involving only summations and finding absolute values. Disparity space image (DSI) is a 3D matrix containing the

computed matching costs for every pixel and for all its potential disparity values. The DSI values for constant disparity value are aggregated inside fix-sized square windows. The dimensions of the chosen aggregation window play an important role in the quality of the final result. Generally, small dimensions preserve details but suffer from noise. On the contrast, large dimensions may not preserve fine details but significantly suppress the noise. Noise suppression is very important for stereo algorithms that are intended to be applied to outdoors scenes. Outdoors images, which is often the case for autonomous navigation tasks, usually suffer from noise induced by a variety of reasons, e.g. lighting differences and reflections. The aggregation windows dimensions used in the proposed algorithm are 13x13 pixels. This choice is a compromise between real-time execution speed and noise cancellation. The AD aggregation step of the proposed algorithm is a weighted summation. Each pixel is assigned a weight depending on its Euclidean distance from the central pixel. A 2D Gaussian function determines the weights value for each pixel. The center of the function coincides with the central pixel. The standard deviation is equal to the one third of the distance from the central pixel to the nearest window-border. The applied weighting function can be calculated once and then be applied to all the aggregation windows without any further change. Thus, the computational load of this procedure is kept within reasonable limits. The DSI values after the aggregation are furthered refined by applying 3D cellular automata (CA). Two CA transition rules are applied to the DSI. The values of parameters used by them were determined after extensive testing to perform best. The first rule attempts to resolve disparity ambiguities. It checks for excessive consistency of results along the disparity axis and, if necessary, corrects on the perpendicular plane. The second rule is used in order to smoothen the results and at the same time to preserve the details on constant-disparity planes. The two rules are applied once. Their outcome comprises the enhanced DSI that will be used in order the optimum disparity map to be chosen by a simple, non-iterative WTA step.

Reliability-Based Fusion Algorithm. The output of the utilized custom stereo correspondence algorithm for each image pair is not only a disparity map but a certainty map as well. That is, for every pixel of the disparity map (depth metric for the corresponding pixel in the reference image) a certainty measure is calculated indicating the likelihood of the pixel's selected disparity value to be the right one. The certainty measure is calculated for each pixel (x, y) as can be seen in (1).

$$cert(x,y) = \left| SAD(x,y,D) - \frac{\sum_{z=0}^{d-1} SAD(x,y,z)}{d} \right| \tag{1}$$

According to this, the certainty *cert* for a pixel (x, y) that the computed disparity value D is actually right is equal to the absolute value of the difference between the minimum matching cost value $SAD(x, y, D)$ and the average matching cost

value for that pixel when considering all the d candidate disparity levels for that pixel. What the aforementioned measure evaluates is the amount of differentiation of the selected disparity value with regard to the rest candidate ones. The more the disparity value is differentiated, the most possible it is that the selected minimum is actually a real one and not owed to noise or other effects. The certainty calculation for all the pixels leads to a greyscale certainty map as shown in the third row of Fig. 3 and Fig. 5. In these images the lighter value of a pixel indicates greater certainty about its computed disparity value.

For each pixel, the disparity value having the largest certainty is chosen, among the three candidates. This technique, gets the best parts of each one of the three preliminary disparity maps and fuses them in one final disparity map. The resulting final depth map is obviously more reliable than any of its producing ones.

3 Experimental Results and Applications

The proposed quad-camera algorithm has been applied to two tetrads of images in order to evaluate its performance.

The most common image set for evaluating stereo correspondence algorithms is the Tsukuba data set. While typical stereo algorithms require two horizontally displaced pictures of this set, the proposed method requires four images. The Tsukuba data set consists of multiple images of the scene captured by a camera grid with multiple, equally spaced horizontal and vertical steps. The choice of the images captured by four cameras belonging to a square perfectly satisfies the demands imposed by the proposed algorithm. Figure 3 depicts the reference, i.e. up-left image, in the first column and the three target images, i.e. up-right, down-left and down-right in the second column. The third and the fourth columns show the certainty and disparity maps calculated for the image pairs consisting of the single reference image and the corresponding target ones. The fifth column of the figure shows the fused final disparity map on the top and the ground truth disparity map on the bottom.

Figure 4 shows the experimental results of the proposed quad-camera algorithm (left), the computationally equivalent simple stereo algorithm (middle) and the utilized single stereo algorithm applied on the horizontal stereo pair (right). The first row shows the calculated disparity maps. The second row shows the maps of pixels with absolute computed disparity error bigger than 1 shown in black. Finally, the third row presents maps of signed disparity error where the middle (50%) gray tone equals to zero error. It is obvious that the simple stereo algorithm, shown in the rightmost column suffers from noise. The usual confrontation of this issue is to enlarge the utilized 13x13 pixel aggregation window during the respective stage. However, window enlargement generally leads to loss of detail and coarse results, as shown in the middle column. This version of the algorithm utilizes a 23x23 pixel aggregation window, which results in triple computational load. Obviously, both of these treatments lack the results' quality of the proposed method. The final result of the proposed algorithm requires

Fig. 3. Algorithm's Steps and Results for the Tsukuba Data Set. (Column 1) the Reference Image (Up-Left), (Column 2) the Three Target Images (Up-Right, Down-Left, Down-Right), (Column 3) the Certainty Maps for the Horizontal, Vertical and Diagonal Pair, (Column 4) the Computed Disparity Map for Each Stereo Pair, (Column 5) the Fused (Top) and the Ground Truth (Bottom) Disparity Maps.

roughly the same computational power as the algorithm in the middle column. The outcome is that the proposed quad-camera algorithm achieves better results than its computationally equivalent simple two-camera stereo counterpart and the simple initial stereo algorithm.

The percentage of pixels whose absolute disparity error is greater than 1 in the non-occluded, all, and near discontinuities and occluded regions are presented in Table 1. The presented percentages refer to the three initially computed stereo pairs (namely the horizontal, vertical and diagonal pair), the final fused result of the proposed system and, finally, the computationally equivalent two-camera utilized stereo correspondence algorithm.

As shown in Table 1 there are cases where the results of the fusion process are marginally worse than those of an initial step. However, the image pair direction that provides the optimum results and should be considered as the most reliable and useful can not be initially anticipated. Moreover, the optimum direction is arbitrary and, therefore, there is little chance to coincide with any of the available three in the proposed system throughout the whole scene. However, the goal of the fusion system is to identify the best disparity value for every pixel. Thus, the results will be roughly as, or occasionally even more, accurate as the best initial results. On the other hand, the final disparity map is, in any case, far more reliable than the initial ones, since it has gone through a validation procedure, guaranteed by (1).

The proposed algorithm has been also applied to a virtual scenery. A virtual quad-camera system was inserted to the virtual room shown in the two first columns of Fig. 5 and the demanded tetrad of images was captured. The room

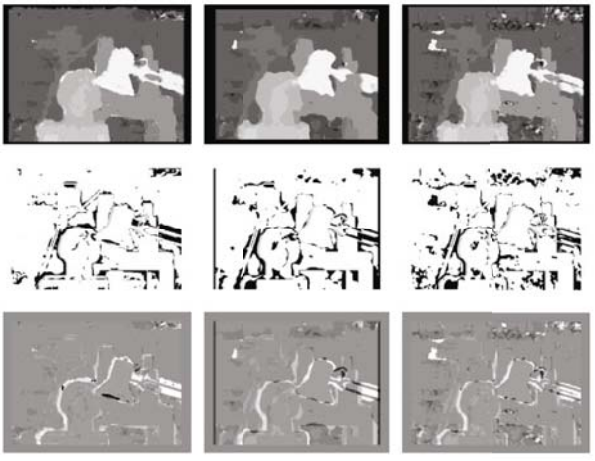


Fig. 4. Results of the Proposed Fusion System (Left), the of the Computationally Equivalent Simple Stereo Algorithm (Middle) and the Preliminary Simple Stereo Algorithm Applied on the Horizontal Image Pair (Right). From Top to Bottom: the Computed Disparity Maps, Pixels with Absolute Computed Disparity Error Bigger than 1 and Maps of Signed Disparity Error.

Table 1. Percentage of Pixels whose Absolute Disparity Error is Greater than 1 in Various Regions for the Tsukuba Pairs

Pair	**Non-occluded (%)**	**All (%)**	**Discontinuities (%)**
Horizontal	16.2	18.1	29.9
Vertical	12.5	13.8	35.1
Diagonal	10.7	12.4	32.3
Proposed	**10.8**	**12.6**	**31.5**
Equivalent	15.8	17.6	33.9

scene was chosen as it is a complex and demanding one, having both regions with fine details and low-textured ones. Moreover, the repetitive pattern of the books, in the background, is a challenging element for the stereo correspondence algorithms. Figure 5 depicts the reference i.e. up-left image in the first column and the three target images i.e. up-right, down-left and down-right in the second column. The third and the fourth columns show the certainty and disparity maps calculated for the image pairs consisting of the single reference image and the corresponding target ones. Finally, the fifth column of the figure shows the fused final disparity map.

The availability of reliable depth maps is the cornerstone of many computer vision as well as robotic applications. Figure 6(a) shows a screenshot of the 3D reconstructed Tsukuba scene. The depth map of Fig. 3 obtained using the

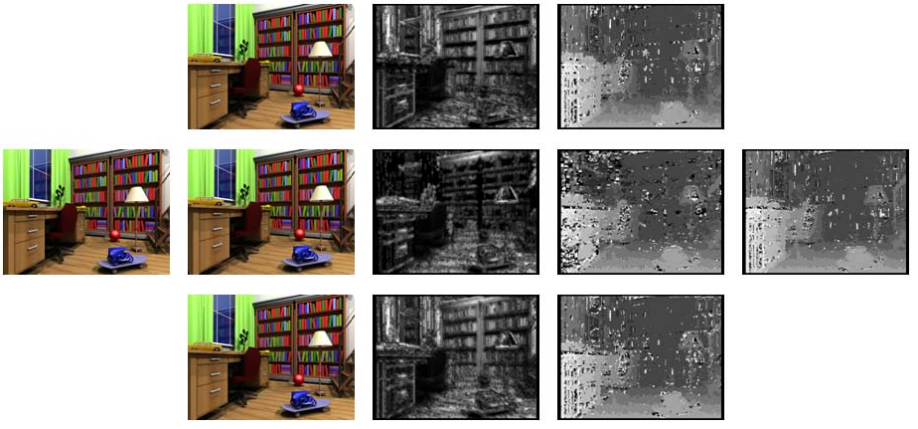


Fig. 5. Algorithm's Steps and Results for a Synthetic Room Scene. (Column 1) The Reference Image (Up-Left), (Column 2) the Three Target Images (Up-Right, Down-Left, Down-Right), (Column 3) the Certainty Maps for the Horizontal, Vertical and Diagonal Pair, (Column 4) the Computed Disparity Map for Each Stereo Pair, (Column 5) the Final Fused Depth Map.

(a) (b)

Fig. 6. Application Results Obtained Using the Caluclated Depth Maps. (a) View of the Reconstructed Tsukuba Scene and (b) Obstacle Detection in the Virtual Room Scene.

proposed method was utilized in order to add the third dimension's information to the reference image. Thus, a 3D model of the scene was reconstructed and a computer user can virtually navigate around the scene. On the other hand, Fig. 6(b) shows an obstacle detection application based on the availability of reliable depth map. Stereo vision can be used by autonomous robotic platforms in order to reliably detect obstacles within their movement range and move accordingly. The previously obtained depth map of Fig. 5 was used for the calculation of the v-disparity image. Using the Hough transformation the floor plane

was calculated and the obstacles were detected. This result is useful for any path-planning algorithm.

4 Conclusions and Discussion

In this work a depth computing system has been proposed aimed for autonomous robotics applications. The system utilizes a square formation of four identical cameras capturing the same scene. Selecting one of the images of each tetrad as reference, three image pairs result. Each pair is processed by a simple and rapid, custom stereo correspondence algorithm which results in an initial disparity map, as well as a certainty map. A fusion process evaluates the three initial disparity maps according to their certainty and produces the final combined disparity map.

Autonomous robotic applications demand reliable depth estimations obtained in real-time frame rates, having at the same time limited computational resources. The proposed system substitutes the computational complexity with a special sensor configuration. However, the demanded configuration can be easily and cost efficiently be achieved. The presented results exhibit a fair compromise between the objectives of low computational complexity and result's reliability.

The accuracy of local algorithms in various regions of a scene is strongly correlated to the orientation of the depicted objects in that particular region towards the orientation of the correspondence search procedure. That is, the depth discontinuities are more discriminable when they are oriented vertically to the correspondence search direction. This conclusion is based on the inherent way local algorithms operate and can be confirmed by the preliminary disparity maps, presented in the fourth row of Fig.3 and Fig. 5. The proposed system has the advantage of being able to adapt to various objects' orientations. The result is that the final fused disparity map is at least as accurate as the most accurate of the initial disparity maps and at the same time much more reliable than any of them. Moreover, the structure of the proposed software architecture is ideal for execution on the nowadays widely available quad-core processors. Each one of the identical but separate stereo correspondence searches can be assigned to a core, while the fourth core will supervise the whole procedure.

Acknowledgments. This work was supported by the E.C. funded research project for vision and chemiresistor equipped web-connected finding robots, "View-Finder", FP6-IST-2005-045541.

References

1. Hariyama, M., Takeuchi, T., Kameyama, M.: Reliable stereo matching for highly-safe intelligent vehicles and its vlsi implementation. In: IEEE Intelligent Vehicles Symposium, pp. 128–133 (2000)
2. Murray, D., Little, J.J.: Using real-time stereo vision for mobile robot navigation. Autonomous Robots 8(2), 161–171 (2000)

3. Sim, R., Little, J.J.: Autonomous vision-based robotic exploration and mapping using hybrid maps and particle filters. Image and Vision Computing 27(1-2), 167–177 (2009)
4. Schreer, O.: Stereo vision-based navigation in unknown indoor environment. In: Burkhardt, H.-J., Neumann, B. (eds.) ECCV 1998. LNCS, vol. 1406, pp. 203–217. Springer, Heidelberg (1998)
5. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International Journal of Computer Vision 47(1-3), 7–42 (2002)
6. Torra, P.H.S., Criminisi, A.: Dense stereo using pivoted dynamic programming. Image and Vision Computing 22(10), 795–806 (2004)
7. Muhlmann, K., Maier, D., Hesser, J., Manner, R.: Calculating dense disparity maps from color stereo images, an efficient implementation. International Journal of Computer Vision 47(1-3), 79–88 (2002)
8. Di Stefano, L., Marchionni, M., Mattoccia, S.: A fast area-based stereo matching algorithm. Image and Vision Computing 22(12), 983–1005 (2004)
9. Yoon, S., Park, S.K., Kang, S., Kwak, Y.K.: Fast correlation-based stereo matching with the reduction of systematic errors. Pattern Recognition Letters 26(14), 2221–2231 (2005)
10. Nalpantidis, L., Sirakoulis, G.C., Gasteratos, A.: Review of stereo vision algorithms: from software to hardware. International Journal of Optomechatronics 2(4), 435–462 (2008)
11. Ruigang, Y., Welch, G., Bishop, G.: Real-time consensus-based scene reconstruction using commodity graphics hardware. In: 10th Pacific Conference on Computer Graphics and Applications, pp. 225–234 (2002)
12. Schirmacher, H., Li, M., Seidel, H.P.: On-the-fly processing of generalized lumigraphs. In: EUROGRAPHICS, pp. 165–173 (2001)
13. Yang, J.C., Everett, M., Buehler, C., Mcmillan, L.: A real-time distributed light field camera. In: Eurographics Workshop on Rendering, pp. 77–86 (2002)
14. Wilburn, B., Smulski, M., Lee, K., Horowitz, M.A.: The light field video camera. In: Media Processors, pp. 29–36 (2002)
15. Zitnick, C.L., Kanade, T.: A cooperative algorithm for stereo matching and occlusion detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(7), 675–684 (2000)
16. Mingxiang, L., Yunde, J.: Trinocular cooperative stereo vision and occlusion detection. In: IEEE International Conference on Robotics and Biomimetics, pp. 1129–1133 (December 2006)
17. Nalpantidis, L., Sirakoulis, G.C., Gasteratos, A.: A dense stereo correspondence algorithm for hardware implementation with enhanced disparity selection. In: Darzentas, J., Vouros, G.A., Vosinakis, S., Arnellos, A. (eds.) SETN 2008. LNCS (LNAI), vol. 5138, pp. 365–370. Springer, Heidelberg (2008)

Surface Reconstruction of Engine Intake Ports with Mixed Constraints

Kun Mo and Zhouping Yin*

State Key Laboratory of Digital Manufacturing Equipment and Technology, Huazhong University of Science and Technology, Wuhan, 430074, P.R. China
Fax: 86-27-87559415
yinzhp@mail.hust.edu.cn

Abstract. This paper presents a novel algorithm called mixed constraints reconstruction (MCR) method for building the surface of engine intake ports from scanning point clouds. The key idea of MCR method is to incorporate the fluid flow as a global constraint to approximate the point clouds with implicit surface representation. By combining the distance field and the fluid velocity generated from point clouds as mixed constraints, a new minimal-like model and its variational level set equation are introduced. The fluid flow velocity is obtained by solving Navier-Stokes equation, where a method for confirming the boundary condition from an extended boundary is also proposed. This method can reconstruct the final surface more reasonably and smoothly, with good fluid flow testing results. It needs not any additional geometric information and mesh construction. In the end, an example from real case is given to demonstrate the effectiveness of this method.

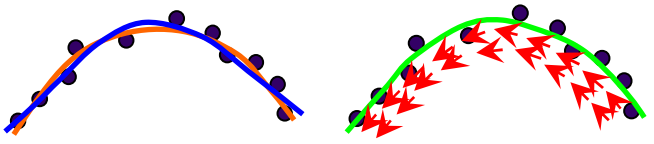
Keywords: Surface reconstruction, Mixed constraints, Engine intake ports, Level set equation, Fluid flow constraint.

1 Introduction

The surface modeling of the engine intake ports deserves much attention in automotive application due to their importance of multi-cylinder internal combustion engine. In most case, this process widely employs the surface reconstruction technique from scanning point clouds, because it is often hard to obtain or develop the surface model of computer representation. However, successful application of this technology still involves many practical problems. For instance, the point clouds by scanning device are often noisy and non-uniform. Due to the inaccessibility during the scanning, some holes and sparse zones exist in point clouds. With such point clouds, it is difficulty to fit the surface accurately and fill holes reasonably. In addition, it is more worth to mention that the surface of engine intake ports is not only a simple shape with geometric property, it is also restricted by certain surface style with physical performance. For instance, the surface of engine intake ports could be a stream surface [1, 2] which is defined according to required flow speed of the fuel under constant air pressure. A small shape changes can influence the fuel speed and

* Corresponding author.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 917–928, 2009.



(a) (b)

Fig. 1. The same point clouds can generate different result surfaces, only the one which both satisfy the geometric and physical constraints is the best result

lower the power of combustion engine. Therefore, for reconstructing such high quality surface, it should not only generate a high accurate geometric shape, but also attempt to keep its required physical attribution. Since there is no unique way to generate a resulting surface based on geometric information (ill-posed problem), it is not wise to only use the geometric standard to evaluate the surface model. As shown in Fig. 1 (a), the black circles represent the point clouds with noise. It can be reconstructed by many methods (represented as blue and red lines). Of course, they all satisfy the accuracy in geometric standard (such as minimal sum of the distance between each points and the surface). But only the one, which can also satisfy the physical constraints, is the optimal result. As shown in Fig.1 (b), the red arrows represent a fluid velocity field under required work conditions. The result surface (represented by green line) should not only reach the geometric accuracy but also be a stream surface whose tangent direction is accordance with the given velocity field. In real case of industry, this respect is often ignored in surface modeling. As a result, to guarantee the physical performance, the sub processes, such as analyzes by computational simulation or physical experiments, are require for modifying the shape of resulting surface.

The known surface reconstruction methods can be generally classified into two categories, explicit and implicit method. The most widely used explicit method is NURBS [3], which is the main method in many computer aided design (CAD) software of industry. But this process has to introduce some trimming and patching operations on a set of NURBS patches. Confirming the continuity between each patch is a tedious and artificial work. Some other explicit methods, such as Delaunay triangulations [4], alpha shapes [5] or Voronoi diagrams [6, 7], often get in trouble with the noisy point clouds. The resulting surface may need to be smoothed [8] or refit (e.g. [5]) in subsequent processing. In contrast, the implicit methods can more conveniently provide a watertight 2-manifold surface approximation by computing the iso-surfaces to best fit the data. But most of them need the oriented point clouds, such as radial basis functions (RBF) method [9] and multi-level partition of unity (MPU) method [10]. The normals of point clouds are required to solve the matrix of the ideal RBF and aslo is needed by MPU method to locally fit surface. The implicit surface can also be obtained by solving partial differential equation (PDE), such as Poisson method [11]. However, it also requires the normals of point sets. In real case, estimating the normals of point clouds, which are often noisy and contain defective parts, is an ill-posed problem. Some other methods can reconstruct surface without normals, such as the moving least squares (MLS) methods [12] and Voronoi based method [13]. The two methods provide the robust algorithms but they have to

construct a complex data structure and cost much computation time. Some methods use the signed distance function to represent the surface, such as the method proposed in [14]. The signed distance function can also be cumulated into volumetric grids [15]. But they are troubled with the noise or sparse point sets because the connection relationship of these areas is hard to confirm.

The level set method [16] has been recently introduced in the field of surface reconstruction. The main idea is to evolve the interface of the object by a velocity field until it reaches a certain configuration. Base on this theory, a novel implicit reconstruction algorithm, called mixed constraints reconstruction (MCR) method, is presented in this paper. It introduces a fluid flow velocity, govern by Navier-Stokes (N-S) equation, as a globally constraints to adjust the approximation to the point clouds with geometric constraint. We create a new minimal-surface-like model and derive its variation, where the mixed constraints actually are the combined velocity of corresponding level set equation. The boundary condition of N-S equation refers the practical test in real case, which can reflect the design instinct to a certain extent. A detailed process of confirming the condition is also proposed. Solving the level set equation is time-consuming, because a re-initialization step is requires after several time steps. Thus the numerical technique for level set method without re-initialization is extended. With MCR method, the resulting surface not only has the geometric accuracy but also is the stream surface according with the velocity field of the design model. It realizes a global smooth and guarantees the physical performance.

The main process and the details of MCR method are explained in section 2. Section 3 describes the numerical implementation. In section 4, an example from real case and its analyses is given. The conclusion of MCR method is summarized in section 5.

2 Introduction of MCR Method

The main process of MCR method is illustrated in Fig.2 and represented as 2D situation for expressing simply. Our goal is to reconstruct a water-tight and smooth implicit surface from a scanning point clouds with both geometric accuracy and fluid flow accordance. The process begins with point clouds (see Fig.2 (a)), which may contain holes (represented by a defective part in the corner) and noise. Two mixed constraints are used for reconstruction, the one is the distance function generated from point clouds (see Fig. 2 (b)), which is used to guarantee the geometric accuracy; the other is the fluid flow velocity (see Fig. 2 (d)) which is to confirm a global smoothness and accordance with fluid flow. To obtain this velocity, we need to confirm a boundary for solving N-S equation. It can be extended from distance field (see Fig. 2 (c)) and the practical fluid flow test data are adopted as the boundary condition. By mixing the two velocity field, we introduce a minimal-like model and deriving its variation as a PDE, which is then translated as a level set equation and the mixed constraints is actually its evolving velocity. For solving the PDE numerically, we extend a non re-initialization technique. Final, the resulting surface is visualized by marching cube method [17] (see Fig. 2 (e)).

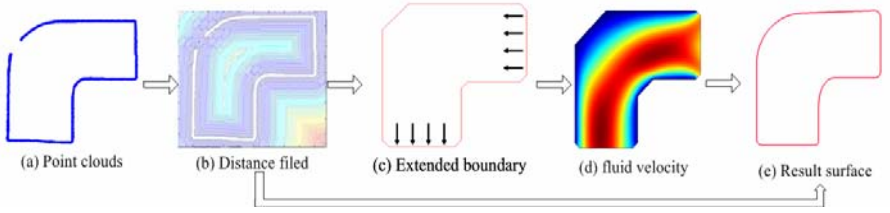


Fig. 2. The general process of MCR method

2.1 Basic Model and Equation

The reconstruction of engine intake port surface could be described as follow, given a set of distinct point clouds $S = \{(x_i, y_i, z_i), i = 1, 2, ...n\}$ representing an unknown surface M in R^3, find a surface M' that is a reasonable approximation to M and the tangent vector of M' is in accordance with the velocity direction of a fluid passing through. Let vector $\boldsymbol{x} = (x, y, z)$ denotes the notation of the point of surface M'. Define $d(\boldsymbol{x}) = dist(\boldsymbol{x}, S)$ the Euler distance function to S. To reconstruct a surface faithfully to the point clouds S, it is equally to obtain function $\Phi(\boldsymbol{x})$, which is the resulting surface minimizing the following object energy,

$$E(\Phi) = \lambda_1 \int_{\Phi} d^2(\boldsymbol{x}) ds + \lambda_2 \int_{\Phi} n \cdot V ds \tag{1}$$

where n is the unit normal vector of $\Phi(x)$ and $V = \dfrac{U}{|U|}$ is the unit velocity vector of fluid pass the surface which is computed by N-S equation,

$$\begin{cases} \dfrac{\partial U}{\partial T} + (U \cdot \nabla)U + \nabla p - \upsilon \nabla \cdot (\nabla U) = F & \text{in } \Phi \\ \text{div}(U) = 0 & \text{in } \Phi \\ p = p_0 & \text{on } \Phi \end{cases} \tag{2}$$

where $U(u, v, w, t)$ stands for the velocity with its x, y and z component respectively u, v and w. $p(x, y, z, t)$ donates the pressure of the fluid. υ and F is the kinematic viscosity of the fluid and the body force respectively. $p = p_0$ is the initial boundary condition which is set previously according to real case. The second term on the right side of Eq.(1) is the integral of the projection from normal vectors of surface to velocity vectors of fluid. Minimizing this term means that the normal and velocity are perpendicular each other and the tangent vector of Φ is in accordance with the fluid velocity. The coefficient $\lambda_1 + \lambda_2 = 1$ is the positive parameters which balances the influence of the geometric and velocity term. To obtain $\Phi(x)$, it is equally to solve the Euler-Lagrange equation of Eq.(1),

$$\lambda_1[2d(\boldsymbol{x})\nabla d(\boldsymbol{x})\cdot n + d^2(\boldsymbol{x})\kappa] + \lambda_2 \cdot \nabla V = 0 \tag{3}$$

where κ is mean curvature of Φ.

2.2 Variational Level Set Formulation

In most cases, it is impossible to solve Eq.(3) directly. As described in [16], looking for the resulting surface that minimizes Eq.(3) is similar to solve a time evolution paradigm, which can be described as the gradient flow,

$$\frac{\partial \Phi}{\partial t} = -(2\lambda_1 d(\boldsymbol{x})\nabla d(\boldsymbol{x})\cdot n + \lambda_1 d^2(\boldsymbol{x})\kappa + \lambda_2 \cdot \nabla V)n \tag{4}$$

The term $\nabla d(x)\cdot n$ in Eq.(4) describes the surface is attracted by the gradient of $d(\boldsymbol{x})$ and $d^2(\boldsymbol{x})\kappa$ is a smooth element for surface tension. The term ∇V represents the auxiliary constraints by the fluid flow. Since the method employs the implicit surface reconstruction, the surface $\Phi(\boldsymbol{x})$ can be represented as implicit function on volumetric region Ω,

$$\begin{cases} \Phi(\boldsymbol{x},t) > 0 \ \text{ for } \ \boldsymbol{x} \in \Omega \\ \Phi(\boldsymbol{x},t) < 0 \ \text{ for } \ \boldsymbol{x} \notin \Omega \\ \Phi(\boldsymbol{x},t) = 0 \ \text{ for } \ \boldsymbol{x} \in \Gamma = \partial\Omega \end{cases} \tag{5}$$

where the $\Gamma = \partial\Omega$ represent the interface of the region Ω, The local unit normal n and curvature κ to the surface are given by

$$n = \nabla\Phi / |\nabla\Phi| \ , \quad \kappa = div(\nabla\Phi / |\nabla\Phi| \ ) \tag{6}$$

Following the procedure introduced in [18], this problem can be transformed as a variational level set formulation, the energy model in Eq.(1) is expressed as

$$E(\Phi) = \lambda_1 \int_\Omega d^2(\boldsymbol{x})\delta(\Phi)|\nabla\Phi| d\boldsymbol{x} + \lambda_2 \int_\Omega n \cdot V \delta(\Phi)|\nabla\Phi| d\boldsymbol{x} \tag{7}$$

here $\delta(\Phi)$ is the Dirac delta distribution which is the differential of Heaviside function

$$H(\Phi) = \begin{cases} 1, & \Phi \geq 0 \\ 0, & \Phi < 0 \end{cases} \qquad \delta(\Phi) = \frac{dH(\Phi)}{d\Phi} \tag{8}$$

The corresponding gradient flow for $E(\Phi)$ is

$$\frac{\partial \Phi}{\partial t} = \lambda_1(2d(\boldsymbol{x})\nabla d(\boldsymbol{x})\cdot n + d^2(\boldsymbol{x})\kappa)\delta(\Phi) + \lambda_2 \cdot \nabla V \cdot \delta(\Phi) \tag{9}$$

As usual by replacing $\delta(\Phi)$ with $|\nabla\Phi|$ to extend this geometric motion other level sets, Eq.(9) is expanded as a level set equation,

$$\frac{\partial \Phi}{\partial t} = (\lambda_1(2d(\boldsymbol{x})\nabla d(\boldsymbol{x}) \cdot \nabla\Phi / |\nabla\Phi| + d^2(\boldsymbol{x}) div(\nabla\Phi / |\nabla\Phi|)) + \lambda_2 \cdot \nabla(U / |U|))|\nabla\Phi| = v(\boldsymbol{x}, t) \cdot |\nabla\Phi| \tag{10}$$

U is the solution of Eq.(2). The combined constraints in Eq.(1) are derived as a mixed velocity field in Eq.(10) of level set equation.

In classic level set method, the re-initialization step is a necessary operation which can guarantee the level set function always to be a signed distance function. Practically this step is usually finished by solving the following PDE [19]:

$$\Phi_t + S(\Phi_0)(|\nabla\Phi| - 1) = 0 \tag{11}$$

where the Φ_0 is the function to be re-initialized, and the $S(\Phi_0)$ is the sign function. But unfortunately, this step is not always useful when the level set function is far away from a signed distance function. In addition, it is also a time-consuming step. For accelerating the level set evolution, we extend a non re-initialization step proposed in [20], where a regular term is added to characterize how close a function Φ is to a signed distance function in 2d space. The extended regular term is expressed as the following integral

$$E_r(\Phi) = \int_\Omega \frac{1}{2}(|\nabla\Phi| - 1)^2 d\boldsymbol{x} \tag{12}$$

Similarly, its variational level set equation is

$$\frac{\partial \Phi}{\partial t} = div((1 - \frac{1}{|\nabla\Phi|})\nabla\Phi) \tag{13}$$

adding Eq.(13) into Eq.(10), the final level set equation is

$$\frac{\partial \Phi}{\partial t} = (\lambda_1(2d(\boldsymbol{x})\nabla d(\boldsymbol{x}) \cdot \nabla\Phi / |\nabla\Phi| + d^2(\boldsymbol{x}) div(\nabla\Phi / |\nabla\Phi|)) + \lambda_2 \cdot \nabla(U / |U|))|\nabla\Phi| + \xi \cdot (\Delta\Phi - \kappa) \tag{14}$$

where ξ is a positive constant to adjust the penalizing of the deviation of Φ from a signed distance function during its evolution. Thus, in the evolving process, level set function Φ can keep as a signed distance function in a narrowband and the re-initialization procedure is not necessary.

3 Numerical Implementation

3.1 Discretization of Equation

For discretizing Eq.(14), the up-wind scheme described in [21] is used. Since the direction of ∇V is not along the normal direction of the boundary of Φ, $\nabla V \cdot \nabla\Phi(\boldsymbol{x}, t)$ is discrete as

$$\nabla V(\boldsymbol{x}) \cdot \nabla\Phi(\boldsymbol{x}, t) = \sum \max(V_i, 0) \cdot D^{-i} + \sum \min(V_i, 0) \cdot D^{+i} \tag{15}$$

where $V_i(i = x, y, z)$ represent the components of ∇V on the axis of coordinates. The term $d^2(\boldsymbol{x})\kappa$ is along the normal of the boundary, $\kappa \cdot \left|\nabla \Phi(\boldsymbol{x}, t)\right|$ is discreted as

$$\kappa \cdot \left|\nabla \Phi(\boldsymbol{x}, t)\right| = (\max(\kappa, 0) \cdot \nabla^+ + \min(\kappa, 0) \cdot \nabla^-) \tag{16}$$

where $\nabla^+ = \sqrt{\sum \max(D^{-i}, 0)^2 + \sum \min(D^{+i}, 0)^2}$,

$\nabla^- = \sqrt{\sum \min(D^{-i}, 0)^2 + \sum \max(D^{+i}, 0)^2}$,

and $D^{+i} = (d_{i+1} - d_i)/dh$, $D^{-i} = (d_i - d_{i-1})/d_i$, $i = x, y, z$.

The mean curvature κ can use center-difference scheme. The discrete time step size should satisfy the inequality $dt_{levelset} < \dfrac{h}{\max\{|d_x|, |d_y|, |d_z|\}}$.

3.2 Compute Distance Field and Set Initial Function

It is no need to compute the distance function $d(\boldsymbol{x})$ in the whole domain. Only a narrowband surround the point clouds is sufficient to evolve the level set function. In this paper, the fast marching method [22] is employed with a narrowband (width $H = 4h$, h is the grid size) in this paper. The details can refer to [21]. With this distance function in narrowband, the initial surface for level set equation can't be chosen any shape. The outer level contour of the distance function $d(\boldsymbol{x}) = \varepsilon$ that encloses the data set is a very good initial choice, since it is a most likely shell-like structure which has same topology geometry as the final surface and the computation cost is not expensive.

3.3 Confirm Fluid Flow Velocity

A grid-based N-S solver is employed to obtain the velocity field $U(u, v, w, t)$ of Eq.(2). It uses an accurately explicit second-order Runge-Kutta method and the pressure of in and out ports is set as constant from real case. The key of constructing fluid flow velocity in MCR method is to define its boundary condition. As shown in Fig.2 (c), the arrows is the direction of the fluid flowing, the pressure is set on the boundary of the ports. It can refer the requirement of the automotive car. Except the in and out ports, the other boundary is set as fixed boundary condition, where the convection is not exit. In theory, the velocity field should be update in every time step during the evolution of $\Phi(\boldsymbol{x}, t)$. But it would cost so much computational time. As the initial surface proposed in 3.3 is nearly a final resulting surface, the affection of the shape changes during evolving to the fluid velocity is not significant. The initial surface is set as a fixed boundary. From the numerical experiment in subsequence, it can be seen that the velocity by the initial surface is sufficient to adjust the resulting surface as a global constraint.

4 Numerical Example and Analyses

In this paper, MCR method has been implemented using Matlab 7 coupled Visual C++ on a PC with 2000 MHz CPU and 1 GByte memory. An engine intake ports model from an automotive car in real case is taken as the example (see Fig. 3). Fig. 3(a) is the real model of engine intake ports and Fig. 3(b) is the computer model of this port, which is used as a stander model for evaluating MCR method.

In first example, MCR method is applied on point clouds, sampling from computer model (Fig. 3 (b)) with normal density. Fig.4 shows this reconstruction process. The

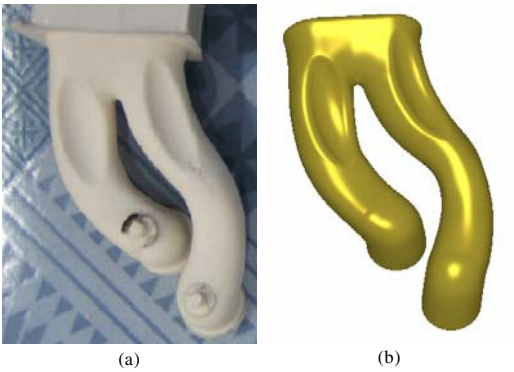


Fig. 3. Engine intake ports example. (a) The physical model. (b) The computer model.

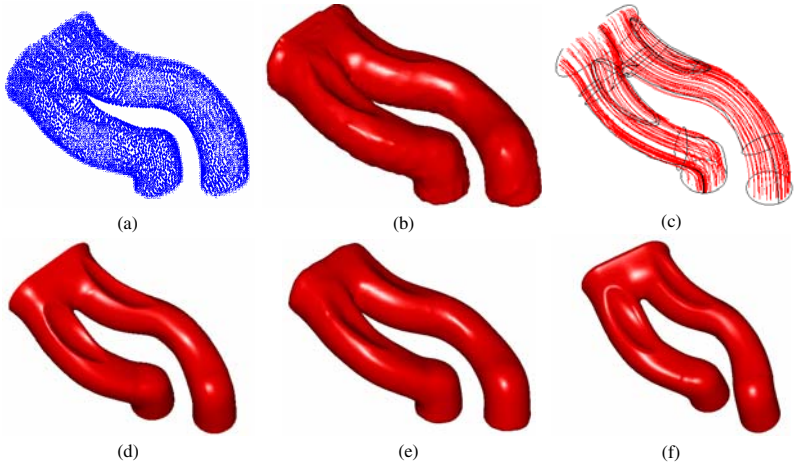


Fig. 4. (a) The point clouds. (b) The initial surfaces. (c) The streamlines derive from N-S equation. (d) The final surfaces with coefficient $\lambda_1 = 0.9, \lambda_2 = 0.1$. (e) The final surface with coefficient $\lambda_1 = 0.1, \lambda_2 = 0.9$. (f) The final surfaces with coefficient $\lambda_1 = 0.5, \lambda_2 = 0.5$.

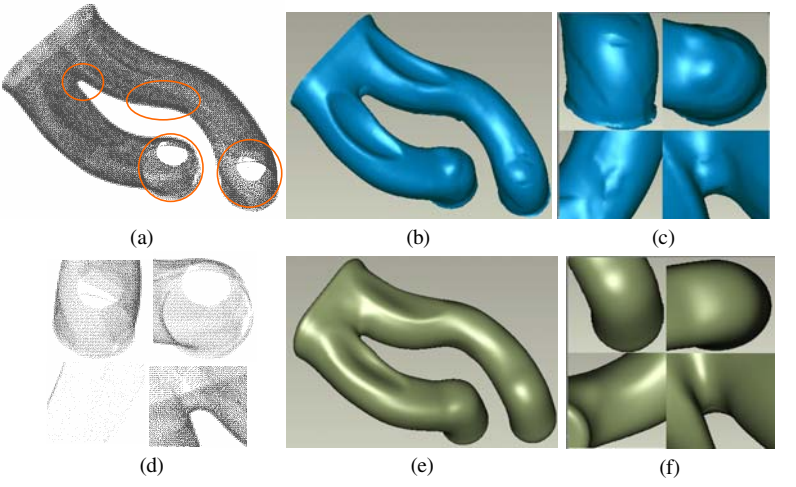
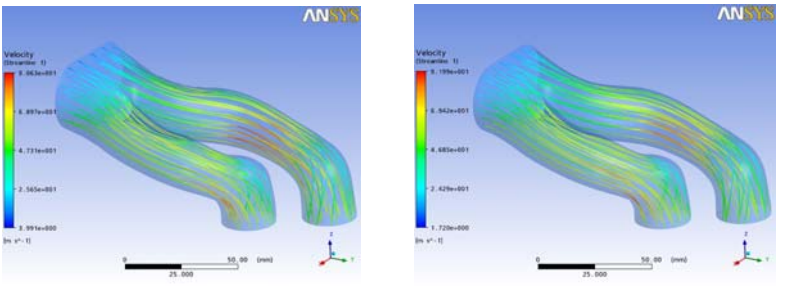


Fig. 5. (a) The point clouds. (b) The result surface by Zhao's method. (c) The details of (b). (d) The details of point clouds. (e) The surfaces by MCR method with coefficient $\lambda_1 = 0.9, \lambda_2 = 0.1$. (f) The details of (e).

number of point clouds (Fig.4 (a)) is 154890 and the volumetric grid is set as $94 \times 126 \times 84$ (with grid size $h = 0.5mm$). A constant pressure ($p = 1433 p_a$) is set on the initial surface (Fig. 4 (b)) and the velocity by solving Eq.(2) is shown in Fig.3 (c) as streamlines. The balance parameter, λ_1 and λ_2, should be choose according to the practical situation. When the point clouds are much noisy, λ_2 should be bigger than λ_1. We show the different result with series of coefficients λ_1 and λ_2 from Fig.3 (d) to Fig.3 (f) to demonstrate their effect to the final surface. The results demonstrate that the surface with bigger λ_2 is much smoother in each fillet zone. As this point clouds contains no noise, $\lambda_1 = \lambda_2 = 0.5$ is a better result (see Fig. 4 (f), the average error is 0.02mm).

MCR method can also handle the point clouds from real case with noise and defective parts. Fig.5 (a) shows the example of real model (see Fig.3 (a)) by scanning device. From the details (see Fig. (d)), it can be seen that the point clouds are noisy and overlapping due to the inaccuracy of sampling. This point clouds also contain holes in the zone for coupling air valve. A comparison between MCR method and the one propose in [16] is done to demonstrate the effective of fluid flow constraints. Fig. 5 (b) is the result by the method in [16] and its detail is shown in Fig. 5 (c). We can see the resulting surface is locally smooth but irregular in global regions and some jagged places exist in the final surface. The result surface and the details by MCR method is shown in Fig. 5 (d) and Fig. 5 (e) respectively. Because the point clouds contain noise and holes, thus, we have to use bigger λ_2 ($\lambda_1 = 0.1, \lambda_2 = 0.9$) to smooth the surface and fill the holes. Since the fluid flow velocity is a global factor and influenced not much by the noise and holes, thus, it can keep the whole surface smooth more reasonably.



(a) (b)

Fig. 6. (a) The velocity field of the original surfaces. (b) The velocity field of the surfaces reconstructed by MCR method ($\lambda_1 = 0.3, \lambda_2 = 0.7$).

Table 1. The comparison of fluid flow performance between original model and resulting shape

Surface Model	Maximum Speed (m/s)	Average Speed Rate(m/s)
Origianl Model	90.6325	52.8408
Resulting surface by MCR method	91.9871	53.2140

A comparison between the resulting surface by MCR method (Fig. 5 (c)) and the original surface model (Fig. 3 (b)) by the ANSYS software (Fig. 6) is also performed. Because the fluid flow constraint is set by the N-S equation, an ideal resulting surface by reconstruction should have the approximate fluid flow velocity as the original model. Thus, its physical performance can be guaranteed at least. The comparison is given in Table 1 (with $p = 1433 p_a$). From Table 1, we can see that the two surface has the nearly the same velocity property under the same condition.

5 Conclusion

In this paper, we propose MCR method based on implicit surface to reconstruct the surface of engine intake ports and apply this method on practical examples. Its main contribution is to incorporate the fluid velocity, referring the practical requirements, as a global constraint to fit the surface. It can smooth the resulting surface and fill holes more reasonably. In addition, the physical performance can also be guaranteed. From the examples, we can see the advantage of coupling fluid flow constraint, in contrast, the conventional reconstruct method, which only focus on the geometric information, can also be improved. As the fluid flow constraint is generated from N-S equation with physical property, the resulting surface by MCR method has nearly the same physical performance as the original one according to the comparison. In addition, to solve level set equation, we also extend the level set method without re-initialization technique from 2d to 3d space. This method can be extended straightforward into surface reconstruction of pipeline type surface or the surface with certain fluid performance, such as propeller blade of aerospace.

Acknowledgments

The work is supported by the Natural Science Fund of China (NSFC) (Project No. 50835004 and 50625516).

References

1. Hultquist, J.P.M.: Constructing stream surfaces in steady 3D vector fields. In: Proceedings of the 3rd conference on Visualization 1992. IEEE Computer Society Press, Boston (1992)
2. Jarke, J.v.W.: Implicit stream surfaces. In: Proceedings of the 4th conference on Visualization 1993. IEEE Computer Society, San Jose (1993)
3. Piegl, L., Tiller, W.: The Nurbs Book. Springer, Berlin (1997)
4. Jean-Daniel, B.: Geometric structures for three-dimensional shape representation. ACM Trans. Graph. 3, 266–286 (1984)
5. Chandrajit, L.B., Fausto, B., Guoliang, X.: Automatic reconstruction of surfaces and scalar fields from 3D scans. In: Proceedings of the 22nd annual conference on Computer graphics and interactive techniques. ACM, New York (1995)
6. Nina, A., Marshall, B., Manolis, K.: A new Voronoi-based surface reconstruction algorithm. In: Proceedings of the 25th annual conference on Computer graphics and interactive techniques. ACM, New York (1998)
7. Amenta, N., Choi, S., Kolluri, R.K.: The power crust, unions of balls, and the medial axis transform. Computational Geometry 19, 127–153 (2001)
8. Ravikrishna, K., Jonathan Richard, S., James, F.O.B.: Spectral surface reconstruction from noisy point clouds. In: Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing. ACM, Nice (2004)
9. Carr, J.C., Beatson, R.K., Cherrie, J.B., Mitchell, T.J., Fright, W.R., McCallum, B.C., Evans, T.R.: Reconstruction and representation of 3D objects with radial basis functions. In: Proceedings of the 28th annual conference on Computer graphics and interactive techniques. ACM, New York (2001)
10. Yutaka, O., Alexander, B., Marc, A., Greg, T., Hans-Peter, S.: Multi-level partition of unity implicit. ACM Trans. Graph. 22, 463–470 (2003)
11. Michael, K., Matthew, B., Hugues, H.: Poisson surface reconstruction. In: Proceedings of the fourth Eurographics symposium on Geometry processing. Eurographics Association, Cagliari (2006)
12. Marc, A., Johannes, B., Daniel, C.-O., Shachar, F., David, L., Claudio, T.S.: Point set surfaces. In: Proceedings of the conference on Visualization 2001. IEEE Computer Society, San Diego (2001)
13. Alliez, P., Cohen-Steiner, D., Tong, Y., Desbrun, M.: Voronoi-based variational reconstruction of unoriented point sets. In: Proceedings of the fifth Eurographics symposium on Geometry processing. Eurographics Association, Barcelona (2007)
14. Hugues, H., Tony, D., Tom, D., John, M., Werner, S.: Surface reconstruction from unorganized points. In: Proceedings of the 19th annual conference on Computer graphics and interactive techniques. ACM, New York (1992)
15. Brian, C., Marc, L.: A volumetric method for building complex models from range images. In: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques. ACM, New York (1996)

16. Zhao, H.-K., Osher, S., Merriman, B., Kang, M.: Implicit and Nonparametric Shape Reconstruction from Unorganized Data Using a Variational Level Set Method. Computer Vision and Image Understanding 80, 295–314 (2000)
17. Jin, G., Wang, Q., Shen, Y., Hao, J.: An improved marching cubes method for surface reconstruction of volume data, Dalian, China, pp. 10454–10457 (2006)
18. Zhao, H.-K., Chan, T., Merriman, B., Osher, S.: A Variational Level Set Approach to Multiphase Motion. Journal of Computational Physics 127, 179–195 (1996)
19. Osher, S., Sethian, J.A.: Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations. Journal of Computational Physics 79, 12–49 (1988)
20. Chunming, L., Chenyang, X., Changfeng, G., Fox, M.D.: Level set evolution without re-initialization: a new variational formulation. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005, vol. 1, pp. 430–436 (2005)
21. Osher, S., Fedkiw, R.: Level set methods and dynamic implicit surfaces. Springer, New York (2003)
22. Sethian, J.: A Fast Marching Level Set Method for Monotonically Advancing Fronts. Proceedings of the National Academy of Sciences of the United States of America 93, 1591–1595 (1996)

Binocular Based Moving Target Tracking for Mobile Robot

Yingkui Du[1,2], Baojie Fan[1,2], Jianda Han[1], and Yandong Tang[1]

[1] State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Science, 110016 Shenyang, China
[2] Graduate School of the Chinese Academy of Science, 100039 Beijing, China
{dyk,fanbaojie,jdhan,ytang}@sia.cn

Abstract. Moving target tracking is an important application of computer vision. A binocular based method is presented for mobile robot to track target reliably under the effect of occlusion, transform and rotation in unstructured environment. Point features are extracted for representing the target and environment background under middle distortion, and then are matched and tracked through consecutive stereo frames by our improved MNCC algorithm. The point features are reconstructed and utilized to estimate the relative motion by Least-Square algorithm. Because the relative motion between the point features of target and robot is inconsistent to that of environment background and robot, the point features of environment background and the errors in feature tracking are removed by RANSAC algorithm. Experiment results validate the efficiency of our method.

Keywords: Mobile robot, binocular vision, target tracking, motion estimation.

1 Introduction

Faithfully tracking of moving target is a critical ability of environment perception for autonomous robot in unstructured environment. In the traditional monocular-based methods, moving target is tracked through two-dimensional video input and target is tracked only depended on image information. In the window-based methods [1], [2], [3], [4], the target is tracked by maximizing the correlation corresponding or minimizing cost function of correlation but is always sensitive to illumination and occlusion. In another way, some kinds of features such as edge [5], corners [6] and lines [7] are extracted as constraints or judgments to handle transform and rotation. But transform, scale and occlusion are still critical problem for the monocular based methods.

Binocular vision, or called as stereo vision, can obtain the depth information of environment easily. Combining image information and depth information, the stereo vision is very appealing to moving target tracking with the support of low price commercial hardware. In [8], [9], [10], [11], [12], moving objects are detected and tracked by the inconsistency of the flow data through the two consecutive frames. This type of method is efficient to distinguish the moving target from surrounding

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 929–935, 2009.

environment background, but is sensitive to the noise of depth data. Unfortunately, there are large numbers of errors of dense stereo matching in real-time running stereo system in generally.

In our method, we extract Harris corners as sparse point features to represent the whole environment stereo image because this type features are stable to moderate distortion and rotation. We can track the target faithfully and sign the position in image accurately by tracking only a few number of point features belongs to the target even if the target is seriously occluded. The features are matched and tracked by our improved MNCC algorithm [13], which is more efficient to reduce the effect of illumination. RANSAC algorithm is utilized to determine whether the point features belongs to the targets and reject the outliers in feature matching and tracking. Furthermore, the computing cost is greatly reduced because there are only spare feature matching rather than dense matching.

The paper is organized as follows. Section 2 describes the process of faithful point features matching and tracking. In Section 3, the relative motion is estimated. In Section 4, RANSAC algorithm is utilized to reject the outliers in target tracking. Experiment results validate our method in section 5. Summary and future works are discussed in the last section.

2 Feature Matching and Tracking

Point features are extracted by Harris operator [14]. The extracted point features are matched and tracked through consecutive stereo frames by computing correlation of window-based intensity. For fast and reliable matching, the coordinate of the point features are rectified by epipolar constraint to find the best match pair of the features along the epipolar lines.

2.1 Feature Matching

There are many existing methods to solve the correspondence problem. In our method, the correspondence is computed by our improved MNCC algorithm that ratio of intensity of pixel and mean intensity of template are used to reduce the effect of illumination.

The best match is determined by bidirectional consistency constraint. Let p_{lc} denote a corner in left image and p_{rc} is a matching candidates in right image. The correspondence C of p_{lc} and p_{rc} is computed by,

$$C = \frac{2\sum_{(x,y)\in W_p} \Delta I_{lc}(x,y)\Delta I_{rc}(x+\Delta x, y+\Delta y)}{\sum_{(x,y)\in W_p} \Delta I_{lc}(x,y)^2 + \sum_{(x,y)\in W_p} \Delta I_{rc}(x+\Delta x, y+\Delta y)^2} \tag{1}$$

where,

I_{lc} and I_{rc} are the intensity of the pixel in left and right image,

$$\Delta I_{lc} = (\frac{I_{lc}(x,y)}{\overline{I}_{lc}} - 1),$$

$$\Delta I_{rc} = (\frac{I_{rc}(x+\Delta x, y+\Delta y)}{\overline{I}_{rc}} - 1),$$

W_p is the widow of template.

2.2 Feature Tracking

In generally, the assumption that the target moves irregular is reasonable. So we must fix a region large enough for point feature tracking. The correlation is still computed by our improved MNCC algorithm to reduce the effect of illumination and the best feature tracking is determined. Most of outliers are rejected by the bidirectional consistency constraint. Then the errors of point feature tracking are rejected secondly by the relative motion estimation and RANSAC algorithm.

3 Relative Motion Estimation

3.1 Reconstruction of Point Feature

The projection of world point $P_i(x_i, y_i, z_i)^T$ in accurately calibrated stereo cameras is computed by equation (2),

$$p_i = MP_i = K(R \mid t)P_i \tag{2}$$

where, K is intrinsic matrix and $(R \mid t)$ is the extrinsic matrix.

The reconstruction of P_i is obtained by equation (3),

$$P_i = (A_i^T A_i)^{-1} A_i^T b_i \tag{3}$$

where,

$$A_i = \begin{bmatrix} c_{l31}u_{li} - c_{l11} & c_{l32}u_{li} - c_{l12} & c_{l33}u_{li} - c_{l13} \\ c_{l31}v_{li} - c_{l11} & c_{l32}v_{li} - c_{l12} & c_{l33}v_{li} - c_{l13} \\ c_{r31}u_{ri} - c_{r11} & c_{r32}u_{ri} - c_{r12} & c_{r33}u_{ri} - c_{r13} \\ c_{r31}v_{ri} - c_{r11} & c_{r32}v_{li} - c_{r12} & c_{r33}v_{ri} - c_{r13} \end{bmatrix},$$

$$b_i = \begin{bmatrix} c_{l14} - c_{l34}u_{li} \\ c_{l24} - c_{l34}v_{li} \\ c_{r14} - c_{r34}u_{ri} \\ c_{r24} - c_{r34}v_{ri} \end{bmatrix}$$

$c_{lij} \in M_l$, $c_{rij} \in M_r$, M_l and M_r are the project matrix of left and right camera.

3.2 Relative Motion Estimation

The change of rotation and translation are denoted as R and T respectively, and the relative position of world point P_i and robot in current time and previous time are denoted as P_{ci} and P_{pi}. The error of motion estimation is computed as follows,

$$e_i = P_{ci} - RP_{pi} - T \tag{4}$$

The rotation matrix R and the translation matrix T are estimated by minimizing the errors,

$$M(R,T) = \min\left(\sum w_i e_i^T e_i\right) \tag{5}$$

where, $w_i = \left(\det\left(\sum_{pi}\right) + \det\left(\sum_{ci}\right)\right)$, the $\sum_{pi}$ and $\sum_{ci}$ are the uncertainty covariance matrix of reconstruction of P_{ci} and P_{pi} respective.

From (4) and (5), the function M can be rewritten as equation (6),

$$M(R,T) = \sum w_i \left(P_{ci} - RP_{pi} - T\right)^T \left(P_{ci} - RP_{pi} - T\right). \tag{6}$$

4 Target Tracking

4.1 Outliers Rejecting by RANSAC

To reduce the effects of the occlusions, we use RANSAC (Random Sample Consensus) algorithm [15] to reject the matched and tracked point features of the static environment as outliers. The features of the moving target are distinguished from the errors of feature tracking by a second RANSAC processing. The two stages are similar. The latter is summarized as the following steps:

(1) Reconstructing matched and tracked features of the target as data set S;
(2) Drawing a sample set s_i for relative motion estimation from S randomly;
(3) Estimating the relative motion between sample set s_i and robot;
(4) Computing the error e_j of each point feature of S by equation (4). If $e_j e_j^T$ less than threshold T, corresponding feature is correctly tracked;
(5) Repeating step2, step3 and step4;
(6) After a certain number of iterations, finding out the best motion estimation with the maximum number of correctly tracked point features.

4.2 Target Determining

At the beginning of target tracking, we define a rectangular region to denote the target. O is the centre, W is the width and H is the height. During the movement of the target, its region is changed. As the target is tracked, the target region in current

stereo image is also determined and the features of the target must be updated for the target tracking in the next frame. In our method, the features of the target are updated by the matched features of the target region in current stereo images.

Denoting the coordinate of the point features of the target on previous stereo image as S_p and corresponding coordinate of tracked features in current stereo image as S_c, the change of O is determined by the difference of mean coordinate of S_p and S_c.

The maximum distance among the features of S_p on row is denoted as L_{pr} and that on column as L_{pc}. Similarly, those of S_c are denoted as L_{cr} and L_{cc} respectively. The change of W and H are computed by,

$$\Delta W = W\left(1-\frac{L_{cr}}{L_{pr}}\right),\ \Delta H = H\left(1-\frac{L_{cc}}{L_{pc}}\right) \tag{7}$$

5 Experiments

A toy car is tracked under large occlusion and rotation in lab environment. The binocular cameras are mounted 1500mm above the ground and the baseline is about 160mm. The parameters of single camera are that the focus is 8mm and the size of pixel is 7.4 m.

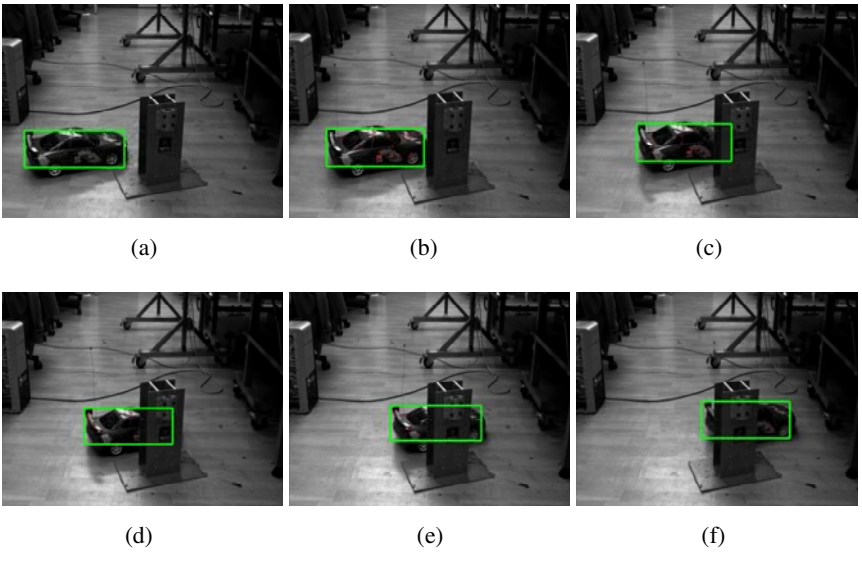


Fig. 1. A toy car is tracked under large occlusion. (a) Target is selected in green rectangular. (b)-(f) The results of the toy car tracking.

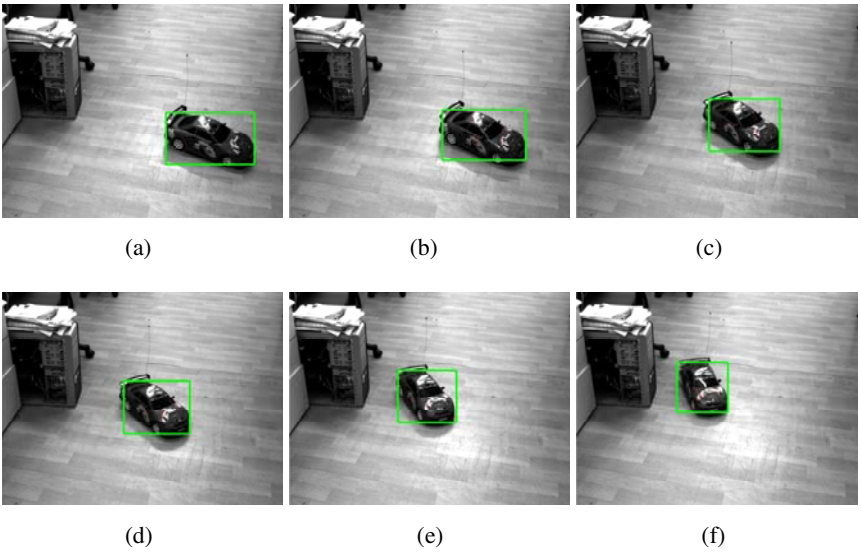


Fig. 2. A toy car is tracked under large rotation. (a) Target is selected in green rectangular. (b)-(f) The results of the toy car tracking.

The results of occlusion experiment are shown in Fig. 1. During the moving of the toy car, the most serious occlusion is larger than 50 percent. The point features of the toy car are tracked correctly and the target region are determine accurately without the effect of large occlusion. Then the point features are updated for target tracking in next stereo frame. In Fig. 2, the toy car are tracked under large rotation. It is about 60 degree that the toy car rotated during its moving. The toy car is still correctlyt tracked because the point features are stable for moderate rotation and are updated between the consecutive stereo frames.

6 Summary

A binocular-based method is presented to track moving target reliably. We represent the target and environment background by point features to reduce the effect of occlusion and rotation. After matching and tracking the point features, the relative motion accurately estimated by Least Square algorithm and the moving target are correctly tracked by twice RANSAC. As experiment results shown, our method is robust and efficient to the effects of large occlusion and rotation. It is necessary for our method that there are enough correctly tracked point features. But it is difficult in some extreme condition such as the target is occluded completely. So we will improve our method to deal with this case in our future work.

Acknowledgments. This work is funded by the Natural Science Foundation of China (60805046) and the Knowledge Innovation Program of Chinese Academy of Sciences (07A1210101). We would like to thank Sun Jing and Zhu Linlin for their advice and support.

References

1. Scharstein, D., Szeliski, R.: A Taxomony and Evaluation of Dense Two-Frame Stereo Correspondence Algorithm. Int. J. Comput. Vis. 47, 7–42 (2002)
2. Fitch, A.J., Kadyrov, A., Christmas, W.J., Kittler, J.: Fast Robust Correlation. IEEE Tran. Image Processing. Vis. 14(8), 1063–1073 (2005)
3. Babu, R.V., Perez, P., Bouthemy, P.: Robust Tracking with Motion Estimation and Local Kernel-based Color Modeling. Image and Vision Computing. Vis. 25, 1205–1216 (2007)
4. Hager, G.D., Belhumeur, P.N.: Efficient Region Tracking With Parametric Models of Geometry and Illumination. IEEE Tran. Pattern Analysis and Machine Intelligence. Vis. 20(10), 1025–1039 (1998)
5. Zhu, G., Zeng, Q., Wang, C.: Efficient Edge-based Object Tracking. Pattern Recognition. Vis. 39, 2223–2226 (2006)
6. Shi, J., Tomasi, C.: Good Features to Track. In: Proc IEEE Conf. Comput. Vis. Pattern Recognit., pp. 593–600 (1994)
7. Lowe, D.: Robust Model-based Motion Tracking through the Intergration of Search and Estimation. Int. J. Comput. Vis. 8, 113–122 (1992)
8. Bae, K.H., Koo, J.S., Kim, E.S.: A New Stereo Object Tracking System using Disparity Motion Vector. Optics Communications. Vis. 221, 23–35 (2003)
9. Parrilla, E., Ginestar, D., Hueso, J.L., Riera, J., Torregrosa, J.R.: Handling Occlusion in Optical Flow Algorithms for Object Tracking. Computers and Mathematics with Applications. Vis. 56(3), 733–742 (2008)
10. Harville, M.: Stereo Person Tracking with Adaptive Plan-View Templates of Height and Occupancy Statistics. Image and Vision Computing. Vis. 22, 127–142 (2004)
11. Rabe, C., Franke, U., Gehrig, S.: Fast Detection of Moving Objects in Complex Scenarios. In: Proc. IEEE Symposium. Intelligent Vehicles, Istanbul Turkey, pp. 398–403 (2007)
12. Ess, A., Leibe, B., Schindler, K., Gool, L.V.: A Mobile Vision System for Robust Multi-person Tracking. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 1–8 (2008)
13. Du, Y., Han, J., Tang, Y.: Feature Matching and Tracking for Visual Odometry of Moblie Robot. In: Proc. International Conf. Information Computing and Automation, Chendu, China, pp. 927–930 (2007)
14. Harris, C., Stephens, M.: A Combined Corner and Edge Detector. In: Proc. British Machine Vis. Conf., pp. 147–151 (1998)
15. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM. Vis. 24(6), 381–395 (1981)

Robust Calibration of a Color Structured Light System Using Color Correction

Xu Zhang and Limin Zhu

Robotics Institute, Shanghai Jiao Tong University,
No.800 Dongchuan Road, Minhang District, Shanghai 200240, China
zhulm@sjtu.edu.cn

Abstract. In this paper, a new method for calibration of the structured light system is presented. Three code words, i.e. red green and blue squares, are used to create the M-array pattern image. Then the color calibration is conducted to deal with the color crosstalk. The color correction can make the decoding process more robust. After that, the camera is calibrated and the 3D corresponding points are also computed using the calibrated camera. Finally, the projector calibration is carried out. The experiment results on our structured light system show that our method is correct and precise.

Keywords: Structured light system, calibration, pseudo-random array.

1 Introduction

The calibration of the structured light system (SLS) is a necessary step in order to extract accurate metric information from 2D images. It often includes two separate stages: camera calibration and projector calibration. Camera calibration has been extensively studied over the years in the computer vision and photogrammetry communities[1,2]. The main difficulty of calibrating the SLS is the projector calibration. Because the projector can't 'see' images like camera and just project the image, determining the correspondence between the projector image and the 3D points is more difficult. The general method[3,4,5,6] to solve the problem is using a calibrated camera to calibrate the projector. The specific pattern image is projected onto the calibration object by the projector. Then, the modulated pattern by the calibration object is captured by the camera.

Joaquim Salvi and Jordi Pagès classified the pattern codification techniques to three categories[7]: time-multiplexing, direct codification and neighborhood codification. The time-multiplexing strategy is based on temporal coding and includes binary codes, n-ary codes, gray code, gray code + phase shifting and hybrid methods. Because the bits of the codeword are multiplexed in time, multiple images are needed. Furthermore, these techniques always encode a single axis. If you want to know the information of two spatial axes, they have to be applied on x axis and y axis respectively. So these techniques will lead to heavy burden of capturing lots of images to determine the corresponding points for just one pose of the 3D calibration object. In the direct codification, each pixel

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 936–946, 2009.

is labeled by a unique code word. These technologies need either a large range of color values or introduce periodicity. In the neighborhood codification technologies, such as non-formal codification, De Bruijn sequences and M-arrays, the codeword that labels a certain point of the pattern is obtained from a neighborhood of the points around it. The strategies based on non-formal codification are generated intuitively without using any mathematical coding theory. The strategies based on De Bruijn sequences are always encode a single axis, so at least two images are needed to get the information of two axes. The strategies based on M-arrays[8,9] take advantage of the theory of perfect maps to encode a unique pattern. Because the M-array techniques code two axes, just one image is enough to known the space information of a certain point. The code words can be represented by different color or shapes. The shape codes lead to bother for decoding, because many shape patterns have to be recognized. However, the color code words are easily implemented and decoded. So the color information is applied in many structured light systems[10,11,12,13].

In an ideal world, the light from each projector color channel reaches only its correspondingly colored sensor pixel (for example, red light is seen only by red pixels). In practice, however, each projector color channel has some influence on all three sensor channels, a phenomenon known as color crosstalk. In some extremely situation, the maximum channel of the camera image is different from the projected monochromatic light. So we have to model this color crosstalk and carry out color correction before decoding the code words.

In this paper, a new method for calibration of the structured light system is presented. Three code words, i.e. red green and blue squares, are used to create pattern image which refers to the pseudo-random array. Then the color calibration is conducted to deal with the color crosstalk. The color correction can make the decoding process more robust. After that, the camera is calibrated and the 3D corresponding points are also computed using the calibrated camera. Finally, the projector calibration is carried out. The experiment results on our structured light system show that our method is correct and precise. The paper is structured as follows. Section 2 introduces the proposed method to calibrate the SLS. Section 3 gives the experiment results. Finally, the conclusion is presented.

2 The Proposed Method for the SLS Calibration

2.1 The Framework of the SLS Calibration

Our calibration method of the SLS includes ten steps:

Step 1: Generate the color images.
Step 2: Generate the M-array code words and create the pattern image and gray image.
Step 3: Project the color images one by one and capture images.
Step 4: Project the pattern image and gray image in turn on the planar target under each pose and capture the corresponding images.
Step 5: Color calibration.

Step 6: Corner detection on the image for camera and projector calibration.
Step 7: Color correction for decoding.
Step 8: Camera calibration.
Step 9: Compute the 3D corresponding points.
Step 10: Projector calibration.

2.2 The M-Array Code Words

The pseudo-random arrays have a great interesting property, the window property. This means that if a window of prescribed size, say $k_1 \times k_2$, is slid over the array, each possible nonzero $k_1 \times k_2$ array is seen through the window exactly once. So each different submatrix is associated with an absolute position in the array and the correspondence between the camera image and projector image is determined.

In this article, Galois field with 3 elements is chosen, denoted by $GF(3)$. The elements of $GF(3)$ are $0, 1, 2$, and $+, -, \times, \div$ are carried out mod 3. We ask the reader to accept this fact[14]: there exists primitive polynomials of degree n for every n to generate pseudo-random sequences with the size of $3^n - 1$. Pseudo-arrays can be constructed by folding pseudo-random sequences[15]. Take a number of the form $N = 3^{k_1 k_2} - 1$ such that $n_1 = 3^{k_1} - 1$ and $n_2 = N/n_1$ are relatively prime and greater than 1.Talbe 1 shows the possible pseudo-random arrays, where n is from 1 to 20. Because the projector image resolution in our SLS is 1024×768, the dimension of the array can not be too small and the widow should be nearly square. In these potential pseudo-random arrays, we choose $n = 9$ with the 3×3 square window.

We can readily get the primitive polynomial over $GF(3)$, $x^9 + x^7 + x^5 + 1$,from the combinatorial object server[16]. We have the iteration function to generate the pseudo-random sequences a.

$$\begin{aligned} a_i &= -a_{i-2} - a_{i-4} - a_{i-9} \\ &(10 \le i \le 3^9 - 1) \\ a_1 &= 1; \quad a_j = 0 \\ &(2 \le j \le 9) \end{aligned} \tag{1}$$

We use a to fill up a 26×757 array, by writing a down the main diagonal and continuing from the opposite side whenever an edge is reached. We just use the 19×25 sub matrix, Fig 1 shows the code words and the code image, where red is 0, green is 1, and blue is 2, the size of square is 20 pixels.

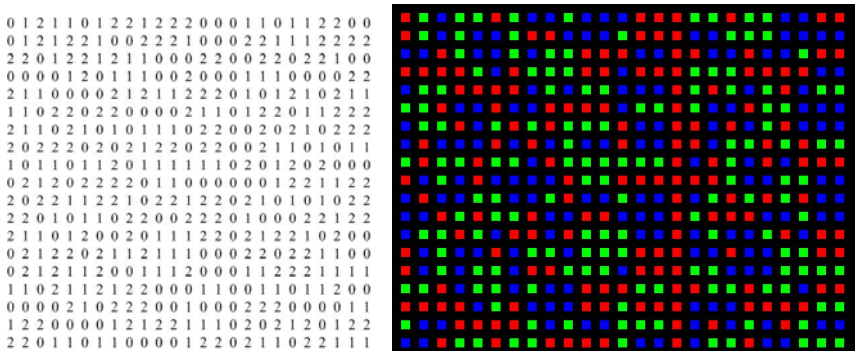
2.3 The Color Calibration and Color Correction

After the illumination pattern from the previous subsection is projected onto the calibration object (the planar target), a camera records an image of the reflected light. In an ideal world, the light from each projector color channel reaches only its correspondingly colored sensor pixel (for example, red light is seen only by red pixels). In practice, however, each projector color channel has some influence on all three sensor channels, a phenomenon known as color crosstalk. Because

Table 1. The possible pseudo-random arrays from 1 to 20

n	$N = n_1 \times n_2$	$k_1 \times k_2$	n	$N = n_1 \times n_2$	$k_1 \times k_2$
1	$*$	$*$	11	$3^{11} - 1 = 2 \times 88573$	1×11
2	$*$	$*$	12	$3^{12} - 1 = 80 \times 6643$	4×3
3	$3^3 - 1 = 2 \times 13$	1×3	13	$3^{13} - 1 = 2 \times 797161$	1×13
4	$*$	$*$	14	$3^{14} - 1 = 8 \times 597871$	2×7
5	$3^5 - 1 = 2 \times 121$	1×5	15	$3^{15} - 1 = 2 \times 7174453$ $3^{15} - 1 = 26 \times 551881$ $3^{15} - 1 = 242 \times 59293$	1×15 3×5 5×3
6	$3^6 - 1 = 8 \times 91$	2×3	16	$*$	$*$
7	$3^7 - 1 = 2 \times 1043$	1×3	17	$3^{17} - 1 = 2 \times 64570081$	1×17
8	$*$	$*$	18	$3^{18} - 1 = 8 \times 48427561$ $3^{18} - 1 = 728 \times 532171$	2×9 6×3
9	$3^9 - 1 = 26 \times 757$	3×3	19	$3^{19} - 1 = 2 \times 581130733$	1×19
10	$*$	$*$	20	$*$	$*$

Note: $*$ means there is not possible pseudo-random arrays.



(a) (b)

Fig. 1. The code image a The code matrix; b The corresponding code image

the surface of calibration object is spectrally uniform, we just need measure a projector-camera color crosstalk matrix that indicates how much each projector channel influences each camera channel.

Generally, for the input of each projector color channel, the output illumination of the projector doesn't only have its corresponding monochromatic light. For example, the illumination is $I_R(\lambda)$ for the red channel 1, where λ is the spectra. So if the input pixel on the projector is $[r_p, g_p, b_p]$, the output illumination is the sum of a red beam, a green beam and a blue beam, which can be denoted as $I^{'}(\lambda) = r_p \cdot I_R(\lambda) + g_p \cdot I_G(\lambda) + b_p \cdot I_B(\lambda)$. The $I_R(\lambda)$,$I_G(\lambda)$, and $I_B(\lambda)$ are nonlinear functions. We use the approximation model instead of it.

$$\begin{bmatrix} r^{'} \\ g^{'} \\ b^{'} \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix} \cdot \begin{bmatrix} r_p \\ g_p \\ b_p \end{bmatrix} \tag{2}$$

Caspi et al[17]introduce the approximation that the reflectance of the scene is constant within each bands.

$$\begin{bmatrix} r_c \\ g_c \\ b_c \end{bmatrix} = \begin{bmatrix} \mu_1 & 0 & 0 \\ 0 & \mu_2 & 0 \\ 0 & 0 & \mu_3 \end{bmatrix} \cdot \begin{bmatrix} r^{'} \\ g^{'} \\ b^{'} \end{bmatrix} \tag{3}$$

Because the surface of calibration object is spectrally uniform, we have $\mu_1 = \mu_2 = \mu_3$.

Similarly, for the monochromatic light, the captured pixel by the camera doesn't only have its corresponding channel. For example, the pixel on the camera image is $[r_{r'R}(r_c) \quad g_{r'G}(r_c) \quad b_{r'B}(r_c)]^T$ for the red incident light. So if the incident light is $[r_c \quad g_c \quad b_c]^T$, the output pixel is the sum of three monochromatic light.

$$\begin{bmatrix} r \\ g \\ b \end{bmatrix} = \begin{bmatrix} r_{r'R}(r_c) + r_{g'R}(g_c) + r_{b'R}(b_c) \\ g_{r'G}(r_c) + g_{g'G}(g_c) + g_{b'G}(b_c) \\ b_{r'B}(r_c) + b_{g'B}(g_c) + b_{b'B}(b_c) \end{bmatrix} \tag{4}$$

Equation (4) also can be simplified to linear equation.

$$\begin{bmatrix} r \\ g \\ b \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix} \cdot \begin{bmatrix} r_c \\ g_c \\ b_c \end{bmatrix} \tag{5}$$

From (2)(3)(5), taking into account the ambient camera readings $[r_0 \quad g_0 \quad b_0]^T$, we have this equation.

$$\begin{bmatrix} r \\ g \\ b \end{bmatrix} = \mu_1 \cdot \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \cdot \begin{bmatrix} r_p \\ g_p \\ b_p \end{bmatrix} + \begin{bmatrix} r_0 \\ g_0 \\ b_0 \end{bmatrix} \qquad A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \tag{6}$$

A is the projector-camera color crosstalk matrix. Without generality, we just set $\mu_1 = 1$. It means the reflectance of the calibration object is the reference unit.

In order to known the A and the vector $[r_0 \quad g_0 \quad b_0]^T$ in (6), we do the following color calibration.

Firstly, the calibration object is acquired an image under just the ambient light. The vector $[r_0 \quad g_0 \quad b_0]^T$ can be readily obtained. Secondly, the input pixel on the projector is just one channel, for example, red channel $[r_p \quad 0 \quad 0]^T$. The corresponding column in the matrix A can be obtained. In this article, three values 85,170,255 for each channel are used. The least square method is applied to calculate the optimal A.

Once we known the A and the vector $[r_0 \quad g_0 \quad b_0]^T$, it is possible to recover the projection input$[r_p \quad g_p \quad b_p]^T$, given the camera reading$[r \quad g \quad b]^T$.

$$\begin{bmatrix} r_p \\ g_p \\ b_p \end{bmatrix} = A^{-1} \cdot \begin{bmatrix} r - r_0 \\ g - g_0 \\ b - b_0 \end{bmatrix} \tag{7}$$

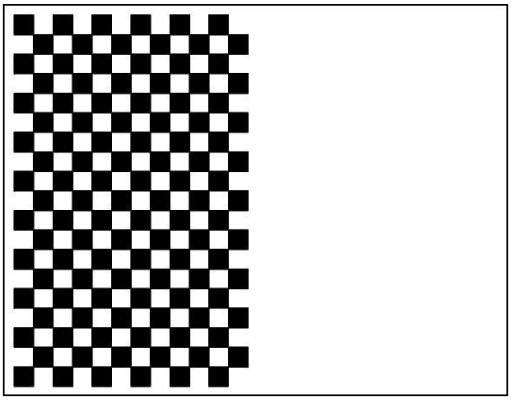


Fig. 2. The planar target

2.4 The Feature Point Detection and the Code Decoding

In this article, the calibration object is the planar target shown in Fig 2. The left is a chessboard pattern containing $18 \times 11 = 198$ corner points. The size of square is $20mm \times 20mm$. The right is blank as the projecting screen.

The feature point for the camera calibration can be readily detected using the sub pixel corner detection method[18]. However, for the projector calibration, it is bothersome. Because the image is color and the corner detection can only deal with gray image, we transform the color image to the gray image using the model HSV, then the sub pixel corner is detected.

The color image has to be corrected using (7) before decoding. After color correction, the image is transformed to the space of the projector input pixel. The code word can be recognized, for the maximum channel is the corresponding code.

2.5 The Camera and Projector Calibration

For the camera calibration, the 3D points and corresponding 2D points are all obtained. The camera can be calibrated using Zhang's method[19]. However, for the projector calibration, there are just camera image points and corresponding projector image points. The 3D corresponding points on the 3D plane have to be obtained. For a calibrated camera, each pixel on the camera image corresponds to a line in the 3D space. Since the 3D plane has been known through camera calibration, the intersection of the line and the 3D plane is the 3D corresponding point for the pixel. So the 3D corresponding points can be computed readily. Now, the projector can be calibrated using the same method as the camera.

3 Experiments

A portable structured light system shown in Fig 3 is developed in our lab. It is composed with a camera with the 1024×1024 resolution and a Digital Light

Fig. 3. The portable structured light system

Procession (DLP) projector with the 1024×768 resolution. The proposed calibration method of the SLS is implemented on the SLS.

3.1 The Performance of Corner Detection

In one process of the SLS calibration, 100 images are captured for just one pose of the planar target. Because the method of the corner detection can't directly deal with the color image, the original color image has to be transformed to a gray image. The first method is transforming the color image to gray image using the HSV model. The second method is extracting the channel corresponding to the maximum projector channel from the color images. The corners for the 100 images are obtained using the same sub pixel corner detection[18] on two kinds of gray images from the same color images. The standard variance is computed for each point. The results are shown in Fig 4. From Fig 4, we can see, the results from the first method have higher precision than ones from the second method. So in this article, the corners are detected on the gray image transformed from the color image using the HSV model.

3.2 The Performance of the Decoding Method Using Color Correction

Our proposed decoding method with color correction is compared with directly decoding basing on the maximum channel without color correction. Figure 5 gives a decoding failure example from the second method on one image. We can see the code word in the second row and the sixth column is decoded as 0, because the red channel is a little larger than the green channel in the original color image. However, in nature, it is 1 under the green light projection. After color correction, the green channel becomes prominent and it is easy to get the

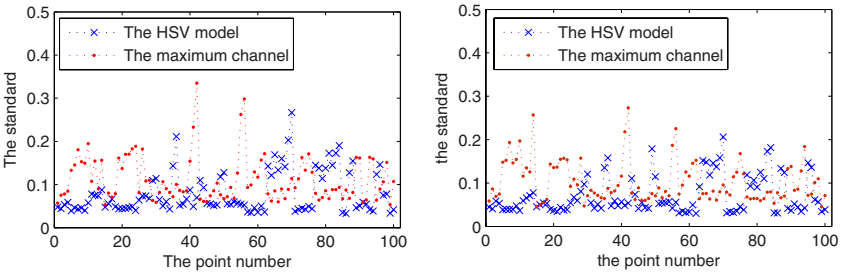


Fig. 4. The standard variance of the corners from two kinds of gray images

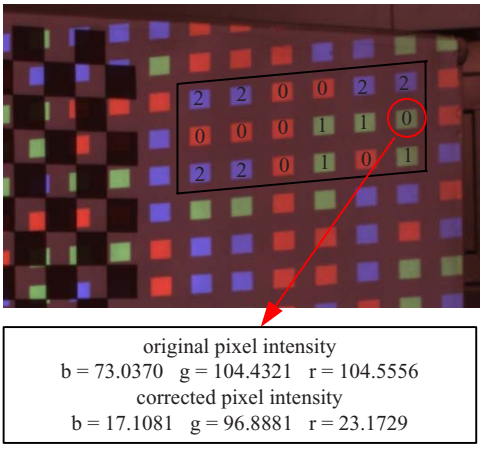
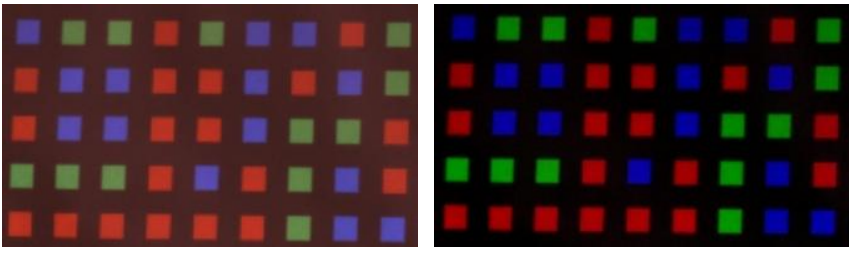


Fig. 5. One example of failure of decoding without color correction

right codeword, 1. The effect of the color correction can be seen in Fig 6. Figure 6 a is the original image and Fig 6 b is the result after color correction. Through comparison, we can see that the color from the corresponding monochrome light is enhanced and the contrast of Fig 6 b is higher.

In this experiment, the performance of two decoding method is compared. One process of the SLS calibration includes 5 couple images from five different poses of the planar target and the process of the SLS calibration is done 20 times. The ambient lights are different, and the apertures are also tuned. Two decoding methods are implemented on the same color images. The times of failure, obtaining wrong code words, are calculated. The results are present on Table 2. It is very clear that our method with color correction can reduce the failure times rapidly.



(a) (b)

Fig. 6. The color correction result. (a) the original image (b) the image after color correction.

Table 2. The failure times from two method in 100 images

	The method without color correction	Our method with color correction
Failure times	19	0

Table 3. The camera and projector calibration results

camera		
	Value	Three times standard variance
$(\alpha \quad \beta)$	(2214.35, 2229.11)	(17.3888 16.4209)
$(u_0 \quad v_0)$	(530.867, 475.213)	(12.4782 21.097)
k	(-0.1021 0.1694 -0.0045 -0.0054)	(0.0367 0.4790 0.0024 0.0023)
projector		
	Value	Three times standard variance
$(\alpha \quad \beta)$	(1976.91, 1968.92)	(16.0278 14.7516)
$(u_0 \quad v_0)$	(548.093, 785.27)	(12.0697 24.9612)
k	(-0.1838 0.2702 -0.0306 0.0007)	(0.0407 0.0886 0.0051 0.0018)

3.3 The Calibration Results

Five couple images are acquired and the camera and projector are calibrated using zhang's method [19]. Table 3 shows the results. we can see the three times standard variance is very small relative to the parameter values and the precision of the parameters is very high.

4 Conclusion

In this paper, a new method for calibration of structured light system is addressed. Three code words,i.e. red green and blue squares are used to create pattern image which refers to the M-array. In order to deal with the color crosstalk,

the color calibration and the color correction are carried out before decoding the codewords. The experiment results show that our proposed algorithm with color correction is more robust in the process of decoding. The camera and projector parameters in our real SLS are also computed using the proposed method. The results show that our method is correct and precise.

Acknowledgement

This work was partially supported by the National Natural Science Foundation of China under grants 50775147 and 50821003, the National Key Basic Research Program under grant 2007CB714005, and the Science & Technology Commission of Shanghai Municipality under grant 07JC14028.

References

1. Clarke, T., Fryer, J.: The development of camera calibration methods and models. Photogrammetric Record 16(91), 51–66 (1998)
2. Salvi, J., Armangue, X., Batlle, J.: A comparative review of camera calibrating methods with accuracy evaluation. Pattern recognition 35(7), 1617–1635 (2002)
3. Zhou, F., Zhang, G.: Complete calibration of a structured light stripe vision sensor through planar target of unknown orientations. Image and Vision Computing 23(1), 59–67 (2005)
4. Legarda-Sáenz, R., Bothe, T., Jüptner, W.: Accurate procedure for the calibration of a structured light system. Optical Engineering 43, 464 (2004)
5. Song, Z., Chung, R.: Use of LCD Panel for Calibrating Structured-Light-Based Range Sensing System. IEEE Transactions on Instrumentation and Measurement 57(11), 2623–2630 (2008)
6. Zhang, S., Huang, P.: Novel method for structured light system calibration. Optical Engineering 45, 083601 (2006)
7. Joaquim, S., Jordi, P., Joan, B.: Pattern codification strategies in structured light systems. Pattern Recognition 37(4), 827–849 (2004)
8. Petriu, E., Bieseman, T., Trif, N., McMath, W., Yeung, S.: Visual object recognition using pseudo-random grid encoding. In: Proceedings of the 1992 lEEE/RSJ International Conference on Intelligent Robots and Systems, 1992, vol. 3 (1992)
9. Kiyasu, S., Hoshino, H., Yano, K., Fujimura, S.: Measurement of the 3-D shape of specular polyhedrons using anM-array coded light source. IEEE Transactions on Instrumentation and Measurement 44(3), 775–778 (1995)
10. Hugli, H., Ma, G.: Generation and Use of Color Pseudo Random Sequences for Coding Structured Light in Active Ranging. Proc. SPIE Industrial Inspection 1010, 75–82 (1988)
11. Monks, T., Carter, J.: Improved stripe matching for colour encoded structured light. In: Fifth International Conference on Computer Analysis of Images and Patterns, pp. 476–485. Springer, Heidelberg (1993)
12. Zhang, L., Curless, B., Seitz, S.: Rapid shape acquisition using color structured light and multi-pass dynamic programming. In: The 1st IEEE International Symposium on 3D Data Processing, Visualization, and Transmission, Citeseer, pp. 24–36 (2002)

13. Fechteler, P., Eisert, P.: Adaptive color classification for structured light systems, pp. 1–7 (June 2008)
14. Berlekamp, E.: Algebraic coding theory, New York (1968)
15. MacWilliams, F., Sloane, N.: Pseudo-random sequences and arrays. Proceedings of the IEEE 64(12), 1715–1729 (1976)
16. Ruskey, F., Cattel, K.: Combinatorial Object Server (2003)
17. Caspi, D., Kiryati, N., Shamir, J.: Range imaging with adaptive color structured light. IEEE Transactions on Pattern analysis and machine intelligence 20(5), 470–480 (1998)
18. Deriche, R., Giraudon, G.: Accurate corner detection: An analytical study. In: Proceedings of Third International Conference on Computer Vision, 1990, pp. 66–70 (1990)
19. Zhang, Z.: A flexible new technique for camera calibration. IEEE Transactions on pattern analysis and machine intelligence 22(11), 1330–1334 (2000)

Close Range Inspection Using Novelty Detection Results

Muhammad Fahmi Miskon and Andrew R. Russell

Intelligent Robotic Research Center (IRRC), Monash University, Australia
{muhammad.miskon,andy.russell}@eng.monash.edu.au

Abstract. This paper presents a close range inspection strategy which requires minimum information from mobile novelty detection results. We propose the use of the estimated position of the detected anomaly to assist path planning for close range inspection. This strategy allows the robot to have more inspection coverage of the surrounding perimeter of the object which in turn will increase the performance of detection using sensors with limited work range. A down sampled laser range finder measurements are used to show that the strategy is robust and work with any range sensors including one that is noisy, inaccurate and has low angular resolution. Experimental results show that the close range inspection brings the robot near the perimeter of the anomalous object and also achieved total close range inspection coverage.

Keywords: Novelty detection, mobile robots, inspection strategy, multi-sensor synergy.

1 Introduction

Surveillance using a mobile robot has many advantages over using static sensor alone. Static sensors are positioned at fixed locations where in many situations, a vast number of the sensors would be required to cover the whole surveillance area [1]. This makes it impractical to use the type of sensors which have a very limited working range [2-4]. The use of a mobile robot overcomes this limitation by allowing the sensors to be brought closer to the object of interest.

Sensing from a close range also benefits any kind of sensors. In general, the closer a sensor is to the source of the entity that it can sense, the more sensitive it will be. For example, chemical concentration is more diluted the further it is from the source. The same with radiation level, magnetic field, ambient light or other entities. Even with images from a camera, a close range snapshot will highlight more detail features such as the texture of the object. Apart from bringing the sensors close to the object of interest, the mobile robot could also diversify the angle of perception of the sensors. A camera would certainly benefit from this, as any object usually looks different from different angles. Other example includes the situation where the wind flows influence the chemical plume in a specific direction.

Motivated by these benefits, we propose a novel approach to perform a close range inspection by using a mobile robot and a range sensor. The approach utilizes information from the anomaly position estimation strategy which was presented in our previous work [5, 6]. When using the proposed approach, close range inspection is

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 947–956, 2009.

performed only when an anomalous object is detected using sensors which have a larger work range and can provide position information such as range sensors.

The work presented in this paper is part of a project which investigates the challenges and benefits of implementing novelty detection (i.e. a mechanism to highlight unusual condition in the environment) on a mobile robot. For this reason, the work in this paper is constrained by the limited information from the result of novelty detection. To prove that our method is robust and requires minimum information, we have down-sampled the angular resolution of the laser range finder sensor that is used by the robot to detect anomalies. The contribution of this paper is a robust method to perform close range inspection of a detected anomaly.

The rest of this paper is organized as follows. Section 2 presents the related work. This is followed by an overview of the system given in Section 3. Section 4 discussed the close range inspection strategy. Experimental results are given in Section 5. This paper is concluded in Section 6.

2 Related Work

The scope of the work described in this chapter is focused on how a mobile novelty detection system that has been described thus far could fully utilize its capabilities and overcome its limitations in order to perform a close range inspection. These make the work in this chapter unique. Thus far, the robot is capable to perform novelty detection on extended areas, locate the location of the source of novelty using the repetitive observation strategy and filter noises using the false positive filter. On the other hand, the robot also has some limitation. As it is designed for mobile robots or standalone novelty detection systems with minimum requirement, it is expected that the robot will use non sophisticated and noisy sensors that don't require too much processing power, small storage and are not expensive. As mention before, this is one of the reasons why the laser range finder that is used in this thesis is down sampled in order to emulate these sensors. The discussion in this section will be based on these capabilities and limitations, as well as the nature of the task which is the close range inspection.

Close range inspection requires a navigation strategy. One of the simplest navigation methods is bug algorithm and its variations [7-13]. A primary advantage of bug algorithm is that minimum information is needed, that is a start and a goal position and sensor measurements for obstacle avoidance. A more advanced navigation approaches involves planning of the path. Some of the popular approaches for path planning includes visibility graph (e.g. A*[14], Dijkstra [15] algorithm), grid based (e.g. distance transform [16]), potential fields [17] and sampling based. These methods require additional information that is the position of the robot and the space that it occupies. Unlike the algorithms in the bug family [18], path planning requires some prior knowledge of the obstacles before the robot move toward a goal.

Traditionally the objective of the path planning is to determine the best path to use between start and goal points. Focus is usually given to finding the lowest cost path from the robot's start state to the goal state. Cost can be in terms of the distance travelled, energy spent etc. However, unlike the common objective of this type of path planning, the objective of close range inspection is to maximize the inspection

coverage. Coverage path planning [19] is closer to the problem at hand because it gives more attention to the layout of the path itself. This type of path planning place emphasis on the space swept out by the robot. Examples of robotic application that benefit from coverage path planning are lawn mowing [20], harvesting[21], mine hunting [22] and floor cleaning [23], where in order to accomplish their task, they must cover all points in the environment. Coverage path planning usually involves breaking down the target region into cells through a process called decomposition. The problem is then simplified to visit all cells within the target region [24, 25]. This is done to provide some form of guarantee or measureable proof of either the completeness or optimality of the coverage.

Close range inspection is unique in its objectives and in the challenges that arise in achieving them. First, unlike any path planning strategy, the aim is to travel as close to and to cover as much of the perimeter of the anomalous object as possible. Secondly, there is no prior information available for performing the navigation, including the start and goal position. Thus, for whatever navigation approaches that is employed by the close range inspection strategy, the robot need to autonomously determine the start and goal position. By using novelty detection and repetitive observation strategy, some information could be acquired about the approximate position of the object. Assuming that no other prior information is available, reactive algorithm similar to bug navigation method is the best option. However, some modification is needed to achieve the objective of close range inspection, which is fundamentally different form bug algorithm.

3 System Overview

An overview of the close range inspection system use for the close range inspection is given in Fig. 1. As mention earlier, the close range inspection depends on the

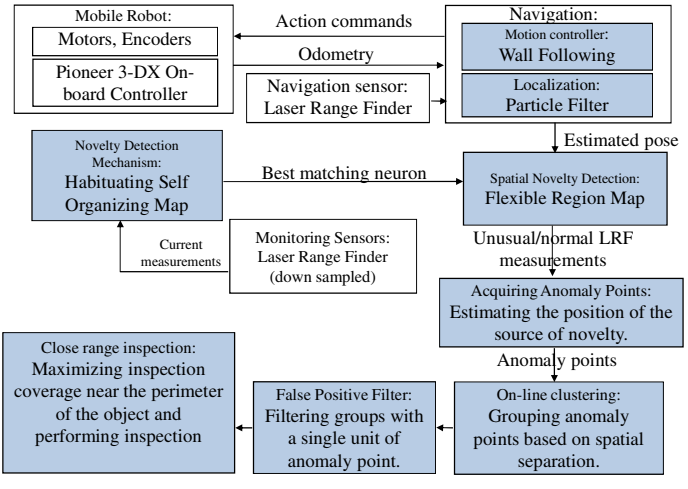


Fig. 1. System overview

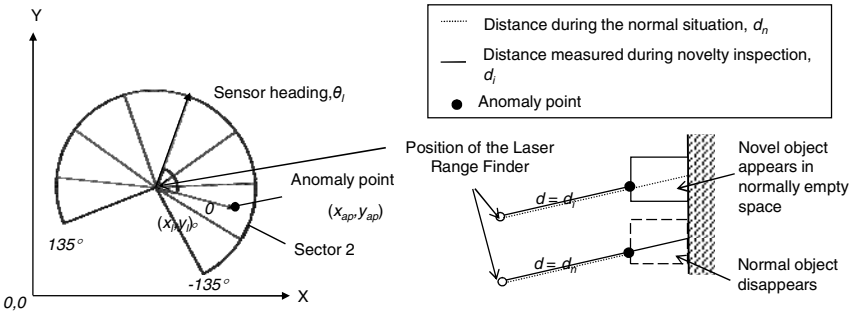


Fig. 2. Input to novelty detector (left) and estimation of anomalous source position (right)

information from the results of novelty detection that was developed in our previous work. In brief, novelty detection works by taking a current measurement and compare it with measurement that is normally perceived at the same location. Fig. 2 shows a snapshot of a laser range finder scan which is divided into 8 sectors to emulate noisy sensor with coarse angular resolution (such as an array of 8 sonar or infrared sensors). Each sector contains an equal number of laser measurements. All the range measurements in each sector are averaged and the average forms the input vector to the Habituating Self Organizing Map (HSOM). HSOM neuron that best match the input measurement becomes the input to the spatial novelty detector together with the pose of the robot/sensor. The position of the anomaly (anomaly point) is estimated using input component (or sector average) that has the highest difference from the normal neuron (see Fig. 2). Repeated observation is performed to confirm the presence of the anomaly. The distribution of the anomaly points also represents the shape of the surface of the object. A detailed explanation of this part of the system is presented in [5, 6].

4 Navigation Strategy

The objective of the close range inspection is to bring the robot close to the vicinity of the anomalous object. Since the distribution of the anomaly points approximately represents the position of the surface of the object, the anomaly points are used to guide the close range navigation. Once a true detection is confirmed, the robot determines the nearest and the furthest anomaly points from its current position. These points are used interchangeably as the start and goal position. The start and the goal position are represented as a point in 2-d space and the robot task is to move between these two points. However, this is not practical in a real world scenario because the anomaly points could be position inside an object, the robot positioning accuracy is not 100%, the robot size is bigger than a point and the robot is restricted to within a safety distance from the object. For this reason a distance threshold, D_T is introduced so the robot is practically on the points if it is within a certain radius from the points.

There is a possibility that the robot moves further away from its goal position when it reacts to obstacles. This is especially true when the robot tries to return back to the starting position after it has reached the goal position. For example, in Fig. 3 if there is a wall on the left side of the anomalous object, avoiding obstacles (the wall) by using a left wall following behavior will make the robot moves further away from the object. To solve this problem, a threshold, D_E is introduced. If the robot misses its target destination or moves further away from its target destination by D_E, it will stop the action and perform the next step of the algorithm. This is necessary to guarantee the termination of the close range inspection.

The detail of the navigation algorithm is as follows (see illustration in Fig. 3):

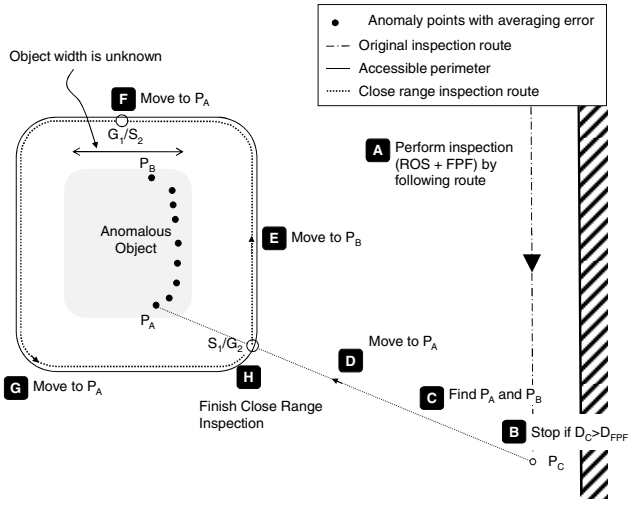


Fig. 3. Close Range Inspection Strategy

Notations:

P_C - Position where the robot confirms the presence of a true anomaly and starts close range inspection.

P_G - Goal position

P_A - Nearest anomaly point position from P_C.

P_B - Furthest anomaly point position from P_C.

P_R - Current robot position.

D_{AB} - Distance between P_A and P_B.

D_{LDA} - Distance travel since robot last detect an anomaly point which belong to a confirm detection.

D_{FPF} - Threshold distance for confirming a true positive detection.

D_{RG} - Distance between P_R to P_G

$D_{RG'}$ - Current distance between P_R to P_G

D_{RA} - Initial distance between P_R and P_A.
$D_{RA'}$ - Current distance between P_R and P_A.
D_{RB} - Initial distance between P_R and P_B.
$D_{RB'}$ - Current distance between P_R and P_B.
D_{RO} - Current distance between P_R and object.
D_{Travel}- Robot's distance travel.
D_E - Threshold distance for stopping the robot as it moves pass or further away from the goal.
D_T - Threshold to indicate that robot has reach the target position.
D_S - Minimum safety distance between P_R and object.
Reach goal - $D_{RG'} < D_T$
Pass goal - $D_{Travel} > D_{RG} + D_E$
Move away from goal - $D_{RG'} > D_{RG} + D_E$
Safe - $D_{RO} > D_S$

Parameter values (note: distance to target position is measured from the center of robot):
D_E = 500mm.
D_T = 300mm.
D_S = 600mm.

Main()

```
While Perform common (non close range) inspection
  If (D_LDA > D_FPF)
    P_C = P_R
    Close Range Inspection()
    Move to P_C
  EndIf
EndWhile
```

Navigate(G=A/B,W=left/right)

```
Set D_RG = D_RG'. Reset D_Travel = 0.
Turn to P_G
While (Not Pass goal AND Not Move away from goal)
  If (obstacles)
    Follow W wall
  Else
    Move straight to P_G
  EndIf
  Operate close range sensor
EndWhile
```

```
Close Range Inspection() (CRI)
  Find P_A and P_B.
  While (Safe AND Not Reach goal)
    Navigate(A,left)
  EndWhile
  Navigate(B,left)
  If (Pass goal)
    Navigate(A,left)
    If (Pass goal)
      Terminate CRI
    Else (i.e. Move away from goal)
      Navigate(A,right)
      If (Pass goal OR Move away from goal)
        Terminate CRI
      EndIf
    EndIf
  Else (i.e. Move away from goal)
    Navigate(B,right)
    If (Pass goal)
      Navigate (A,right)
      If (Pass goal OR Move away from goal)
        Terminate CRI
      EndIf
    Else (i.e. Move away from goal)
      Terminate CRI
    EndIf
  EndIf
```

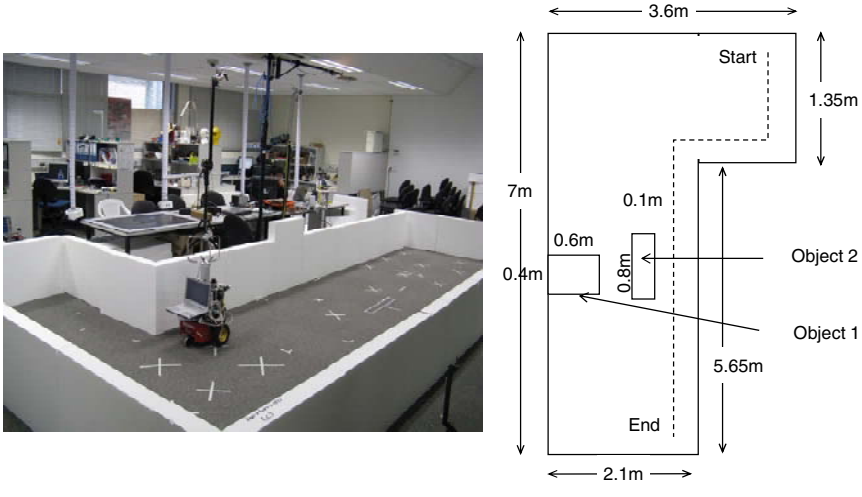
5 Experimental Results

In order to analyze the performance of the close range inspections, experiments were conducted in an environment shown in Fig. 4. The robot employed a wall following behavior for navigation and localized itself using particle filter. The objectives of the experiments are to analyze the performance of the strategy by using the measure of coverage given by Equation (1) where $p_{covered}$ is the distance covered by the robot during close range inspection and $p_{accesible}$ is the actual accessible area by the robot for close range inspection. The algorithm was tested in three different scenarios, when an object was partially accessible, when an object is fully accessible and when a normal object is missing from its location. Prior to this, the robot was trained to learned normal condition of the environment, where in the first two scenario, the environment is normally empty whereas in the last scenario, the environment normally has object 2 (see Fig. 4) at its center.

$$Coverage = \frac{P_{covered}}{P_{accessible}} \times 100\% \quad (1)$$

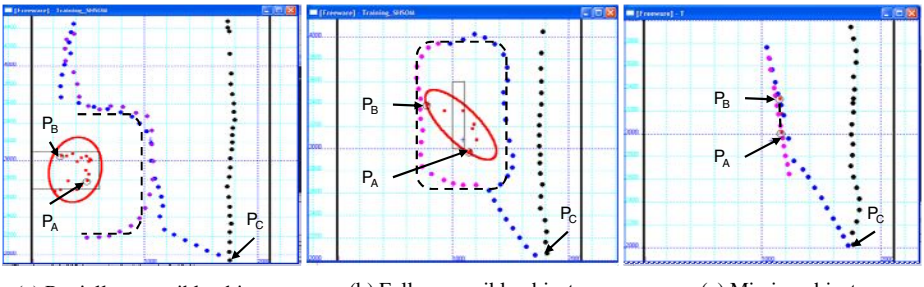
Fig. 5 shows the results of the experiments in the different scenarios. In the figures, anomaly points are represented as the red square dots. As we can see, by using the same algorithm and parameter settings, the robot successfully achieved 100% coverage in all three situations. However, in the case of a missing object, the threshold D_E made the robot overly inspect the nearby areas.

This is the tradeoff in choosing different value of D_E. A higher D_E value guarantees total coverage but the robot might over inspect. On the other hand a lower D_E value ensures robot navigate only near the vicinity of the anomalous object at the expense of the possibility of not achieving 100% coverage.



(a) Image of the L-shape environment (b) Dimension of the L-shape environment

Fig. 4. Experimental setup



(a) Partially accessible object (b) Fully accessible object (c) Missing object

Fig. 5. Experimental results; the close range inspection paths. The black dots represent the trail of the robot on its usual route. Blue and purple dots represent the trail of the robot on its close range inspection route. Different colors (blue and purple) are used to clarify movement of robot in different directions.

6 Conclusion

The close range inspection strategy described in this paper uses minimum information from novelty detection results using range sensors and a mobile robot. It is a novel approach to perform an autonomous surveillance task. This approach opens up the possibility of using many types of sensors which were previously neglected for surveillance tasks because of their limited working range. Experimental results show that a full close range coverage could be achieved using the close range inspection strategy.

Other than surveillance, other application of the close range inspection strategy is autonomous learning. The robot can gather data of new object using many of its sensors from a close range. The robot can also see the new object from different viewing points by following the close range inspection route. Future work includes the possibility of using sensors other than the laser range finder in order to estimate the anomaly position and to perform further close range inspection.

References

1. Saidi, A., Chiricescu, S., Norris, J.M., Schuette, M.A.: Method and System for Gas Leak Detection and Localization. In: Publication, U.S.P.A. (ed.), vol. US 2008/0168826 A1. Motorola, Inc., United State of America (2008)
2. Miskon, M.F., Russell, R.A.: A novel electromagnetic signature sensor for mobile robots. In: IASTED International Conference on Robotics and Application, Weurzberg, Germany (2007)
3. Lilienthal, A., Zell, A., Wandel, M., Weimar, U.: Experiences Using Gas Sensors on an Autonomous Mobile Robot. In: Proceedings of EUROBOT 2001, 4th European Workshop on Advanced Mobile Robots, pp. 1–8. IEEE Computer Press, Lund (2001)
4. Anderson, J.L., Cantu, A.A., Chow, A.W., Fussell, P.S., Nuzzo, R.G., Parmeter, J.E., Sayler, G.S., Shreeve, J.M., Slusher, R.E., Story, M., Trogler, W., Vekatasubramanan, V., Waller, L.A., Young, J., Zukoski, C.F.: Existing and Potiential Standoff Explosives Detection Techniques. The national Academic Press, Washington (2004)
5. Miskon, M.F., Russell, R.A.: A Repetitive Observation Strategy for Recognizing a True Anomaly and Estimating its Position. In: Austalasia Conference on Robotics and Automation, Canberra (2008)
6. Miskon, M.F., Russell, R.A.: Mapping Normal Sensor Measurement Using Regions. In: IEEE International Conference of Industrial Technology, Melbourne, Australia (2009)
7. Lumelsky, V.J., Stepanov, A.A.: Dynamic path planning for a mobile automaton with limited information on the environment. IEEE Trans. Automat. Contr. 31, 1058–1063 (1986)
8. Sankaranarayanan, A., Vidyasagar, M.: A new path planning algorithm for moving a point object amidst unknown obstacles in a plane. In: Proc. of the IEEE Int. Conf. Robot. Autom., vol. 3, pp. 1930–1936 (1990)
9. Sankaranarayanan, A., Vidyasagar, M.: Path planning for moving a point object amidst unknown obstacles in a plane: a new algorithm and a general theory for algorithm development. In: Proc. of the IEEE Int. Conf. on Decision and Control, vol. 2, pp. 1111–1119 (1990)
10. Kamon, I., Rivlin, E.: Sensory-based motion planning with global proofs. IEEE Trans. Robot. Autom. 13, 814–822 (1997)

11. Noborio, H.: A path-planning algorithm for generation of an intuitively reasonable path in an uncertain 2-D workspace. In: Proc. of the Japan-USA Symposium on Flex. Autom., vol. 2, pp. 477–480 (1990)
12. Noborio, H., Maeda, Y., Urakawa, K.: A comparative study of sensor-based path-planning algorithms in an unknown maze. In: Proc. of the IEEE/RSI Int. Conf. on Intelligent Robots and Systems, vol. 2, pp. 909–916 (2000)
13. Kamon, I., Rivlin, E., Rimon, E.: TangentBug: a range-sensor based navigation algorithm. J. Robot. Res. 17, 934–953 (1998)
14. Ko, W.S., Senevieatne, L.D., Earles, S.W.E.: Space representation and map building-A triangulation model to pathplanning with obstacle avoidance. In: Proceedings of the 1993 IEEE/RSJ International Conference on Intelligent Robots and Systems, vol. 3, pp. 2222–2227 (1993)
15. Foux, G., Heymann, M., Bruckstein, A.: Two-dimensional robot navigation among unknown stationary polygonalobstacles. IEEE Transactions on Robotics and Automation 9, 96–102 (1993)
16. Jarvis, R.A.: Collision-Free Trajectory Planning Using the Distance Transforms. Mechanical Engineering Trans. of the Institution of Engineers, Australia 10 (September 1985)
17. Bräunl, T.: Embedded Robotics. Springer, Berlin (2003)
18. Ng, J., Braunl, T.: Performance Comparison of Bug Navigation Algorithms. Journal of Intelligent and Robotic Systems 50, 73–84 (2007)
19. Choset, H.: Coverage for robotics – A survey of recent results. Annals of Mathematics and Artificial Intelligence 31, 113–126 (2001)
20. Huang, Y.Y., Cao, Z.L., Hall, E.L.: Region filling operations for mobile robot using computer graphics. In: Proceedings of the IEEE Conference on Robotics and Automation, pp. 1607–1614 (1986)
21. Ollis, M., Stentz, A.: First results in vision-based crop line tracking. In: IEEE International Conference on Robotics and Automation (1996)
22. Land, S., Choset, H.: Coverage path planning for landmine location. In: Third International Symposium on Technology and the Mine Problem, Monterey, CA (1998)
23. Colegrave, J., Branch, A.: A case study of autonomous household vacuum cleaner. In: AIAA/NASA CIRFFSS (1994)
24. Zelinsky, A., Jarvis, R.A., Byrne, J.C., Yuta, S.: Planning paths of complete coverage of an unstructured environment by a mobile robot. In: Proceedings of International Conference on Advanced Robotics, Tokyo, Japan, November 1993, pp. 533–538 (1993)
25. Lumelsky, V., Mukhopadhyay, S., Sun, K.: Dynamic path planning in sensor-based terrain acquisition. IEEE Trans. Robotics Automation 6, 462–472 (1990)

The Vision System of the ACROBOTER Project

Rigas Kouskouridas, Nikolaos Kyriakoulis, Dimitrios Chrysostomou, Vasileios Belagiannis, Spyridon G. Mouroutsos, and Antonios Gasteratos

Democritus University of Thrace, School of Engineering,
Vas. Sofias 12, 67100 Xanthi, Greece
{rkouskou,nkyriako,dchrisos,vm6872,agaster}@pme.duth.gr,
sgmour@ee.duth.gr

Abstract. The ACROBOTER project aims to develop a radically new locomotion technology that can effectively be used in a home and/or in a workplace environment for manipulating small objects autonomously. It extends the workspace of existing indoor service robots in the vertical direction, whilst its novel structure allows covering the whole volume of a room. For the adequate accomplishment of demanding manipulating tasks, its vision system must provide vital visual information concerning the position of the robot in the 3D working space and the topology of possible objects/obstacles in robot's trajectory. Thus, the proposed system is capable of: estimating robot's pose in the room; reconstruct the 3D working space and; recognize objects with remarkable efficiency. In this work, initially, we present the basic structure of ACROBOTER and its vision system and, we also evaluate the aforementioned functions.

Keywords: Pose Estimation, 3D Reconstruction, Object Recognition, Vision Systems.

1 Introduction

In the last few years, a remarkable increase of autonomous robots' usage in domestic workplace environments has been discerned. A wealth of research is devoted in building new frameworks capable of assisting people in everyday life. Moreover, industries address all their efforts to developing machines for substituting humans in house chores such as tidying child's bedroom after a party or collecting clothes before they enter the washing machine. The need of robots working closely to human beings makes a necessity the usage of intelligent systems. One of the most interesting evolutioned system is the ACROBOTER (Autonomous Collaborative Robot to Swing and Work in Everyday EnviRonment) whose vision system is presented in this work.

The main idea underlying the ACROBOTER project is the development of new locomotion technology for manipulating small objects and/or assisting humans in their movements or exercises. This new type of mobile robot is able to move fast and in any 3D direction in an interior environment. Due to the fact that it extends the workspace of existing robots in the vertical direction, it is

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 957–966, 2009.

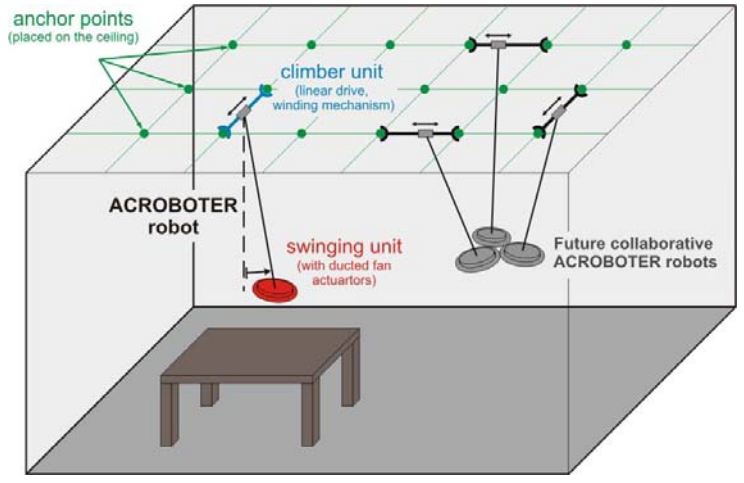


Fig. 1. Concept of the ACROBOTER project

able to operate on the top of tables, wardrobes and can be used for manipulating objects placed on shelves, tables, work surfaces or on the floor.

The conceptual idea of the ACROBOTER project is the development of a completely new technology of a wire suspended robot as it is shown in Fig. 1. Specifically, the whole system is divided into several sub-systems whose architecture, apart from the vision's one, is not presented in this work. The moving platform depends on the anchor points-units placed in a raster fixed to the ceiling of the room. The pendulum-like structure shown in the figure corresponds to the swinging unit (SU) that hangs on a wire, whilst the necessary vertical movements are provided by a winding mechanism (WM) placed on the climber unit (CU). The horizontal transportation of the robot are fulfilled by the CU combined with a linear drive for the fine adjustment of SU's position between the anchoring points. In addition, the SW is equipped with ducted fan actuators that provide a free motion inside a conic volume. Therefore, ACROBOTER's ducted fan system is also used for the posing and as well as stabilization of the robot's motion, and as a result, the unit can fly freely around a suspension point.

As far as the vision system is concerned, it is responsible for three vital tasks that affect directly the overall efficiency of the project. Especially, the vision part emphasizes in the estimation of the SU's pose, the reconstruction of the 3D working environment of the robot and the recognition of objects found in the scene. It consists of five Grasshopper cameras dispersed in different positions. The Grasshopper camera manufactured by Point Grey Research is able to capture images up to 1280 X 960 pixels resolution and it is connected to the PC via a firewire port, using the IEEE 1394b transfer protocol. As far as 3D reconstruction and object recognition are concerned, they are based on four cameras installed in the four corners of the room shown as shown in Fig. 1. The last camera, which provides vital data to compute the SU's pose, is mounted onto

the CU and is fixated towards the ground. The aforementioned demanding tasks are adequately fulfilled by developing and implementing techniques beyond the state-of-the-art.

The remainder of this paper is structured as follows: In Section 2 we give an overview about the related work in the areas of pose estimation, 3D reconstruction and object recognition. ACROBOTER's pose estimation method is presented in Section 3, where both the hardware architecture and the first experimental results are shown. In addition, the 3D reconstruction technique adopted along with the corresponding results are demonstrated in Section 4. Furthermore, essential information concerning the object recognition scheme and how it is trained for the purposes of the ACROBOTER, is presented in Section 5. The last Section 6, concludes with some final notes and an outlook to the future work.

2 Related Work

The essential idea underlying the ACROBOTER project is this new type of autonomous mobile platform, which is designed to move fast and in any direction in an indoor environment. Furthermore, the main challenge is to navigate around any kind of obstacles such as stairs, doorsteps and various other everyday objects that can be found in a room. The ACROBOTER concept outperforms existing developments that require a robot to climb walls and ceiling, due to the fact that it extends significantly the vertical working space of any state-of-the-art similar robotics application. The MATS robot [1], walking robots like Honda's ASIMO or even the Care-o-Bot from IPA proved to have many drawbacks compared to ACROBOTER. The only system or mobile platform that approximates the design or the concept of the presented project is the Flora ceiling based service robot [2]. The latter uses some kind of telescopic arms for the navigation of the working unit into the 3D working space and electromagnetic force for the stabilization of the moving cart on the ceiling.

As far as pose estimation is concerned, it is achieved by using two sensors, a visual and an inertial one. The visual sensor is mounted on the CU and is fixated towards the ground, where the SU operates, and computes the extrinsic parameters of the SU in real-time. An IMU senor lies on the top of the SU and its measurements are fused with the camera ones. This visuo-inertial fusion has been used to many robotics applications as their complimentarity provides efficiency and robustness to the system. The visuo-inertial applications can be divided into three categories, namely the correction, the colligation and the fusion one [3]. The correction deals the readings of the camera as the desired values, while the IMU readings are the estimated ones and vice versa. Concerning the colligation type of integration, a combination of all the measurements takes place. Usually there is a control loop to verify, the camera measurements with the IMU ones and vice versa [4]. Finally, the fusion category takes both measurements into account. Gemeiner *et al.* have estimated the egomotion and the environmental structure by assessing a binocular vision system integrated with an inertial sensor

[5]. Shademan *et al.* [6] have demontrated a comparative study for fusing inertial sensors with visual ones via EKF and iterated extended Kalman filter (IEKF). The main difference of the pose estimation subtask (PES) in the ACROBOTER project with the aforementioned work is that the visual sensor has a different reference frame from the IMU. Furthermore the fact that the visual sensor is constantly moving increases the subsystem's complexity to a high degree. The PES architecture used in the ACROBOTER project is selected after examination described in [7].

The key framework where the 3D Reconstruction task of the ACROBOTER is based on is the computation of the representation of every object in an everyday room using wide baseline stereo techniques. Previous approaches on this field encompass the use of small correlations windows among the views [8] or optimization techniques like graph-cuts [9]. The drawback of using such small image windows and optimization techniques is that are very sensitive to illumination changes and textureless surfaces. Recently, methods that rely on the simplicity and discriminative power of feature detectors and descriptors have been presented [10] and have been used mostly for depth estimation applications.

Recognizing objects in a scene is fundamental task in image understanding and still constitute one of the most challenging tasks. Every pattern recognition technique is directly related with the decryption of vital visual information contained in the natural environment. During the past few years, researchers emphasized in building new recognition frameworks based on appearance features with local estate. Algorithms of this field extract features with local extent that are invariant to possible illumination, viewpoint, rotation and scale changes. In addition, several techniques that enforce the crucial role of local features in demanding pattern recognition application were presented [11]. One of the most efficient object recognition scheme, that is adopted for the purposes of ACROBOTER, is the Scale Invariant Feature Transform (SIFT) [12]. The latter was selected among a set of high-level algorithms to describe patterns and objects. Furthermore, by using information derived from SIFT, we were able to estimate the distance between camera and objects found in a scene.

3 ACROBOTER's Pose Estimation Task

In the ACROBOTER project the pose estimation task provides the orientation and the position vectors of the SU. The PES, is composed of two senors and two computers. A camera is mounted onto the CU and is fixated towards the ground, while a IMU sensor is placed on the top surface of the SU. Both sensors are connected to the respective subsystem computers, while their communication is built on top of sockets. The ACROBOTER system operates in real-time, demanding so, the PES to be able to operate also in high frequencies. In order to meet the requirements of low processing times the PES utilizes the C# OpenCV (Open Source Computer Vision) libraries [13] for the vision tasks, which exhibit low computational load and, thus, they achieve real-time operation. The IMU sensor readings, are acquired by a program code which is implemented in the

C# environment, as well. The measurements of the sensors are fused by the extended Kalman filter (EFK) as it is capable to deal with non-linear problems, and its recursive nature eliminates the errors [14].

The ACROBOTER[1] pose estimation subsystem is depicted in Fig. 2(a). The two computers are connected with each other, through a high speed network protocol as it is illustrated in Fig. 2(b). The visual pose estimation measurements are fused at the SU PC. The EKF is utilized in a function which has two inputs, one for each sensor. The inputs to the EFK are vector states, which includes the positions and the rotations of the three axis. The filter's error covariances, R and Q, concerning the measurement and the process noises, respectively, are tuned to have a ratio of 10 (R/Q=10). These tuning variables determine whether the output follows more the IMU's measurements or the camera's ones. Depending on the reliability of the sensors, the value of Q can be increased, to give more weight to the visual pose, or on the contrary can be decreased, in order the IMU's pose estimates to gain more weight. However, there is a trade off between the cumulative drift error of the inertial unit sensors and the possible occlusions to the field of view regarding the camera. As a result, the tuning process of the filter is highly application dependent and it is mostly a trial and error one.

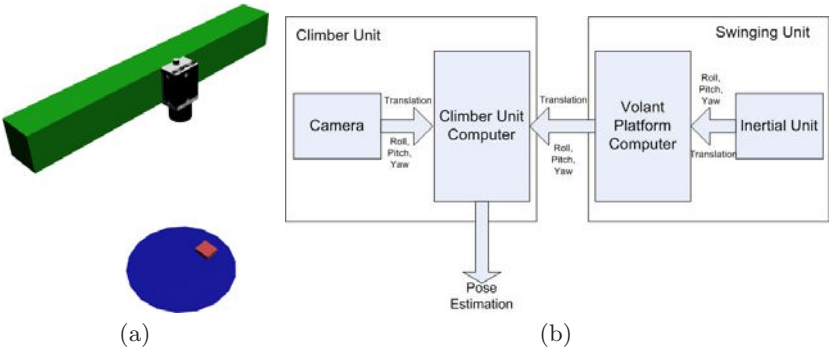


Fig. 2. (a) The ACROBOTER system hardware architecture. The orange box represent the inertial unit for SU, blue cylinder. (b) The proposed system architecture.

The first prototype for the assessment of the visual pose estimation system was performed by identifying a chess board with known geometry. The algorithm identifies the SU by recognizing all the features, i.e. the corners of the chess board squares. The final rotation and translation matrices are given by the OpenCV libraries. Although there is not going to be used a chess board as a feature for the ACROBOTER, the same procedure applies. The landmarks are going to be an arrangement of markers with known geometry. Thus, instead of having as

[1] The ACROBOTER project is still under development. There are copyright restrictions for demonstrating the proposed parts into images. The only parts that can be illustrated are the camera and the inertial sensors.

features the square corners, the markers will be the new features. In order to test the efficiency of our system we mounted the camera on a pan-tilt mechanism and rotated the camera to known angles. The translation measurement where tested with a laser sensor. The accuracy the system is demonstrated in Table 1(a).

Table 1. Overall results for the rotations (left) and for the translations (right). The pan-tilt and the laser sensor measurements are used as the ground truth.

	Pan-Tilt	Camera	IMU	EFK
Roll	14.66	13.99	14.83	14.62
Pitch	4.47	4.58	4.36	4.42
Yaw	9.08	8.84	8.93	8.89
Accuracy	100%	96.7%	98.2%	98.8%
Error	0	0.34	0.15	0.1

	Laser	Camera	IMU	EFK
x	56.21	55.87	57.60	56.47
y	34.82	33.31	33.26	33.39
z	76.63	76.07	77.33	76.51
Accuracy	100%	98.11%	97.4%	98.33%
Error	0	0.80	1.22	0.64

4 ACROBOTER's 3D Reconstruction Task

Reconstruction of surfaces from multiple images has been a central research problem in Computer Vision for a long time. Early previous work in this area focused on developing stereo algorithms mostly for binocular camera configurations. More recently, however, due to significant advances in computational power, vision systems using multiple cameras are becoming increasingly feasible and practical. Multi camera systems like the *Vi Room* which is a low cost synchronized multi camera system developed by [15] and the *3D Room* developed by [16] are systems able to capture multiple synchronized images of indoor scenes. They were developed mainly for tracking applications without providing any knowledge for the structure of the 3D environment.

Due to the wide baseline nature of the 3D reconstruction subtask of the ACROBOTER project, where only four cameras mounted on the four respective corners of the ceiling must provide the coordinates of every object inside the scene, voxel-based algorithms can not deliver accurate results and robust measurements. Thus, a system similar to [17], where point-wise similarity measures for two consecutive views are used, has been expanded to trifocal plane in order to meet the requirements of the task.

Our system gradually combines the strengths of a point-wise similarity measure and a discriminant feature descriptor to deal with the huge amount of information and acquire the desirable result. First of all, we extract features from four views. Due to the unique position of the cameras we need an algorithm that can deal with extreme image tranformations and extract robust feature points. Thus, the use of an algorithm which can provide both quality features and robust matching among them is considered. The ASIFT [18] descriptor can provide the great results of the SIFT algorithm even when affine transformations will occur. The features of different images are then compared using the similarity function of SIFT and lists of potential matches are established. Based on these matches

the relations between the views are computed. We first compute the relationship among the two images which can give us imformation about the exact position of the points among those two views. The next step is the application of trilinear constraints in order to process additional information from another camera. The trifocal tensor is used because of its unique characteristic to transfer corresponding points of the two views to the corresponding point in a third view. Thus, we attain robust point correspodences between the three views of the four cameras in the room. Finally, we compute trifocal tensors over the four triplets of images and a dense point cloud is gradually produced, providing exact information about the position of the objects inside the scene.

In Fig 3 an example of three views from our tests is illustrated. Three wide baseline views are shown at top left side which are the original images from the cameras. Below them the corresponding points between them are depicted and the final 3D point cloud after the application of the trifocal tensor is shown on the right part of the figure.

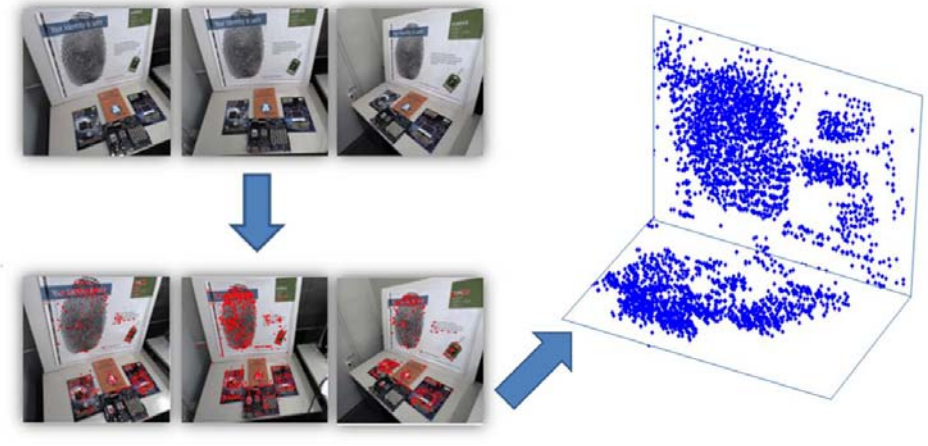


Fig. 3. Three views of the cameras on the top left, three views showing the correspodences between them on the bottom left and the final 3D point cloud after the trifocal tensor is applied

5 ACROBOTER's Object Recognition Task

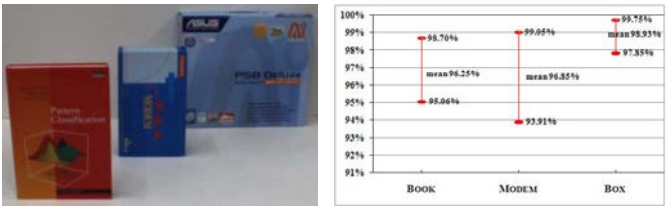
The main idea of local appearance-based recognition methods is the decryption of locally visual information. Thus, the indispensable visual distinctiveness of an object in a scene is ensured by locally sampled descriptions. The efficiency of algorithms based on features is directly related to the maintenance of this regional-based data. Furthermore, the two main submechanisms of such frameworks are the detectors and descriptors of areas of interest. The latter provide special attributes such as, insensitivity against rotation, scale, illumination or

viewpoint changes. Generally, the main idea behind interest point detector is the pursuit of points or regions with unique information in a scene. These spots or areas contain data that distinguish them from others in their local neighborhood. The most important methods of this field are Harris Corner detector [19], Scale Invariant Feature Transform (SIFT) [12] and Speeded Up Robust Features (SURF) [20]. On the other hand, the descriptor organizes the information collected from the detector in a discriminating manner. Thus, locally sampled feature descriptions are transformed into high dimensional feature vectors. From time to time, several approaches that implement a descriptor have been proposed. The most important are the Moment Invariants [21], Gradient Location and Orientation Histogram (GLOH) [22] and SURF [20]. For the purposes of ACROBOTER SIFT was adopted and expanded in order to estimate the distance between camera and objects found in a scene.

ACROBOTER's object recognition task is based on the four cameras installed in the four corners of the room. The main idea behind the proposed method is to maintain SIFT's properties whilst we make an attempt to enhance them. Thus, we have constructed a large database containing images from several objects. With a view to database's enrichment, these objects were photographed from different viewpoints and distances from the camera. Moreover, we used SIFT's matching sub-procedure to build an on-line scene search engine. Estimations derived from this engine are taken into account for the position estimation task. Initially, for each image in the database keypoint features are extracted, using SIFT. Then, the training session, where each object is photographed at different distances from the camera that are stored for further exploitation, takes place. Afterwards, the matching sub-procedure of SIFT is performed, where one image representing the scene is compared with several others (one per sample), representing the object from different viewpoints. Moreover,the features' centers of mass in both images are calculated. Next, the distance, measured in pixels with the use of Euclidian Distance, of each keypoint from the center of mass is calculated. Finally, by taking into account camera's distance from the object during the training session we are able to estimate objects' distance from the camera.

In Fig. 4(a) a scene that contains three different objects (e.g. a book, a modem's box and a motherboard's box), is illustrated[2]. With a view to reader's better understanding, objects found in the scene are referred as book, modem and box, respectively. During the training session, where the system remained offline, each object was captured separately from different viewpoints under varying illumination and geometrical (distance from the camera) conditions. Afterwards, and when the system started, the on-line scene search engine came into operation. The proposed method was evaluated through exhaustive tests containing several scenes and objects. The results of the recognition process were remarkable (approximately 98%), whilst the ones concerning the position estimation procedure for the three aforementioned objects are shown in Fig. 4(b).

[2] The ACROBOTER project is still under development. Due to the limited space only a part of the experimental results is demonstrated.



(a) (b)

Fig. 4. (a) The three objects (book, box, modem) used for evaluation purposes. (b) Accuracy of ACROBOTER's position estimation algorithm.

6 Conclusion

The basic structure and the vision system of the ACROBOTER project has been presented in this paper. The ACROBOTER's system is capable of: estimating robot's pose in the room; reconstruct the 3D working space and; recognize objects with remarkable efficiency. For each aforementioned process we implemented methods that are beyond the current state-off-the-art. Moreover, although the project is ongoing, the first experimental results prove that the overall efficiency of the developing vision system ranges at very high standards. As a result, and with a view to the integration with other systems of the project, we believe that the ACROBOTER will make a great breakthrough at the field of autonomous mobile assistant robots.

References

1. Balaguer, C., Gimenez, A., Huete, A., Sabatini, A., Topping, M., Bolmsjo, G.: The MATS robot: service climbing robot for personal assistance. IEEE Robotics & Automation Magazine 13(1), 51–58 (2006)
2. Sato, T., Fukui, R., Morishita, H., Mori, T.: Construction of ceiling adsorbed mobile robots platform utilizing permanent magnet inductive traction method. In: Proceedings of 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2004), vol. 1 (2004)
3. Xu, D., Li, Y.F.: A new pose estimation method based on inertial and visual sensors for autonomous robots. In: IEEE International Conference on Robotics and Biomimetics, 2007. ROBIO 2007, pp. 405–410 (2007)
4. Zufferey, J.C., Floreano, D.: Fly-inspired visual steering of an ultralight indoor aircraft. IEEE Transactions on Robotics 22(1), 137–146 (2006)
5. Gemeiner, P., Einramhof, P., Vincze, M.: Simultaneous motion and structure estimation by fusion of inertial and vision data. Int. J. Rob. Res. 26(6), 591–605 (2007)
6. Shademan, A., Janabi-Sharifi, F.: Sensitivity analysis of EKF and iterated EKF pose estimation for position-based visual servoing. In: Proceedings of 2005 IEEE Conference on Control Applications, 2005. CCA 2005, pp. 755–760 (August 2005)
7. Kyriakoulis, N., Karakasis, E., Gasteratos, A., Amanatiadis, A.: Pose estimation of a volant platform with a monocular visuo-inertial system. In: IEEE International Workshop on Imaging Systems and Techniques, IST, pp. 1–6 (2009)

8. Strecha, C., Fransens, R., Gool, L.V.: Combined depth and outlier estimation in multi-view stereo. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 2394–2401 (2006)
9. Kolmogorov, V., Zabih, R.: Multi-camera scene reconstruction via graph cuts, pp. 82–96 (2002)
10. Tola, E., Lepetit, V., Fua, P.: A fast local descriptor for dense matching, Alaska, USA (2008)
11. Liao, M., Wei, L., Chen, W.: A novel affine invariant feature extraction for optical recognition, vol. 3, pp. 1769–1773 (August 2007)
12. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
13. Open Source Computer Vision (OpenCV) home page, `http://sourceforge.net/projects/opencvlibrary`
14. Welch, G., Bishop, G.: An introduction to the kalman filter. Technical report, Chapel Hill, NC, USA (1995) (Revised, July 2006)
15. Svoboda, T., Hug, H., Gool, L.V.: Viroom - low cost synchronised multicamera system and its self-calibration. LNCS, pp. 512–522. Springer, Heidelberg (2002)
16. Saito, H., Baba, S., Kimura, M., Vedula, S., Kanade, T.: Appearance-based virtual view generation of temporally-varying events from multi-camera images in the 3d room, pp. 516–525 (October 1999)
17. Pollefeys, M., Van Gool, L., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J., Koch, R.: Visual modeling with a hand-held camera. International Journal of Computer Vision V59(3), 207–232 (2004)
18. Morel, J., Yu, G.: ASIFT: A New Framework for Fully Affine Invariant Image Comparison. SIAM Journal on Imaging Sciences 2(2), 438–469 (2009)
19. Rothganger, F., Lazebnik, S., Schmid, C., Ponce, J.: 3d object modeling and recognition from photographs and image sequences, pp. 105–126 (2006)
20. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (surf). Computer Vision and Image Understanding 110(3), 346–359 (2008)
21. Van Gool, L., Moons, T., Ungureanu, D.: Affine/photometric invariants for planar intensity patterns, pp. 642–651. Springer, London (1996)
22. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. Mach. Intell. 27(10), 1615–1630 (2005)

Prediction Surface Topography in Flank Milling

Wei Hao[1] and Xiao-Jin Wan[2,*]

[1] Guangdong Vocational College of Mechanical and Electrical Technology
[2] State Key Laboratory of Digital Manufacturing Equipment and Technology of China, Huazhong University of Science and Technology, Wuhan 430074, P.R. China
Tel.:+862787557843
wxj_2001@163.com

Abstract. Machined surface texture is very critical factor since it directly affects the surface quality and part mechanical behavior, especially, the surface roughness. In this paper, a new model is presented to build kinematic relationship between the arbitrary point of the cutter edge and the arbitrary point of the machining feature, the ruled surface texture is first evaluated by means of calculating an deviation value along a normal direction of an arbitrary point on nominal part surface in terms of the proposed model, three dimension profile can also be simulated, two examples are used to verify the feasibility of the proposed model.

Keywords: Flank milling, Ruled surface, Geometric error, surface roughness.

1 Introduction

The qualities of die/mold and machined parts are determined by their surface roughness and the form accuracy. As the life cycle of mold products reduces, lead time reduction together with high quality assurance becomes an important issue in manufacturing industry. In this regard, high speed machining is widely used for the high productivity. In high speed machining, feed per tooth and pick feed are very small compared to those in general machining. As a result, it is necessary to study the surface texture characteristics in the microscopic view. End milling is a prevalent machining process used extensively in the manufacturing industry to make flat surfaces of precision die/mold and other mechanical parts. Simulations of machined surface texture constitute an active research topic in the manufacturing community. Relevant published research work can be summarized as follows. Kline et al. [1] discussed the effects of cutter runout on the shape of the tooth marks in end milling process. Jung et al. [2,3] developed the so-called ridge method to predict the characteristic lines of the cut remainder for a disk tool in the ball-end milling process, and three types of ridges are defined to this end. Imani et al. [4], Imani and Elbestawi [5], and Sadeghi et al. [6] used solid modeling techniques and Boolean operations to

* Corresponding author.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 967–975, 2009.

deal with the geometric simulation of the ball-end milling operations. In summary, many researchers employed discretization and interpolation techniques to simulate the machined surface texture. Elbestawi et al. [7] and Ismail et al. [8] studied the trochoid path for surface generation of end milling. The tool path is discretized into segments to simulate the surface texture. Based on the concepts of parallel reference section levels and elementary linear sections, Bouzakis et al. [9] modeled the workpiece and cutting edge to simulate the texture in the ball-end milling process. Furthermore, Ehmann and Hong [10], Yan et al. [11] and Lazoglu [12] applied a similar method to the ball-end milling process. Li et al. [13] formulated the trajectory equations of the cutting edge relative to the workpiece and simulated the surface texture in the end milling process. Antoniadis et al.[14] determined the machined surface roughness for ball-end milling, based on shape-function interpolation over a number of finite linear segments of the workpiece. Zhang et al. [15] developed a algorithm to predict the machined surface topography and roughness in multi-axis ball end milling of plane and sculptured surfaces. In fact, all above methods aim to plane, though, [15] dealt with sculptured surface, they assume that tool swept path was given in advance, however tool swept path is very different for obtaining, as far as we known, until now, ruled surface topography is not developed, under motivation, we develop model and simulate the ruled surface topography.

This paper presents a general model to simulate the machined surface texture and roughness in milling process. First, the trajectory equation system of the cutting edge relative to the workpiece in the milling process is formulated with an illustration of the height of the cut remainder. Then, numerical methods are developed to solve the equation system for end and flank milling. Finally, some examples are studied to evaluate the texture and roughness. Results are compared to corresponding experimental ones. In addition, the developed algorithm has also the advantage of determining the tool position whenever the machined surface is generated in any desired node. This will be helpful in the prediction of form errors due to machining deformation.

2 Kinematic Model

First, consider an arbitrary contour mill and machined surface illustrated in Fig. 1. Under the premise of disregarding the influence of material properties of the workpiece and the tool, and vibration of the machine system, the machined surface texture mainly depends on the tool geometry, tool path, and spindle runout. In order to facilitate the description of the relative motion relationship between the blade of the cutter and workpiece in the milling process, a set of coordinate systems are established a priori, as shown in Fig. 1. $O_W x_W y_W z_W$ represents a reference coordinate system fixed on the workpiece. $O_A x_A y_A z_A$ is the local coordinate system fixed on the main shaft of the milling machine. $O_C x_C y_C z_C$ is the local coordinate system fixed on the cutter. The cutter revolves round the spindle, i.e., axis $O_A Z_A$, with the angle speed . $O_C x_E^j y_E^j z_E^j$ is defined as the local coordinate

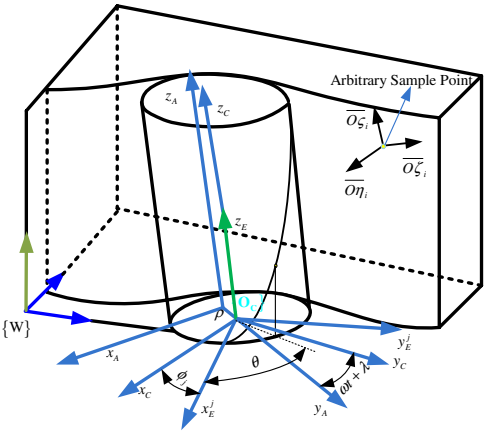


Fig. 1. Coordinate systems configuration in the milling process

system attached to the *j*th cutting edge. $O\zeta_i\xi_i\eta_i$ is the local coordinate system whose origin is located at sample point P on the machined surface with being the normal vector of point P.

In addition, $\mathbf{f}=\left(f_x\ f_y\ f_z\right)^T$ is the feed vector, ϕ_j is the angle between the axis $O_C x_E^j$ and axis $O_C x_C$ .In the coordinate system $O_C x_E^j y_E^j z_E^j$, let and z be curvilinear parameters, then the coordinates of a given point P on the jth cutting edge in terms of the geometrical characteristics of the cutting edge can be expressed as

$$\begin{pmatrix} x_E^{P_{i,j}} \\ y_E^{P_{i,j}} \\ z_E^{P_{i,j}} \end{pmatrix} = \begin{pmatrix} R\cos\left(\theta_{i,j}^P\right) \\ R\sin\left(\theta_{i,j}^P\right) \\ R\theta_{i,j}^P \big/ \tan\left(\gamma_h\right) \end{pmatrix} \tag{1}$$

where, β is the helix angle of the cutter, R is the diameter of the cutter.

For the purpose of transforming the point on the cutter edge into the workpiece coordinate frame, the corresponding transformation procedures are given as follows:

The coordinate frame of the cutter edge $Ox_E^j y_E^j z_E^j$ in cutter coordinate frame $\mathrm{O_w}x_\mathrm{w}y_\mathrm{w}z_\mathrm{w}$ can be expressed as follows:

$${}^C\mathbf{H}_E\left(Z_T,\varphi_j\right)=\begin{bmatrix} \cos\left(\varphi_j\right) & -\sin\left(\varphi_j\right) & 0 & 0 \\ \sin\left(\varphi_j\right) & \cos\left(\varphi_j\right) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2}$$

where, $\varphi_j = \varphi_0 + \dfrac{2\pi(j-1)}{N_t}$, is the position angle of the jth edge in coordinate cutter coordinate frame, φ_0 is an initial value, N_t is the number of the cutter edges. Similarly, the cutter coordinate frame $\mathrm{O}_C x_C y_C z_C$ in the coordinate frame of the main shaft $\mathrm{O}_A x_A y_A z_A$ can be described as follows:

$$ {}^{S}\mathbf{H}_C(t) = \begin{bmatrix} \cos(\omega t+\lambda) & -\sin(\omega t+\lambda) & 0 & -\rho\cos(\omega t+\lambda) \\ \sin(\omega t+\lambda) & \cos(\omega t+\lambda) & 0 & -\rho\sin(\omega t+\lambda) \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3} $$

The coordinate of main shaft $\mathrm{O}_A x_A y_A z_A$ in the coordinate frame of the workpiece can be expressed as follows:

$$ {}^{W}\mathbf{H}_A(t) = \begin{bmatrix} 1 & 0 & 0 & f_x t + p_x^s \\ 0 & 1 & 0 & f_y t - (R+\rho) + p_y^s \\ 0 & 0 & 1 & f_z t + p_z^s \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{4} $$

in order to obtain relative motion relationship of the cutting edge point with respect to the machining feature point on the nominal part surface, the detailed procedure is given as follows:

Firstly, assume that the part surface is regular surface which is a C^2 type of surface and is differentiable or smooth, is parameterized as

$$ \begin{cases} \mathbf{x} = \mathbf{x}(u,v) \\ \dfrac{\partial \mathbf{x}}{\partial u} \times \dfrac{\partial \mathbf{x}}{\partial v} \neq 0 \end{cases} \tag{5} $$

Where, u and v are curvilinear coordinates of a surface, according to intrinsic properties of a surface, tow tangent vectors and normal vector of an arbitrary point on a nominal surface can be obtained as

$$ \begin{cases} \mathbf{t}_{it1}\left(r_{it1}^{11}, r_{it1}^{12}, r_{it1}^{13}\right) = \dfrac{\partial \mathbf{x}/\partial u}{\left\|\partial \mathbf{x}/\partial u\right\|} \\ \mathbf{t}_{it2}\left(r_{it2}^{21}, r_{it2}^{22}, r_{it2}^{23}\right) = \dfrac{\partial \mathbf{x}/\partial v}{\left\|\partial \mathbf{x}/\partial v\right\|} \\ \mathbf{n}_{in}\left(r_{in}^{31}, r_{in}^{32}, r_{in}^{33}\right) = \dfrac{(\partial \mathbf{x}/\partial u)\times(\partial \mathbf{x}/\partial v)}{\left\|(\partial \mathbf{x}/\partial u)\times(\partial \mathbf{x}/\partial v)\right\|} \end{cases} \tag{6} $$

These vectors, in Eq. (8), $\mathbf{t}_{it1} \in \Re^{3\times1}$, $\mathbf{t}_{it2} \in \Re^{3\times1}$ and $\mathbf{n}_{in} \in \Re^{3\times1}$ are used to define unit coordinate frame of an arbitrary point $P_i^f\left(x_i^f, y_i^f, z_i^f\right)$ on a nominal surface, right hand rule is adopted as shown in Fig. 1. the specific form is expressed as follows:

$$\mathbf{t}_{it1} = \overline{O\varsigma_i} \quad \mathbf{t}_{it2} = \overline{O\xi_i} \quad \mathbf{n}_{in} = \overline{O\eta_i} \tag{7}$$

At the same time, they are also used to construct transformation matrix which determine position and orientation of a current feature point coordinate frame in workpiece coordinate frame, three Eular angle (α, β, γ) consists of orientation matrix is given as

$${}^{W}R_f^{P_i}(\alpha,\beta,\gamma) = \begin{bmatrix} c\alpha c\beta c\gamma - s\alpha s\gamma & -c\alpha c\beta s\gamma - s\alpha c\gamma & c\alpha s\beta \\ s\alpha c\beta c\gamma & -s\alpha c\beta s\gamma & s\alpha s\beta \\ -s\beta c\gamma & s\beta s\gamma & c\beta \end{bmatrix} \tag{8}$$

For a given point on a nominal part surface, a corresponding to coordinate frame $O\zeta_i\xi_i\eta_i$, in workpiece coordinate frame, has an corresponding orientation matrix which can be expressed as follows:

$${}^{W}\mathbf{R}_f^{P_i}\left(\mathbf{t}_{it1}, \mathbf{t}_{it2}, \mathbf{n}_{in}\right) = \begin{bmatrix} r_{it1}^{11} & r_{it2}^{21} & r_{in}^{31} \\ r_{it1}^{12} & r_{it2}^{22} & r_{in}^{32} \\ r_{it1}^{13} & r_{it2}^{23} & r_{in}^{33} \end{bmatrix} \tag{9}$$

By combining Eqs. (5)-(9), a transformation matrix which is used to describe position and orientation of workpiece coordinate frame $\mathrm{O_W}x_\mathrm{W}y_\mathrm{W}z_\mathrm{W}$ with respect to a nominal feature point coordinate frame $O\zeta_i\xi_i\eta_i$, on a nominal part surface can be obtained as follows

$${}_f^{P_i}\mathbf{T}_W = \begin{bmatrix} \left({}^{W}\mathbf{R}_f^{P_i}(\alpha,\beta,\gamma)\right)^T & -\left({}^{W}\mathbf{R}_f^{P_i}(\alpha,\beta,\gamma)\right)^T \mathbf{P}_{f_i}^W \\ 0 & 1 \end{bmatrix} \tag{10}$$

where, ${}_f^{P_i}\mathbf{T}_W \in \Re^{4\times4}$ is a transformation matrix, $\mathbf{P}_{f_i}^W \in \Re^3$ is a position vector of a nominal feature point in workpiece coordinate frame $\mathrm{O_W}x_\mathrm{W}y_\mathrm{W}z_\mathrm{W}$.

In side milling ruled surface, the feed velocity is perpendicular to the normal vector of a nominal part surface, a dot product of feed speed $\mathbf{f}$ and normal vector is expressed as follows

$$\mathbf{n}_{in} \cdot \mathbf{f} = 0 \tag{11}$$

By combining Eqs. (1)-(4), Eq. (10) and Eq. (11), an explicit, kinematical trajectory expression of cutter edge point relative to machining feature point can be expressed as

$$\begin{pmatrix} x_{E_{AP}}^{F_{AP}} \\ y_{E_{AP}}^{F_{AP}} \\ z_{E_{AP}}^{F_{AP}} \end{pmatrix} = \begin{pmatrix} \sqrt{\left(r_{it1}^{11}\right)^2+\left(r_{it1}^{21}\right)^2}\left[R\cos\left(\omega t+\lambda-\phi_j-\theta_{i,j}^P+\alpha\right)-\rho\sin\left(\omega t+\lambda+\alpha\right)\right] \\ +r_{it1}^{31}\dfrac{R\theta_{i,j}^P}{\tan\gamma_h}+\mathbf{t}_{it1}\left(\mathbf{f}t+\mathbf{P}_S-\mathbf{P}_N^i\right) \\ \sqrt{\left(r_{it2}^{12}\right)^2+\left(r_{it2}^{22}\right)^2}\left[R\cos\left(\omega t+\lambda-\phi_j-\theta_{i,j}^P+\beta\right)-\rho\sin\left(\omega t+\lambda+\gamma\right)\right] \\ -r_{it2}^{32}\dfrac{R\theta_{i,j}^P}{\tan\gamma_h}+\mathbf{t}_{it2}\left(\mathbf{f}t+\mathbf{P}_S-\mathbf{P}_N^i\right) \\ \sqrt{\left(r_{in}^{13}\right)^2+\left(r_{in}^{23}\right)^2}\left[R\cos\left(\omega t+\lambda-\phi_j-\theta_{i,j}^P+\gamma\right)-\rho\sin\left(\omega t+\lambda+\gamma\right)\right] \\ -r_{in}^{33}\dfrac{R\theta_{i,j}^P}{\tan\gamma_h}+\mathbf{n}_{in}\left(\mathbf{P}_S-\mathbf{P}_N^i\right) \end{pmatrix} \tag{12}$$

Therefore, we obtain a general model of the relative motion of an arbitrary point of the cutting tool with respect to corresponding to an arbitrary point of the machining feature, since the cutting edge has to contact with the machined surface at all through time, and then we can obtain the contact constraint equation system in terms of Eq. (12), their specific forms are given as follows

$$\begin{cases} x_{E_{AP}}^{F_{AP}}\left(R,\omega,N_t,\lambda,\phi_j,\theta_{i,j}^P,\alpha,t\right)=0 \\ y_{E_{AP}}^{F_{AP}}\left(R,\omega,N_t,\lambda,\phi_j,\theta_{i,j}^P,\beta,t\right)=0 \end{cases} \tag{13}$$

in milling, cutting edge trajectories sweeping part surface are a family of cycloid surface associated with definite position relationship each other, therefore, based on the geometrical characteristics of the cutting edge and tool path, the topography corresponds to the $z_{E_{AP}}^{F_{AP}}$ value with such $\theta_{i,j}^P$ and t satisfying the following equation system: by means of the specified initial conditions and solving Eq. (13) and obtaining a deviation along a normal direction of an arbitrary point on nominal part surface, texture can be simulated; the detailed application will be illustrated in cases of two examples in following section.

3 Case Studies

In two simulations, plane side and curve milling as shown in Fig. 2 and Fig. 3 are simulated. Milling conditions listed in Table 1 are selected not only for machining requirement but also for more observability and feedrate is deliberately enlarged. The Prediction results given in Fig. 4 show that the up and down milling processes result in different surface textures. From Table 2, it is noted that the simulated R_a (surface roughness) associated with up milling is less than that of the down milling. This result agrees with the basic knowledge of machining experiences. Finally, effects of various machining parameters, such as cutter radius and feed speed, on the machined surface roughness are investigated. For the up milling process, it is shown in Fig. 4 that an

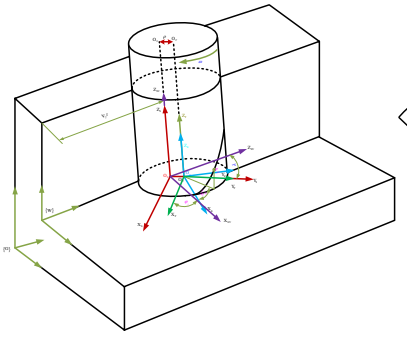


Fig. 2. Plane milling in machining process

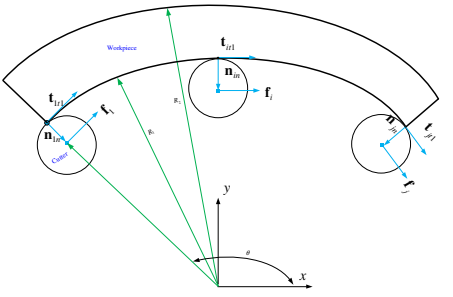


Fig. 3. A curve ruled surface

Table 1. Milling conditions in the end milling process

methods	R	N_t	β	ω	f_t	a_p	a_e	ρ
plane	3	4	40^0	800rev/min	0.168	1mm	21mm	0.004
curve	4	4	40^0	800rev/min	0.85	1mm	10mm	0.002

increase of the cutter radius will lead to a decrease of surface roughness. Contrarily, an increase of the feed rate will lead to an increase of the roughness as shown in Table 3 (R_a is arithmetical average deviation, R_z is ten point height of irregulartlies, R_y is maximum height of the profile). These numerical results conform to practical experiences. It is noted that with the increase of the feed speed and decrease of the cutter radius are the same results. The reason is that the effect of the relative movement of the cutting edge and the workpiece to the surface roughness will increase, correspondingly, with those changes. Consequently, effects of other factors, e.g., vibration and tool wear, will decrease. Another important comment has to be made for the above results. However, some other unconsidered factors are not always sure to increase the roughness. For example, the influence of flank wear on the roughness was studied in [8].

Since curved surface topography is not to be easily observed, all of deviations along normal direction of each sampled point on a nominal surface are up to 10x magnification of the image displayed in Fig. 5 (a) and Fig. 5(b).

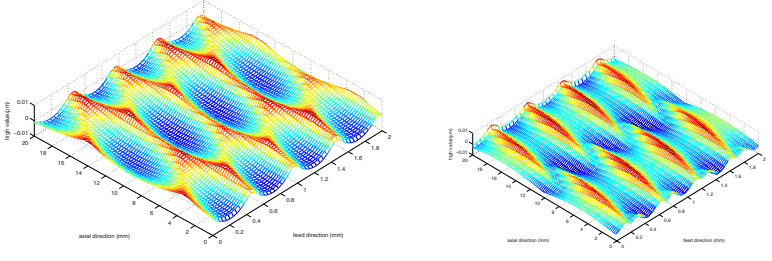
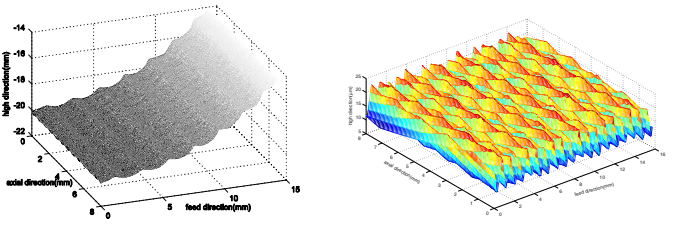


Fig. 4. Down(Up) milling surface texture of plane milling



(a) (b)

Fig. 5. Down milling surface texture of curve milling

Table 2. Simulation results in the end milling process (μm)

Milling type	R_a	R_z	R_y
Up milling	3.616	7.228	7.195
Down milling	11.995	10.669	10.620

Table 3. Simulation results in the end milling process (μm)

Spindle rotation speed	f_t	R_a	R_z	R_y
450	0.15	0.398	0.890	0.890
900	0.3	0.435	1.675	1.675
450	0.3	1.580	3.402	3.402

4 Conclusions

This paper proposed an effective model which is flexible, can be applied to not only plane surface but also curved surface. by means of the model, two simulations are performed to validate the correctness, one is a plane side milling which is used to demonstrate simulation texture agreeing with experiment, the other is a curve ruled surface flank milling which is used to verify the model applying to surface, simulation results are consistent with basic knowledge of machining experiences, which indicate the effectiveness of the proposed model.

Acknowledgements

This work is supported by the '973' National foundation Research Program of China (Grant No.2005CB724100), the National Natural Science Foundation of China (Grant No. 50775080 and 50835004), and the High-Tech '863' Plan Research and Development of China (grant No.2007AA04Z204).

References

[1] Kline, W.A., DeVor, R.E., Shareef, I.A.: The Prediction of Surface Accuracy in End Milling. ASME J. Eng. Ind. 104, 272–278 (1982)

[2] Jung, T.S., Yang, M.Y., Lee, K.J.: A New Approach to Analysing Machined Surfaces by Ball-End Milling, Part I: Formulation of Characteristic Lines of Cut Remainder. Adv. Manuf. Technol. 25, 833–840 (2004)

[3] Jung, T.S., Yang, M.Y., Lee, K.J.: A New Approach to Analysing Machined Surfaces by Ball-End Milling, Part II: Roughness Prediction and Experimental Verification. Adv. Manuf. Technol. 25, 841–849 (2004)

[4] Imani, B.M., Sadeghi, M.H., Elbestawi, M.A.: An Improved Process Simulation System for Ball-End Milling of Sculptured Surface. Int. J. Mach. Tools Manuf. 38, 1089–1107 (1998)

[5] Imani, B.M., Elbestawi, M.A.: Geometric Simulation of Ball-End Milling Operations. ASME J. Manuf. Sci. Eng. 123, 177–184 (2001)

[6] Sadeghi, M.H., Haghighat, H., Elbestawi, M.A.: A Solid Modeler Based Ball-End Milling Process Simulation. Int. J. Adv. Manuf. Technol. 22, 775–785 (2003)

[7] Elbestawi, M.A., Ismail, F., Yuen, K.M.: Surface Texture Characterization in Finish Milling. Int. J. Mach. Tools Manuf. 34(2), 245–255 (1994)

[8] Ismail, F., Elbestawi, M.A., Du, R., Urbasik, K.: Generation of Milled Surfaces Including Tool Dynamics and Wear. ASME J. Eng. Ind. 115(2), 245–252 (1993)

[9] Bouzakis, K.D., Aichouh, P., Efstathiou, K.: Determination of the Chip Geometry, Cutting Force and Roughness in Free Form Surfaces Finishing Milling, With Ball End Tools. Int. J. Mach. Tools Manuf. 43, 499–514 (2003)

[10] Ehmann, K.F., Hong, M.S.: A Generalized Model of the Surface Generation Process in Metal Cutting. CIRP Ann. 43, 483–486 (1994)

[11] Yan, B., Zhang, D.W., Xu, A.P., Huang, T., Zeng, Z.P.: Modeling and Simulation of Ball-End Milling Surface Topology. J. Comput.-Aided Des. Comput. Graphics 13, 135–140 (2001)

[12] Lazoglu, I.: Sculpture Surface Machining: A Generalized Model of Ball-End Milling Force System. Int. J. Mach. Tools Manuf. 43, 453–462 (2003)

[13] Li, S.J., Liu, R.S., Zhang, A.J.: Study on an End Milling Generation Surface Model and Simulation Taking Into Account of the Main Axle's Tolerance. J. Mater. Process. Technol. 129, 86–90 (2002)

[14] Antoniadis, A., Savakis, C., Bilalis, N., Balouktsis, A.: Prediction of Surface Topomorphy and Roughness in Ball-End Milling. Int. J. Adv. Manuf. Technol. 21, 965–971 (2003)

[15] Zhang, W.H., Tan, G., Wan, M., Gao, T.: A New Algorithm for the Numerical Simulation of Machined Surface Topography in Multiaxis Ball-End Milling. Journal of Manufacturing Science and Engineering 130, 1–11 (2008)

Cognition Dynamics: Time and Change Aspects in Quantitative Cognitics

Jean-Daniel Dessimoz

HESSO-Western Switzerland University of Applied Sciences,
HEIG-VD, School of Management and Engineering,
CH-1400 Yverdon-les-Bains, Switzerland
Jean-Daniel.Dessimoz@heig-vd.ch

Abstract. Cognition, and automated cognition i.e. cognitics, have recently been given a theoretical framework, whereby core concepts have been formally defined, along with metric units. Beyond intellectual exploration and cultural interest, this was also motivated by the goal to automate cognition, i. e. to carry over some of the cognitive processes and tasks from humans to machine-based artifacts. First achievements have led to a number of interesting new results. One of them is to make explicit and more evident than in the past the critical role, in cognitics, of time and change quantities, as well as of the coercive level of control actions, along with perturbations and system properties. After a brief reminder of essential definitions, the paper reports on these topics, describing and discussing a number of complementary aspects summarized in title: 1. time and abstraction levels, which lead to various time properties and possibly specific orders of stationarity; 2. time based difference between knowledge and expertise; 3. critical role of time in estimation of information quantities, and thereby in particular in the estimation of complexity and knowledge quantities; 4. explanation, straightforward in this context, of the apparent so-called "paradoxes of experts", in learning and forecasting; 5. importance of time in automation as well, more specifically in loop control, where some time properties of control path (including perception, decision and action phases) relatively to those of system behavior (here the system is the entity to be controlled) are critical for success; 6. time and changes interrelations, with necessity, for quantitative estimation, of considering other factors as well; 7. possible changes of system time properties, in the context of closed loop control, whereby large differences may occur with respect to those in natural (open loop) status, depending on action and perturbation intensities, as well as on possible overall non-stationarities and non-linearities; 8. driving causes for changes, and classical analogies in the context of human psyche and cognition dynamics. The paper illustrates discussions with concrete examples relating to Robocup-at-Home competition tests and applications.

1 Introduction

Progress in automation reaches applications where cognitive abilities are critical. An example of domain where this is evident is the domain of service robots, and in particular systems to help at home, where basic cognitive abilities are necessary in terms

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 976–993, 2009.

of self initiatives, cooperative behavior and, generally speaking, human-robot-interaction. The league "at-Home" of the worldwide "Robocup" project is a particularly interesting field, where practical experiments and theoretical developments can complement each other, with participation of forces stemming from all world horizons [1,2].

Applications require technical excellence, and for this purpose appropriate bases in terms of definitions and metric system had to be developed for machine-based cognition [3,4]. Quantitative aspects have proven critical in a variety of domains; here are three examples: 1. in most applications signal processing techniques and software engineering approaches prove to deliver much more performances than conventional AI methodologies; 2. shortcomings have been made more evident than in the past, even for concepts long established as models and information [5]; 3. now this paper reports on another class of difficulties which become more apparent: traditionally we have kept a rather too static view of matters, especially in cognition, and in fact phenomena involve time and changes, which leads to dynamics and evolution; in this respect again, it helps a lot to be quantitatively oriented.

The reported work is believed to be unique in its commitment to provide a precise framework allowing for quantitative estimations of cognitive entities, with formal connection and full coherence, where applicable, with widely accepted theories and standard units (re. numerous references cited in previous publications (in particular [3-5]). Nevertheless other views about cognitive dynamics can also be found in existing literature [e.g. 6, 7].

The paper is organized as follows. Part 2 briefly reminds the basic definitions in cognition dynamics and automated cognition, cognitics. Part 3 is the main part and introduces a number of new aspects, ranging from elementary multimodal time properties, time varying amounts of complexity and knowledge, time-based difference between knowledge and expertise, to compound consequences, such as paradoxes of experts, stability of automated processes, correspondence between time and changes, as well as the possibility, with appropriate driving causes, to modulate the quantity of some critical time properties. Part 4 provides concrete examples for the points discussed so far, taken from Robocup-at-Home context.

2 Essential Definitions in Cognition and Cognitics

Let us briefly mostly remind basic definitions relevant for cognition dynamics. This will be done in three groups, relating first to central aspects of cognition and dynamics, then to immediate contextual concepts, the second group dealing with information and modeling, and finally the third one addressing basic time properties.

2.1 Cognition Domain and Dynamics

Starting from usual, individual definitions for cognition and dynamics, this section introduces the new cross concept of cognition dynamics, and reminds how cognitics and, rigorously, the concepts of knowledge and complexity have been delineated.

It is generally accepted, in particular, that “in psychology and cognitive science cognition refers to an information processing view of an individual's psychological functions” [8], and that the origin of the word is cognoscere (Latin) i.e. “to know”.

Similarly “dynamics refers to time evolution of physical processes”; it may also refer to “forces and motions” of objects; and the original Greek form of the word (dynamics) means ”power”.

In the paper, “cognition dynamics” will be essentially coherent with these common definitions, mapping in the first domain -cognition- concepts traditionally more typically associated with the physical world in the context of dynamics.

“Cognitics” has been coined in order to possibly make a distinction between human-based and machine-based individuals or systems. Cognitics is in principle reserved to the latter case; it relates to the domain of “automated” cognition.

In an attempt to describe rigorously cognition, which is in particular quite necessary in order to develop cognitics, axiomatic definitions have been elaborated along with an appropriate metric system [5]. This effort to explore cognition beyond purely qualitative aspects develops in the domain of “quantitative cognitics”.

Knowledge is the cornerstone of this rigorous construction. The word refers to the roots of cognition (“to know”); it refers to the property of cognitive systems to process information, and more specifically to deliver (output) information. Quantitatively, knowledge, K, is estimated as a function of the amounts of two flows of information: input information and output information; the unit is the “lin”.

Complexity is the property of requiring a large amount of information for description; quantitatively it can be estimated with the same unit as for information, i.e. “bit”.

Abstraction is the property of cognitive systems that yield less information than they receive. Reciprocally, concretization is the property of systems yielding more information than they receive.

2.2 Information and Modeling

Information and modeling are first classically introduced in the sequel. Then, known shortcomings are reminded. And a new view on the possible role of digital information is presented.

Information has been defined in the middle of 20th century on the basis of a probability calculus; the unit for metric purpose is usually the “bit”.

It turns out that in practice information cannot be estimated in finite terms directly on reality and always requires some modeling.

Models are always of relatively small complexity; nevertheless, specific models may be useful in order to reach pre-defined goals.

Historically, the information theory has been created for the field of machine communication. In this context, information involves relatively small amounts of information. When moving to human context, more than before, reality is in background. As reality is infinitely complex, typically, two humans are schematically required: 1. one to select relevant goals, and 2. the other one to provide appropriate corresponding models (Setting-up the correspondence between model and reality establishes the meaning, and is sometimes referred to as the grounding problem). Between those two schematic humans, digital representations are possible, and as a

matter of fact, they are now getting elaborated in very large quantities, contributing to a whole digital universe.

2.3 Basic Time Properties

Let's briefly remind the definition of some basic time-related concepts which will be referred to in the sequel of the paper: time, time constant and stationarity.

As everyone would agree, "time is a component of a measuring system used to sequence events, to compare the durations of events and the intervals between them, and to quantify the motions of objects" [8]; it is measured in seconds.

The time constant of a system is usually understood as the time it takes for a system to essentially reach its asymptotic response state after a starting excitation[1].

Stationarity is the property of time-invariance.

3 New Aspects in Cognition Dynamics

One of the core meanings of dynamics relates to evolution in time. It may be surprising for some observers that cognition and time might have deep connections with each other. In particular, cognition seems to relate essentially to information, i.e. to a non-physical entity; and on contrary, time seems to be central in the physical world. It will be shown however that in fact time plays a variety of important roles in cognitive domain. In addition, the power aspect of dynamics also reflects in some cognitive properties.

This part introduces a number of new aspects relating to cognition dynamics, ranging widely from some elementary multimodal time properties, time varying amounts of complexity and knowledge, time-based difference between knowledge and expertise, to more complex, compound consequences, such as paradoxes of experts, agility, stability of automated processes, correspondence between time and changes, as well as, finally, the possibility, with appropriate driving causes, to modulate the quantity of some critical time properties.

3.1 Order of Stationarity

Systems and processes may usually be viewed at various levels of abstraction. Depending on the level being considered, stationarity may strongly vary.

In general the lowest levels, more detailed, where no abstraction or very little abstraction is done, tend to vary more than higher levels.

For example, in learning systems, there is by definition no stationarity at lowest level, the level of elementary information flows, as improvements are under way; yet when considering more abstract levels, such as overall system architecture, or information processing paradigm, time invariance, i.e. stationarity typically applies.

There may also be significant variations in terms of time horizon; systems may be stationary for short term behavior and on the contrary not stationary on the long term.

[1] A common case in control is to define the time constant for a first-order system as the time it takes to reach 63% of its asymptotic response to an incoming step signal.

3.2 Time Based Difference between Knowledge and Expertise

Knowledge and expertise are probably the most essential entities in cognition[2]. Time is qualitatively present in the concept of knowledge, and even more present, indeed quantitatively present, in the one of expertise.

Consider here the intuitive definition for knowledge, i.e. the ability to "do right", to deliver the correct information. It is already clear that in principle knowledge is not *quantitatively* related to time. Referring to metric equations, time does not appear. Yet to process and deliver information however does imply time.

Now there are many cases in cognitive processes where time matters, or even is critical. "To do right and fast" is another quality than knowledge; it is the quality of experts. While the unit for knowledge is the "lin", the unit for expertise[3] (synonyms in natural language: skills, competences, know-how, etc.) is the lin per second, "lin/s".

3.3 Perishable Quantities of Complexity and Knowledge

It has been shown that information is, by nature, subjective and perishable. This has consequences on the properties of cognitive entities involving information.

Quantitatively, information is estimated on the basis of a probability. Probability is generally (and hopefully!) essentially related to external circumstances (objective aspect of uncertainties), but in all cases primarily dependent on receiver's expectations (subjective aspects related to receiver's own model). And it is the very role of information to update receiver's model. Therefore expectations vary in time: a repeated message is in principle useless.

Complexity is one of those cognitive concepts inheriting the features of information. For example a first explanation may require a lot of information. By definition, that explanation is complex. Yet it maybe simply repeated: "Idem".

Similarly, a cognitive system may feature a large amount of knowledge, requiring in particular a large amount of incoming information. Yet as time goes, information flows may decrease, and consequently the instantaneous amount of knowledge will also decrease, as quantitatively shown by metric equations.

3.4 About Expert Paradoxes

In AI, some experimentally verified properties of experts have been widely found to be paradoxical, as no evident theoretical background was there to explain them. Two of such paradoxes (probably the main ones) are presented here, along with quantitative cognitics and cognition dynamics rationales in order to help understand the phenomena. The first one is in essence the following one: The more humans know, the more they can learn. And the second one: Experts are worse than beginners at forecasting.

In the physical world, phenomena are often bound by an asymptotic limit and difficulties grow as we get close to it. For example compressing garbage, of filling up a

[2] The focus is set here on knowledge and expertise; nevertheless time and changes play also a role in some other cognitive concepts, such as, in particular, learning.

[3] Re. B-Prize to be done.

container, grow more difficult as progress is done. Yet on the contrary in the cognitive world, the more humans know, the more it seems that they can still learn further. As shown above, the subjective and perishable nature of information is inherited by core cognitive concepts, and in particular by complexity and knowledge. As people get more expert, the uncertain nature of incoming information, i.e. the quantity of incoming information tends to fade. Less complexity; less knowledge required for further processing. Experts have (subjectively) much less information to process than beginners, and therefore it is not a surprise that learning gets in principle easier for them.

Are experts really worse than beginners at forecasting? Maybe yes, and maybe not. If experts are experts, this means that in principle they perform "right and fast". Now forecasting refers to future situations. Therefore a critical parameter is here stationarity: for phenomena that benefit from some order of stationarity, a proven expertise level will tend to be maintained through time. Now, if on the contrary phenomena do not have an appropriate stationarity, it should not be, nor a surprise nor a paradox, that past expertise cannot be carried over through time. If cognitive domains are different in the past from in the future, it may even happen that at times a negative correlation develops, which perfectly explains how experts (experts in the context of historical circumstances) might perform comparatively worse than beginners[4] in the new circumstances.

3.5 Agility

Cognitive systems have a real importance in as much as they are followed by actions. The latter is even more essential for control systems. A useful concept in this context is the one of agility.

Agility combines the notion of time with the one of action (both words, agility and action, share the same Latin root: agere, i.e. in current English to act, to do, to make). In natural language, agility has a connotation of referring to animal or human motions in space. This character may however be generalized by analogy to other forms of action.

Quantitatively, let us define agility of a system as the inverse of its time constant (re. §2.3), which implies 1/s as a unit.

3.6 Stability of Automated Processes

Cognitics denotes the domain of automated cognition. In automation, the necessary approach for controlling systems affected by unpredictable perturbations includes an estimation of system state. This is commonly described as closed-loop control. Closed-loop control however is known to risk instability in some circumstances, and it is shown below that time properties play a crucial role in this regard.

Closed-loop control includes a broad range of situations. Schematically, three classes of situations may be encountered; the first two of them bring stable solutions, and the third one, instability. 1. It may sometimes be very easy to realize the control

[4] In some previous work, mention has already been made in substance that "Hazard may be an acceptable guide in domains that have never been explored before"; which tends to radicalize the paradox under review.

system, and the latter proves very effective. 2. Sometimes, tuning is more difficult, and performances, without being as good (fast, accurate, simple) as in the previous example, nevertheless remain of an acceptable level. 3. The third typical class of situations includes those cases where failure cannot be avoided: either the target system is hardly set into motion, or it enters oscillatory or erratic behavior even for constant target values.

Now an extremely interesting indicator for possible system (in-)stability consists in the ratio, Ar, of two agilities, the one of controller (including perception, decision, action and communication phases) versus the one of controlled load. With Ar larger than 20, a system typically belongs to class 1 above; otherwise the system belongs to class 2, unless Ar is smaller than 2, in which case it falls into the third, unstable, case[5].

Here again, some time properties turn out to be a critical parameter or, at least, provide a critical indicator for successful system behavior in automation and cognitics domain.

3.7 Time and Change

Time is usually considered as a specific dimension of reality and is given its own unit (second, in the SI international unit system). But other views may be useful. For example the ancient Greek Parmenides invites us to consider reality as a permanent whole without any dimensions at all (what is, is); and in this United Nations Year of Astronomy, it is particularly evident that yet another model of reality where time and space are dependent of each other may also have some merits (in astronomy distances are counted in years; and reciprocally far objects are old ones).

Here time is discussed in its interrelations with change.

Intuitively, it appears that time can be defined in reference to change. For example a cycle of natural light change typically somehow defines a day long time duration.

More rigorously, and quantitatively, a change is not sufficient per se to define time, and other factors need be expressed. There are numerous common laws (equations) in physics or in cognition where time appears, and thus which provide as many ways to define time. Time may for example be precisely related to 1. a change in space by a speed factor, 2. a change in energy by a power factor, 3. a change in momentum by a force factor, or 4. a change in knowledge by an expertise factor[6].

3.8 Closed Loop Control, Consequences on Time Properties, and Autonomy

In fact and ultimately, the own ("natural") time properties of a particular system may not always prove to be the relevant criteria for success. This is especially true in the context of perturbations and of control; and even more so in common practical cases where non-stationarities and non-linearities prevail.

Powerful systems are complex and are usually organized as a multiplicity of interconnected subsystems (e.g. hierarchies, cascades, parallel or distributed structures, etc.). In this context, it is regularly verified that an elementary system may behave

[5] This can be straightforwardly deducted from previous publications, where the inverse of this ratio is used, i.e. the ratio of time constants [3].

[6] This is particularly true in differential terms.

with much improved time properties with the help of a dedicated, autonomous associated control system.

By this approach, far from representing absolute constraints, the natural time properties of an element may just be viewed as contingent features, which may drastically be improved at system level by appropriate design and engineering.

The paradigm of granting local autonomy is also very effective in improving the potential agility of control system in the closed loop.

3.9 Driving Causes for Changes in General, Consequences on Time Properties, and Analogies in Human Psyche

It has been reminded that dynamics imply time, changes and power (re. § 2.1). Relations between power, change and time are made here more explicit; then connections are established between power, energy, motion, forces and in general, driving cause for change; finally, classical analogies in the context of human psyche are presented (incl. motivation and emotional forces), which may be pertinent in inspiring yet new concepts in cognition dynamics and quantitative cognitics.

Time has been shown related by a power factor to a change in energy (re. § 3.7). Now energy is defined, in general, as the product of a driving cause by the corresponding effect. For example energy, in physics, may be estimated as the product 1. of a force by a distance, 2. of a voltage by some electric charges, or 3. pressure by a volume.

Notice that by the same token (energy is the product of a driving cause by the corresponding effect), since energy is the product of power by time, we could view power as a general cause for change, and time then becomes simply the corresponding effect!

For describing dynamics in human psyche, people have traditionally used analogies with the physical world: energetic person, powerful argument, etc.; a special place is given to mechanics, forces and motion: emotions and motivations are two words sharing the same Latin root: movere, to move.

Experience confirms that in the cognitive world as in the physical one, time is usually subjectively estimated as a function of changes; and that the factors linking changes to time are not always evaluated right; this principle leads sometimes to large errors in time estimation, especially when driving causes for changes are intense.

4 Examples

This section provides a large number of examples taken in the context of our participation to Robocup-at-Home competitions. First a very brief presentation of our (family of) robot(s) is made. Then examples follow in a sequence similar to the way concepts have been progressively presented above in this paper.

4.1 Brief Presentation of RH4-Y Robot

Developed since 2005, and following previous works in autonomous robots, which can be traced back to 1998, our RH-Y robot can move autonomously and interact cooperatively with humans in context of domestic type.

In terms of hardware, the robot consists typically in a platform mechanically driven in a classic way, i.e. with 2 active wheels, and a passive, free, third one; it carries an arm and hand (and also sometimes a 5 dof Katana arm), an electrical system with batteries, a number of sensors (Axis color camera, Mesa-imaging 3D range-finder, 2 Hokuyo planar laser rangefinders, a microphone) and actuators (Maxon DC servomotors for wheels, arm and hand, lights, sound system), a Fujitsu Siemens supervising computer, an embedded system featuring essentially a double star architecture, the first star being centered on an Ethernet hub and the second one on a USB hub, distributed Galil and Fiveco processors for motion control, and Beckhoff BC9000 PLC.

In terms of software, a key role is played by our Piaget language and environment, which features four progressive levels of complexity for users, extensive possibilities not only in programming and autonomous "run" modes but also in simulation, configuration, and interactive control modes. Implementation code is mostly done in Borland C++ object-oriented language (previous and alternate implementation languages include Pascal, C and C#). Ancillary functions are provided by various commercially available drivers and codes specialized for our COTS pieces of equipment and for elementary word recognition (MS SAPI 4.4). Motion control typically develops at 4 complementary levels, of respective agilities ranking progressively from approximately 1 to 100'000 1/s: programming and supervision, coordination, servo-control, and lowest level, encoder and motor management stage, including amplification and current limiting. Usually, position control is performed, and sometimes velocity control or force control laws are temporarily enacted.

Fig. 1. RH4-Y, version 4 of our cooperating robot for applications at home (photos HEIG-VD / PFG)

Fig. 2. Prof. Minoru Asada, President of Robocup, is watching RH2-Y, which successfully replicates one by one his gestures (moving blue boxes for the human, and white boxes for the robot) during « CopyCat » test, in Atlanta, July 2007. (Goal: teach motions just by showing; Photo Thomas Wisspeinter, Frauenhofer Gesellschaft).

Details, illustrations and videos can be found on our website [9].

4.2 Illustrative Case Study

This section revisits the top, more theoretical part of the paper , this time with a systematic illustration of introduced concepts. It is not the intention of authors to be either exhaustive nor even minimally complete for a real application, in the examples below, but rather to provide simple, representative samples, on the way to proceed. A careful attention must be given to the specific domain addressed by each example.

Service robots and cognitive tasks. In Robocup-at-Home league, the rulebook [e.g. 10] describes some representative tasks for a robot serving at home: "Follow a person[7]" and move according to a few possible orders, "Navigate" autonomously through home towards vocally ordered locations, "Fetch and Carry" a specific object, recognize "Who is who" among a few persons, find some lost object, learn from gestures in a "CopyCat" fashion, etc., leaving in addition some blank slots where novel applications may be presented ("Open Challenge", "Demos" and "Finals").

Quantitative aspects. Quantitatively, the rules provide contrasting levels of explicitness; e.g. the "Who is who" (WiW) test was involving 5 persons in 2007, which is numerically very clear; but no value is given in terms of minimal difference between individuals. Obviously it would be much easier to distinguish 1. a small fat silent boy wearing black attire from a tall, noisy, slim boy wearing yellow, than 2. two monozygotic twins wearing and behaving the same.

[7] Depending on the considered year, variations on the mentioned themes may occur, e.g. for what is presented here as "Follow a person" there is in fact "Follow and Guide"(2007), "Fast-Follow"(2008), "Follow Me"(2009), and yet another title for 2006.

Modeling. Let us consider in more details the who-is-who (WiW) example. Human intuition and expertise suggest that for this *goal*, i.e. in order to successfully perform the task, it is useful to break it in several subtasks. First it is necessary to detect a person, then to estimate a discriminating pattern, and finally to recognize the latter by reference to previously learned references. Let's concentrate on the first subtask – localization of a standing person: it is sufficient to represent a human by the reflective nature of its chest to the laser beam of a planar range finder mounted on the robot, in the vicinity of two 20cm or more empty spaces, one on each side. Considering that the chest is about 50cm wide, about 10 samples are required, with accuracy in position of about 10cm in "width" and distance. Let us in addition consider that the robot will walk around and that an elementary measurement would represent here the scanning of an area of about 1x1 m^2.

Grounding. Rules are described digitally in a book available on the internet in a very adequate manner. Yet for establishing the official correspondence between rulebook representations and actual "homes" and real world circumstances, as in Bremen (2006), Atlanta (2007), Suzhou (2008), as well as surely in the future in Graz etc. further human contributions are/will be required, as typically: Referees, Technical and Organizing Committees, Execs, and Team Leader Meetings.

Information. What is the *quantity* of information corresponding to an elementary measurement in the *domain* just described? The measurement consists in 10 samples which can each have one among 10 significantly different values (1 m with 10cm accuracy). Considering that samples are independent of each other and distances are all of equal *probability*, the quantity of information per measurement, Q, amounts to 33.2 bits.

$$Q = N_{samples} \cdot \log_2(N_{values}) = 10 \cdot \log_2(10) = 33.2[bit] \tag{1}$$

Complexity. Let us consider here two examples in estimation of complexity. 1. The first one addresses the domain of the description of the WiW test. Considering that the description provided in the rulebook consists in 652 words, and that in average a word conveys 10 bit of information, the complexity of this description is of 6520 bit. 2. Considering that the rulebook confines the persons to be recognized in a 4x4 m area of a home, i.e. 16 m^2, the complexity of the domain representing the location of persons, in accordance to above examples (re. § modeling, § information) amounts to 531.5 bits.

Knowledge. Consider the case of detecting the presence of a person in the cognitive domain introduced above (1 square meter of analyzed area, with coarse planar range estimation). As shown above, input information amounts to n_i=33.2 bit. The decision can be considered Boolean, and the probability of positive detection is of 5/16 (5 persons, with an individual area of about 1 square meter, out of a 16 square meter area; therefore the output information amounts to n_o=0.9 bit: \

$$n_o = \sum_{i=1}^{2} p_i \cdot \log_2(\frac{1}{p_i}) = 5/16 \cdot \log_2(16/5) + 11/16 \cdot \log_2(16/11) \cong 0.9 \ \text{[bit]} \tag{2}$$

With these values, the necessary knowledge for locating a person in the domain of an elementary measurement amounts to K= 33 lin.

$$K = \log_2\left(n_0 \cdot 2^{n_i} + 1\right) = \log_2\left(0.9 \cdot 2^{33.2} + 1\right) = 33.04 \quad \text{[lin]} \tag{3}$$

Abstraction. In the current example, detecting a human in an elementary WiW test, the cognitive process features an abstraction index, i_A= 33.2/0.9, amounting to 37.

Concretization. Imagine that a human has been detected at a distance of 1.5 m. The supervising stage will call for a forward motion of the robot by an amount of say 0.5 m. Consider, for simplification purpose here, just the cognitive task of the coordinating stage, moreover in its usual mode where feedback is restricted at higher and lower levels. On the input side, goal is given with a cm accuracy, in a range of about 12m; acceleration, speed limit, and deceleration are given with a 1% accuracy, which amounts in total to n_i=30.2 bit :

$$n_i = \log_2(1200) + 3 \cdot \log_2(100) \cong 30.2 \quad \text{[bit]} \tag{4}$$

On the output side, targets are computed for each of two wheels, interpolating higher level commands at a rate of about 30 samples per second. Considering that the accuracy is also equivalent to the order of 1cm in a 12 m range, and assuming that the motion last for 1 second, output information amounts to n_o=614 bit:

$$n_o = 2 \cdot 30 \cdot \log_2(1200) \cong 614 \quad \text{[bit]} \tag{5}$$

Under those assumptions, the concretization index amounts here to i_C= 614/30.2, i.e. to 20.3.

Time constant. In our motors, the time constant of coils is of about 1 millisecond. Imagine the following experiment: the (DC) motor is mechanically blocked; a voltage of 5V is applied (step signal), and the current is observed to climb towards an asymptotic value (say 2 A), which is reached within a 30% tolerance after 1 millisecond.

Stationarity. Let's consider several cases. 1. In the previous experiment (5V, 2A, etc.), after a short transient, all is stationary: applied voltage, measured current, coil under study. 2. Now consider another case: the motor is let "free, without load", e.g. with the robot wheel off the ground. The 5V voltage supplied to the motor sets the latter into motion, and consequently the 5V voltage of the example is commuted on successive coils by commutator and brushes. From a single coil point of view, the supply is not stationary any longer; it consists in an alternating square signal, with a certain duty cycle (e.g. 2ms/0V; 1ms/5V). At a more abstract level, the behavior may still be qualified of stationary, though in a different meaning: time-invariant velocity; time-invariant supply voltage and duty cycle. 3. Several levels of control higher, and even after integration of random perturbations and dialog components with the walker, there is still another order of stationarity which may qualify the paradigm of our robot, which dictates instantaneous changes in control modes (position, velocity, torque, recovery from errors, etc.), as in the current solution for "Follow Me" test.

Fig. 3. A view in Suzhou (China) of Robocup competition 2008: FastFollow challenge, with RH3-Y following its guide, later on crossing another team, and finally successfully finishing first the walk through home. (Goal: teach the robot a path just by guiding ; Photo HEIG-VD-PFG)

Expertise. While knowledge does not quantitatively depend on time, expertise does. Consider again the cognitive domain of elementary human localization, as studied in above paragraph (re. § Knowledge). For this task, in our system, 0.1 second are necessary to acquire input information and the subsequent computation time is comparatively negligible. The quantity of expertise in this domain, E, amounts therefore to 330.4 lin/s.

$$E = K/\Delta t = 33.04/0.1 = 330.4 \quad \text{[lin/s]} \tag{6}$$

Time-varying complexity and knowledge. As reminded above (re. §2.2 Information and modeling), information is a very perishable good, which has various consequences in cognitive domain. In the WiW test, humans are expected not to move. Therefore in this context the amounts of complexity and of required knowledge necessary for the detection of humans decreases very fast. It has been shown that initially 33.2 bit of information are required to describe the occupancy of 1 m^2 (re §Information); this is the initial value for the complexity of the description. Now we have also seen that the representation for occupancy after decision of the cognitive system consists only in 0.9 bit of information (re §Knowledge), which finally yields 33.04 lin of knowledge in the domain. This is true for the initial 0.1 s (re §Expertise). After that time, everything vanishes: input information drops to 0 bit, for the probability is "1" that identical messages repeat. Complexity drops to 0 bit ("no change"). It is

the same for the output decision (probability=1 for same output) and reassessing the knowledge equation (Equ. 3) with n_i=0 bit and n_o=0 bit yields 0 lin as well.

Expert paradoxes. The first expert paradox mentioned above states "The more experts know, the more they can learn[8]". As seen in the previous paragraph, a WiW beginner would start with raw information, i.e. 33.2 bit of information (ranger data). Now an expert would immediately deduce the detected state, which amounts to only 0.9 bit of information (occupancy). This provides a simpler basis for further cognitive processes, such as searching for other areas and persons with corresponding amounts in terms of incoming information, knowledge, expertise; and learning.

The second paradox introduced above (re. § 3.4 About expert paradoxes) states essentially that "Experts are worse than beginners at forecasting". Consider again the elementary WiW detection of a person. If circumstances change, and the cognitive domain evolves for example because of a new rule introduced, whereby humans should stand sidewise in front of robots instead of facing them, the raw data of beginners would still provide a sound basis for the new task, while the expert "shortcut" just mentioned in the previous paragraph would become obsolete and would systematically fail to detect humans.

As can be seen from these two examples, in the framework of current theory for quantitative cognitics, what used to appear as paradoxical, in fact turns out quite logical.

Agility. It follows from above definitions (re § 3.5 Agility) and for the example case about time constants (re. § *Time constant*) that the agility of our motors is of 1000 1/s.

Stability of automated processes. Three examples are provided here, corresponding each to a different type of strategies to recover from instability; close reference is made here to the agility ratio criterion mentioned above (re. §3.6 Stability). Examples will be taken in our application in the context of the Robocup-at-Home "Follow a person" test. Typically, for such an application there is a complex hierarchy of controls (levels A to E), which are briefly presented here: A. Linear and rotational motions as functions of walker position relative to robot (uppermost level, level 1; 2 parallel controls; multimode closed loop control -forward and rotational, proportional; reverse on/off; coordinated blocking, with recovery; etc.); B. MIMO stage, with inverse kinematics (level 2; 2 input and 2 output signals; parameterized gain matrix); C. Motion law stage (level 3; parameterized accelerations, 2 V-V controls); D. wheel velocity control (level 4; 2 PID closed loop control with encoder management); E. amplifier, current management (level 5; 2 on/off closed loop control). Consider now a first example, relating to level E: In our robots, there has been a time where while apparently behaving normally, motors were hot. Signal observation made it clear that a 20kHz 24 V peak-to-peak wave was present in addition to normal command curve on motor lines. In fact this seems to be quite a common problem, and a standard solution consists in increasing the time constant of the load; this in practice is done by adding an inductance serially to the motor, in the drive circuit. The agility ratio, Ar is thus improved, and brings the system in the class 1 region, i.e. in stable state.

[8] Notice that in the existing theory, learning is by definition the property of a cognitive system to increase its amount of expertise; and quantitatively, it is also the difference in amounts of expertise before and after.

Consider a second example, relating to levels 4 and 1: In both cases, the agility ratio turns out to correspond to class 2 situations as defined in §3.6; in this class of situations, the main parameter is the proportional gain between tracking error and correction command, which in principle should be as high as possible without oscillation (if at higher levels a smooth motion is desired, this is typically ensured by other means, namely a specific motion low – re. level 3). Consider a 3rd example 3: Beginners might imagine to control DC motors (e.g. time constant 1 ms) quite directly from supervising computer, even with the constraint of Ethernet TCP/IP protocol (time constant of about 0.1 s); these values yield an agility ratio Ar of 0.01, which corresponds to class 3, unstable systems; for a stable solution, modifications are required, additional resources are necessary, e.g. as further discussed below in § *Improving natural system time properties with control*.

Time and change interrelation. Generally speaking, a time duration can be estimated on the basis of a change in some other entity, in as much as a known conversion operator is available. Two examples follow: 1. Let us assign differently the role of unknown in the equation for elementary WiW subtask addressed above (re. § Expertise). In practice a time duration of 0.1s would correspond to waiting for a new decision about whether an elementary area is occupied or not, since knowledge amounts to 33.4 lin and expertise to 334 lin/s. 2. Changes in activity have the potential to describe time evolution. For very high resolution, no standard clock was available on commonly available computers. Our Piaget programming environment therefore traditionally features an internal counter ("Ticks"), which is incremented every time that a full cycle of agent activity has been completed, i.e. all agents have used a single individual time slice. The purpose is to statistically estimate very small delays, freely selectable in the range from 1microsecond to 10 milliseconds. This is possible because our agents take in average very thin time slices (about 100 nanoseconds) for their individual subtasks. Conversion from activity change to time duration is made on the basis of a computer and configuration dependant calibration. While dynamic calibration has also been practiced in this application, a static calibration, stationary in the time horizon of a run-time session or more, is usually preferred[9].

Improving natural system time properties with control. Reality is infinitely complex, and therefore it is always critically important to pay attention to the small domain under focus. Two examples follow, where natural time properties are drastically changed by appropriate control strategies; the second one also illustrates the concept of autonomy.

1. In the motor example presented above (re.§ *Time constant*), a 5V constant voltage was applied, and a 1 ms time constant had been measured on the current response, reaching 2A in asymptotic value. These performances per se are rather poor and would limit robot motions to low speed. Imagine now that you have a highly nonlinear control system, which simply, with high agility, switches the supply voltage to + or – Vmax, depending on the target current (2A) being reached (closed loop current control). If Vmax is 50V, i.e. 10 times higher than the voltage in the previous example, the resulting, effective, new time constant will be 10 times shorter! In our

[9] In current, C# development context, a 100 MHz system time basis is being used.

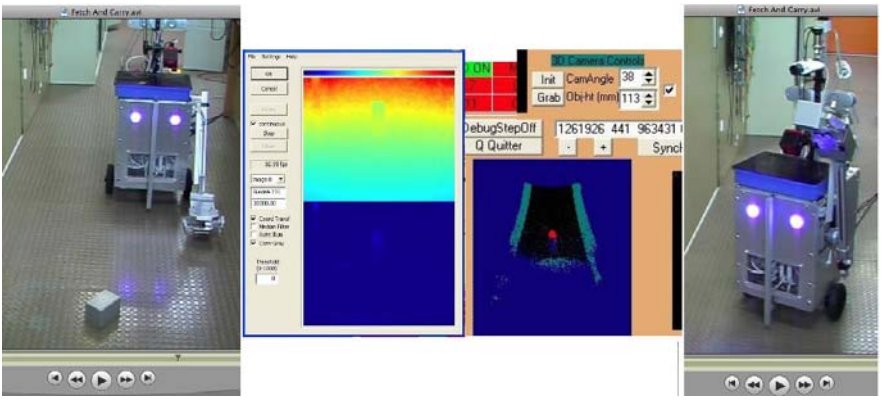


Fig. 4. RH4-Y moves, looking for an object (image 1, from left to right); it perceives its environment with rangefinder (img2), maps data in Cartesian space and detects the object (red spot in img3); finally it fetches and carries it to home user (photo HEIG-VD / AE)

robots, Vmax has usually a nominal value of 24 V; in the most recent platform, it is 48V.

2. Our early robots could rely on PC parallel ports and DOS or early Windows OS, where access to physical addresses, either in memory or on IO ports were fast, simple, and standard. In such a context the real-time control of multiple stepper motors and even DC motors was possible even with implementation on supervising computer, with excellent possibility of integration of low-level or intermediary-level phenomena (smart handling of blocking situations, multi axes coordination in versatile modes), as agility was extraordinary high (10^6 1/s); in recent years, architecture has evolved – communication is now based on Ethernet or USB hubs - with a number of advantages (numerous services of OS and OS-compatible multiple sources, convenient mechanical and electrical connections with external devices), but also a sharp drop in agility[10] (down to 10 1/s). While in the first case stable control could be achieved without extra resources, in the second case a hierarchy is required with an intermediary stage (incl. servo controllers and PLC) between former resources (between computer and DC motors, etc.), to which a certain level of autonomy is granted under normal circumstances.

Driving causes for changes, time properties, and analogies in human psyche. Changes are not always correctly mapped into time duration, and errors can be particularly large when strong emotions occur. As an example of the first, "ordinary" case, during the development of one of our early robots, possibly because of the unusual situation which was involving small, microtechnological scale, hours have been wasted in an effort to correct a bug which in fact did not exist: the change of three rotations of an actuator was psychologically perceived as lasting much longer than the correct 0.8 second which had been programmed. As an example of the second,

[10] Of course some exceptions do exist, where under special conditions much better agilities are theoretically possible, but in as much as possible standard solutions and conservative approaches should be preferred, which are compatible with the quoted value.

"emotionally laden" case, it has been reported that in emergency context such as an accident, because of unusual psychological turmoil occurring, time is temporarily perceived as being strongly shrunk or dilated.

5 Conclusion

Cognition, and automated cognition i.e. cognitics, have recently been given a theoretical framework, whereby core concepts have been formally defined, along with metric units. Beyond intellectual exploration and cultural interest, this was also motivated by the goal to automate cognition, i.e. to carry over to machine-based artifacts some of the cognitive processes and tasks usually performed by humans.

First achievements have led to a number of interesting new results. One of them is to make explicit and more evident than in the past the critical role, in cognitics, of time and change quantities, as well as of the coercive level of control actions, along with perturbations and system properties.

After a brief reminder of essential definitions, the paper has reported on cognition dynamics, describing and discussing a number of complementary aspects: 1. time and abstraction levels, which lead to various time properties and possibly specific orders of stationarity; 2. time based difference between knowledge and expertise; 3. critical role of time in estimation of information quantities, and thereby in particular in the estimation of complexity and knowledge quantities; 4. explanation, straightforward in this context, of the apparent so-called "paradoxes of experts", in learning and forecasting; 5. importance of time in automation as well, more specifically in loop control, where time properties of control path (including perception, decision and action phases) relatively to those of system behavior (here the system is the entity to be controlled) are critical for success; 6. time and changes interrelations, with necessity, for quantitative estimation, of considering other factors as well; 7. possible changes of system time properties, in the context of closed loop control, whereby large differences may occur with respect to those in natural (open loop) status, depending on action and perturbation intensities, as well as on possible overall non-stationarities and non-linearities; 8. driving causes for changes, and classical analogies in the context of human psyche and cognition dynamics.

The paper illustrates discussions with concrete examples relating to Robocup-at-Home competition tests and applications.

References

1. Robocup Federation (9.2.2009), http://www.robocup.org/
2. van der Zant, T., Wisspeintner, T.: RoboCup X: A proposal for a new league where roboCup goes real world. In: Bredenfeld, A., Jacoff, A., Noda, I., Takahashi, Y., et al. (eds.) RoboCup 2005. LNCS (LNAI), vol. 4020, pp. 166–172. Springer, Heidelberg (2006)
3. Dessimoz, J.-D., Gauthey, P.-F., Kjeldsen, C.: Ontology for Cognitics, Closed-Loop Agility Constraint, and Case Study in Embedded Autonomous Systems – a Mobile Robot with Industrial-Grade Components. In: Proc. Conf. INDIN 2006 on Industrial Informatics, August 14-17, p. 6. IEEE, Los Alamitos (2006)

4. Dessimoz, J.-D.: La Cognitique - Définitions et métrique pour les sciences cognitives et la cognition automatisée, Editions Roboptics, Yverdon-les-Bains, Switzerland (August 2008) ISBN 978-2-9700629-0-5
5. Dessimoz, J.-D.: Contributions to Standards and Common Platforms in Robotics; Prerequisites for Quantitative Cognitics. In: International Conference on Simulation, Modeling, and Programming for Autonomous Robots (SIMPAR) 2008. First International Workshop on Standards and Common Platform for Robotics, Venice, Italy, October 3-7 (2008)
6. Dietrich, E., Markman, A.B. (eds.): Cognitive Dynamics: Conceptual and Representational Change in Humans and Machines. Lawrence Erlbaum Associates, Mahwah (2000)
7. De Gyurky, S.: The Cognitive Dynamics of Computer Science, Cost-Effective Large Scale Software Development. Wiley, Chichester (2006)
8. Wikipedia Foundation (5.2. 2009), http://en.wikipedia.org/wiki/Main_Page
9. HEIG-VD website (6.5.2009), http://rahe.populus.ch/rub/4.
10. Nardi, D., et al.: Rulebook, Draft edition (February 2009), http://www.ai.rug.nl/robocupathome/documents/rulebook2009_DRAFT.pdf

Adaptive Cellular Automata Traffic System Model Based on Hybrid System Theory

Yuge Xu, Fei Luo, and Xiaodan Tian

Automation Science and Engineering College, South China University of Technology, Guangdong Guangzhou 510640, China
xuyuge@scut.edu.cn, aufeiluo@scut.edu.cn, linxiaodan@sina.com

Abstract. Traffic system model with multiple lanes switch behaviors is very complex. In order to study the nonlinear and strongly relevant micro traffic system, this paper applied hybrid system theory into micro traffic system model. One novel cellular automata NaSch traffic model based on hybrid system is built up. The traffic flow made up of lots of cars is a hybrid system. The driving behaviors of cars are continuous but the events that affect the cars speed are certainly or randomly discrete events. This paper analyzes the effect of driver switch behaviors to the whole traffic flow. The proposed novel traffic model is simple and easy to expand. This model can simulate the traffic states in the different traffic conditions correctly especially in some complex multiple lanes conditions. Simulation results tell us the presented adaptive traffic system model is correct and effective.

Keywords: Micro traffic system, hybrid system, cellular automata, NaSch traffic model.

1 Introduction

Traffic problems have been studied for a long time to relieve a traffic jam and to increase transportation efficiency. By describing the change of traffic flow both on time and space, traffic simulation system estimates multifarious parameters, which is the technical support of road design and lay-out. The computer simulation system has been applied a lot to relieve the traffic problems and increase transportation efficiency by describe the vehicle behavior.

There're two main modes for traffic simulation, macro traffic model and micro traffic model, according to different research objects. Macro traffic model is used to describe the general characteristic of the system, and the micro model describes objects in detail which is easy to reappearance accurately the traffic condition by computer and become the best way to estimate the real-time traffic system.

In recent years, Cellular Automata (CA) method has been applied to a simulation for traffic flow control. CA is an array of sites (cells) where each site is in any one of the permissible states. The evolution of a site value depends on some rules (the combinational logic) which are functions of the present states of cells. Based on those rules and primal states, CA could construct many complicated systems by a set of

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 994–1003, 2009.

fundamental transformations. This method is so simple and flexible that is considered to be suitable for a simulation of intricate nonlinear system like a traffic flow. So there are a lot of traffic model based on CA model, such as the well-known NaSch model, FI model, and other improved model.

Based on the NaSch model, the hybrid system theory is used in the analyzing and modeling of traffic system. Systems that consist of a combination of continuous dynamics and discrete events are called Hybrid Systems [1-4]. The traffic flow is a typical hybrid system, the moving speed and displacement of vehicles are the continuous states; the stimulations influencing the vehicle behaviors are the certain or random discrete events. During those discrete events, we pay most attention to the effect between the driving way switch behavior and the whole traffic flow. We built up a micro traffic flow model based on Hybrid System-Cellular Automata theory by computer, and simulated the traffic under different condition by changing the rules and parameters of the proposed traffic system model.

2 NaSch Model

The use of the CA approach to model traffic flow along a road was made by Nagel and Schreckenberg [5]. In NaSch model, space, time and velocity are discrete. The space is divided into cell that has a length of 7.5m. Each site may either be occupied by one vehicle, or it may be empty. The integer velocity ranges from 0 to $Vmax$=5 cell/t; Time is taken as the driver reaction time (one second). For the arbitrary configuration, one update of the system consists of the following four consecutive steps, which are performed in parallel for all vehicles.

1) Acceleration: if (v<gap) then v=min{$Vmax$, v+1}
2) Slowing down (due to other cars): if (v>gap) then v=gap
3) Randomization: if v>0, then v=v+1 with probability P.
4) Car motion: x=x+v

From the fact above, NaSch model uses a minimal set of rules to simulated freeway.

There're two kinds of boundary conditions of NaSch model: one is periodicity boundary, the other is open boundary. The periodicity boundary condition creates a kind of idealized CA model. Though there is not any analytic solution because of the concurrent update method, the simulation can approach the approximate analytic solution under the periodicity boundary and provides a strong foundation for analyzing and validating problems. The open boundary means that the vehicles in or out of the road under certain disorder. It's more closed to the veracious states than the periodicity boundary. In recent years, NaSch model and its expanding model, such as metastability model, hysteresis model and slow-to start model, all based on the periodicity boundary, that means the speed of vehicles are 0 at the beginning.

The randomization shows the over braking, and describes well how the traffic switch between blocked and smooth by integrated with the concurrent update method. In the other hand, there is a big defect of NaSch mode, which is that the short distance between cars seldom happens, and it's against to the actual situation [6]. Furthermore, NaSch model only investigate motorcade on one drive way. In that case, the cars just drive after the anterior ones, that means it only study the characters of the motorcade

under constraint and unable to switch the driving way. But actually, to get higher speed and more reasonable traffic density, the switch and overtake behavior happens frequently. It is sure that the switch would cause the speed change of motorcade on both driving ways.

The micro traffic model set up in this paper, focuses on the influence of drive way switch to the speed, density of traffic flow. And the switch not only depends on the traffic status, but also the judgment of driver, which makes it a random, casual event. So we consider the switch behavior as a random event with a certain probability *P*.

3 Adaptive Cellular Automata Traffic System Model

With the Hybrid System-Cellular Automata theory, the adaptive micro traffic flow is constructed as follow:

The drive way is divided into cells that has a length of 7.5m and formed a planar CA. The vehicles come from the input side, the states of each cell include the car existed statement $car = \{0,1\}$, the velocity $v \in [0, v_{\max}]$; position $l \in [0, 7.5]$.

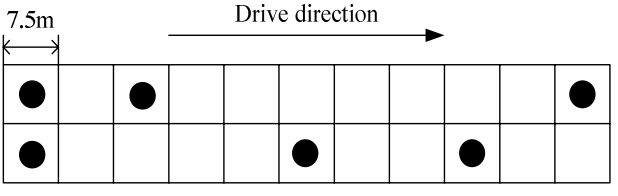


Fig. 1. 2 lanes way model

The rules are defined as follow:

Velocity adjustment: After satisfying the safety request, the velocity should be as large as possible. When the safe velocity $V > v_{\max}$, it means that the road is empty, the velocity should rise to the maximal speed $v_{\max}$ in the acceleration a; when current speed $v > V$, it means the car is too close to the front car, and have to slow down to the safe velocity V immediately.

Drive way switch: when $v < v_{\max}$ and the next drive way isn't taken and offer a possible higher speed, the switch condition is ready and would happen in the probability *P*.

States refresh: when the displacement is over the length of cell, the car goes to the neighbor cell, and the states of cells would be refreshed at the same time.

The acceleration and deceleration of vehicle are decided by the drivers and they changes in a very large range. Acceleration is a relatively slow process, so in this model, we take a fixed rate of change in the rate a. The deceleration of vehicles is directly related to traffic safety. The deceleration range is very big. We assume the vehicle can decelerate to the expected speed immediately under the braking effect.

Assume in our model, the speed limit value is the maximum speed $v_{\max}$ and every car can reach this maximum speed.

In NaSch Model, in order to avoid two cars are in one cell unit at the same time, which means two cars crash, in one update period the car behind should keep its current speed and its driving distance should be less than the distance with the car front. We assume the update period is 1 second and then the safety speed is the distance between two cars. In practice, the calculation of safety speed and safety distance should consider the current speed and the relative speed between two cars. In our model the safety speed is as follows:

$$v_s = v_{i+1} - v_i + \frac{d}{\Delta t} \tag{1}$$

There v_i is the car speed, v_{i+1} is the front car speed, d is the distance between two cars, Δt is the minimum variable time. We can choose higher safety speed when the current speed is high and if the current speed is low, the speed of the car behind should choose lower safety speed.

In this model, we choose open-boundary conditions. It means the cars drive in this model in some probability distribution and the cars that drive away will no longer enter again. Because it's difficult to judge the front road condition, the car which will drives away should keep its current driving state and leave this road.

Assume the initial road state is shown in Fig.2. The cars from left to right are named as *a* to *f*, the top-left corner value means the initial speed. The maximum speed is 20 m/s and acceleration is 5 m/s^2. According to the activity rules, the changes in cars are as follows:

A: If the distance with the front car is very long, the car can accelerate to the maximum speed.

B: If the distance with the front car is very small and the neighbor lane is occupied, so the car would not change lane and should decelerate.

C: If there is no car in front, the car can accelerate the car to the maximum speed.

D: If the distance with the front car is very small and the neighbor lane is empty, the switch condition is ready, the car can choose to switch.

E: If the car is the top one in the line, the car can accelerate the car to the maximum speed.

F: If the car is in the exit, the car should keep current speed and leave.

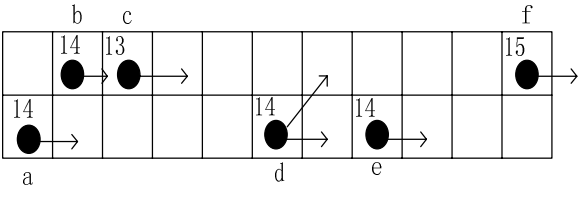


Fig. 2. Traffic states in time t

Based on the above rules, the cars in the model can drive as Fig.3 shown.

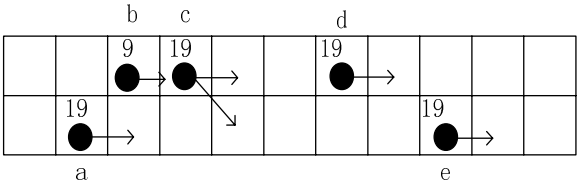


Fig. 3. Traffic states in time t+1

In the traffic model applications, there are two important problems. One is the relationship between density and volume, the other one is the distribution of cars distance and cars speeds. In NaSch model, cars distance update in fixed period T. Serial updated method is replaced of the original parallel update method. The car speed and distance between two cars are variable in time. The distribution becomes reasonable in switch behavior regulation condition and can reflect the real situation correctly.

In all, the proposed adaptive traffic model based on hybrid system is better than traditional NaSch model as mentioned follows:

1. After apply hybrid system theory into traffic modeling, the adaptive traffic model can describe the system dynamic performance completely and detail, the model can reflect the real situation correctly. NaSch model discrete the car speed and displacement in order to simply the model. In the adaptive traffic model, the car speeds and displacements are continuous variables. The car states in neighbor cell and the dynamic discrete events interact each other, it can describe the micro-traffic flow in precise.

2. The speed and displacement of car is continuous variables, the model performance is improved. In NaSch model, car speed is integer and every time displacements change one cell unit length which always set as 7.5m. These two variables always change too much or to small. In adaptive model, the car speeds and displacements change continuously and precise, they needn't to change in fixed rate.

3. The adaptive considers the road switch. The road switch often arises in real traffic system and it is one kind of discrete event to the traffic dynamic system. So considering the road switch, the traffic model is closer to the real traffic system.

4 Simulation Results Analysis

Based on the proposed cellular automata traffic model, we simulate the traffic model with software Matlab 7.0. The simulation steps are as follows:

Step1: Create input traffic flow and initialize the parameters. At first the cars distribute in the road with random probability *Px*. Assign one random speed value to the cars already on the road and the cars which is coming to the road in a allowable speed value range.

Step2: Begin the simulation. From the road exit, search every cell in order to find the position of the cars. The cars leave the road should keep its current speed. Calculate

the distance between the cars that don't leave the road and the front car. Calculate the distance between the car and the front car which is driving in the neighbor lane. If there is no car in the neighbor lane and the distance between the car to the front car, this car switch its road in probability P and calculate the safety speed vs. the safety speed *vs* is the Product of safety speed and update period time.

Step3: Adjust speed. If the current speed is bigger than the safety speed, decelerate to the safety speed. If the current speed is smaller than the safety speed and the maximum speed, accelerate with acceleration a till the speed reach the maximum value. And then keep the maximum speed.

Step4: Adjust position. Calculate the displacement of the car in cell n with the speed calculated with step3. If the displacement is beyond the cell length and the switch condition isn't ready, go into the next cell in current lane. If the displacement is beyond the cell length and the switch condition is ready, go into the next cell in the neighbor lane.

Step5: Record the simulation results and output the information. Repeat from step 2 to step4.

Assume the cars drive in standard road condition and traffic condition. The standard road condition means the road width should be more then 3.65m, the lateral width should be less then 1.75m. The standard traffic condition means the cars are with suitable size, every car keeps suitable distance with the other cars. Assume the random reach probability of cars is 0.5. The speed values of cars are random and obey normal distribution. In our model, the average speed value is 12 m/s. The other system parameter values can be found in table1.

Table 1. Simulation parameters

Lane	2
Length	1000m
Maximum speed	20 m/s
Acceleration	5m s^2
Switch probability P	0.3
Simulation time	3600s

Average speed 14.1281/s Average density 0.8697. The total car number in simulation process: 5118 cars

In Fig.5, the upper curve is the average speed on the road, it reflect that if there are traffic jams on the road. The bottom curve is the car density on the road, it reflect the distribution of the cars on the road. Through statistics the number of cars on the road within one hour, the transportation ability of this road can be evaluated. Based on the traffic parameters in table1, our model simulate on the condition of 1000m section and 2 lanes. The maximum speed is 70 Km/h or 20 m/s. Assume the drivers switch with probability 0.3 if the switch condition is ready.

Modify the traffic model parameters, we also simulate in the conditions of 3 lanes and 4 lanes. We choose three switch probability values, that is P=0, P=0.5 and

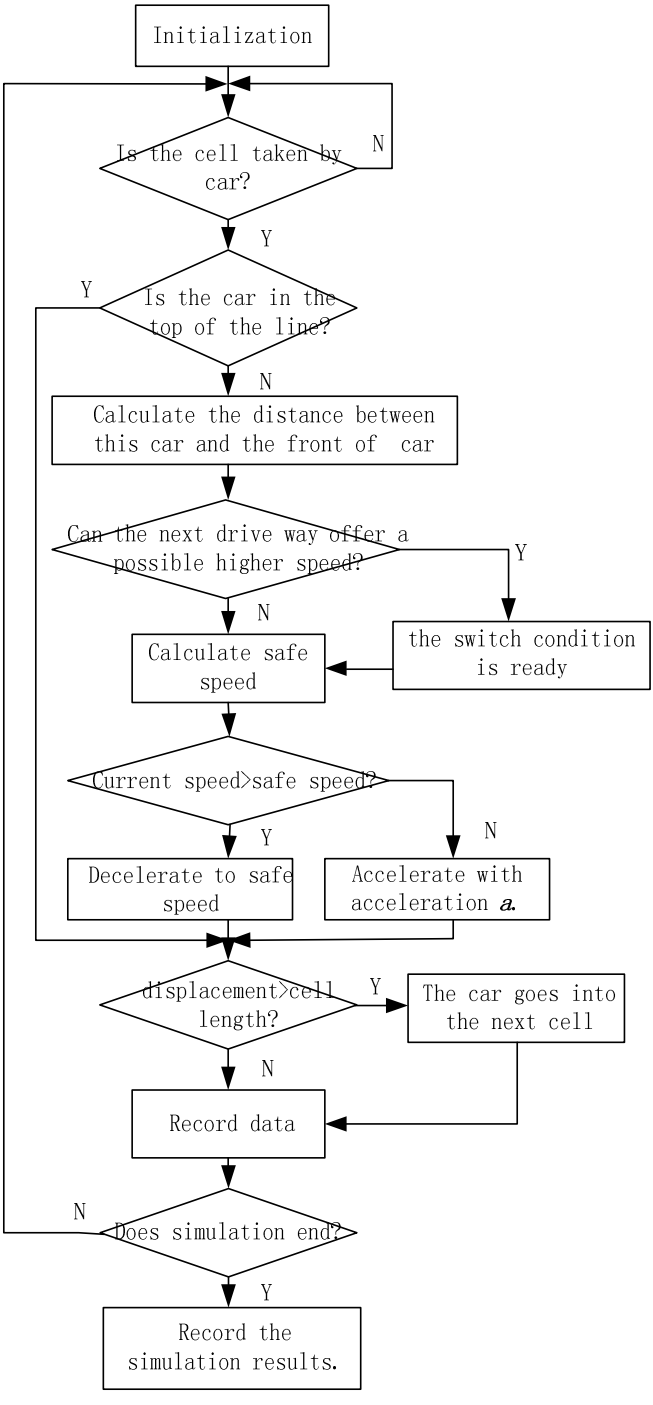


Fig. 4. Simulation flow chart

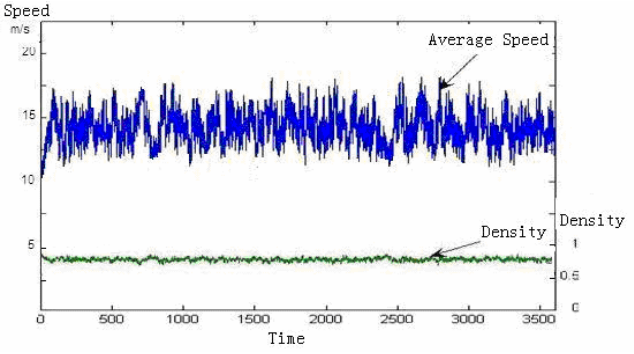


Fig. 5. Simulation results

Table 2. Switch probability *P*=0

	2 lanes	3 lanes	4 lanes
Average speed （m/s	15.0410	15.0452	15.0451
Density	0.9819	0.9900	0.9889
Transportation ability (cars/h)	5318	7780	10431

Table 3. Switch probability *P*=0.5

	2 lanes	3 lanes	4 lanes
Average speed （m/s	15.1983	15.1452	15.2328
Density	0.9546	0.9660	0.9486
Transportation ability (cars/h)	5417	7980	10743

Table 4. Switch probability *P*=1

	2 lanes	3 lanes	4 lanes
Average speed （m/s	15.2172	15.3554	15.2808
Density	0.9401	0.9456	0.9473
Transportation ability (cars/h)	5240	7841	10527

P=1.*P*=0 means there are no switch behaviors happened in the simulation. *P*=0.5 means half of the cars will switch its lane when the switch condition is ready. *P*=1 means all of the cars will switch its lane when the switch condition is ready. The simulation results are shown in table 2 table4.

From table 2 table 4, we can see that when there were no switch behaviors happened, the number of cars didn't effect the cars distribution and speed. The average speeds and densities in lane 2 lane 4 are the same. But when the switch probability increases, switch behaviors will increase, it will affect the dynamic traffic states. In the same condition, average speed increases when switch probability increases, car density decreases when switch probability increases. The traffic jam

Table 5. Comparison of transportation ability in different lanes

	2 lanes		3 lanes		4 lanes	
	TA	Growth (%)	TA	Growth (%)	TA	Growth (%)
P=0	5318	100	7780	146	10431	196%
P=0.5	5417	100	7980	147	10743	198%
P=1	5240	100	7841	149	10527	200%

will decrease when switch behaviors happen. With the same switch probability, when there are different cars going into this model, the transportation abilities are improved in different extent. Based on the transportation ability of 2 lanes, table 5 shows us the comparison of transportation ability in different lanes. TA means transportation ability.

Analyze the simulation results we find that the switch behaviors of drivers in order to reach the maximum speed can optimize the cars on the road in a certain extent. It can improve the driving speed and transportation ability. This phenomenon is more obvious in multiple lanes road such as three or four lanes.

5 Conclusion

With hybrid system theory, we proposed one adaptive cellular automata NaSch traffic model. This novel traffic model improves the performance of the traditional NaSch model. We regard the traffic flow made up of lots of cars as hybrid system. The car driving behavior is continuous but the events that affect the cars speed are certainly or randomly discrete events. In these discrete events, we focus on the effect of driver switch behaviors to the whole traffic flow. The proposed novel cellular automata traffic model is simple and easy to expand. By changing cell rules and parameters simply, this model can simulate the traffic states in the different traffic conditions in detail and correctly especially in some complex multiple lanes conditions.

Acknowledgments. This paper are supported by the Special Research Fund of Ministry of Education of China for College Doctoral Subjects (Projects for Young Scholars) under Grant No.20070561006, The Natural Science Foundation of Guangdong Province China under Grant No.9451064101002853 and the National Natural Science Foundation of China under Grant No.60774032.

References

1. Alur, R., Sontag, E.D., Henzinger, T.A. (eds.): HS 1995. LNCS, vol. 1066. Springer, Heidelberg (1996)
2. Henzinger, T.A., Sastry, S.S. (eds.): HSCC 1998. LNCS, vol. 1386. Springer, Heidelberg (1998)
3. Vaandrager, F.W., van Schuppen, J.H. (eds.): HSCC 1999. LNCS, vol. 1569. Springer, Heidelberg (1999)

4. Lynch, N.A., Krogh, B.H. (eds.): HSCC 2000. LNCS, vol. 1790. Springer, Heidelberg (2000)
5. Nagel, K., Schreckenberg, M.: A cellular automaton model for freeway traffic. J. Phys., 2221–2229 (1992)
6. Jin, Z.: Urban Mixed-Traffic Flow Modeling Techniques Based on Cellular Automaton. Zhejiang University, Huangzhou. PHD (2004)

Study on a Location Method for Bio-objects in Virtual Environment Based on Neural Network and Fuzzy Reasoning

Hongjun Wang, Xiangjun Zou, Changyu Liu, Tianhu Liu, and Jiaxin Chen

Key Laboratory of Key Technology, Agricultural Machine and Equipment (South China Agricultural University), Ministry of Education,
Guangzhou 510642, Guangdong, P.R. China
Tel.: +86-020-88371598; Fax: +86-020-85285036
xtwhj@tom.com, xjzou1@163.com, xjzou1@163.com,
yezhichunac@tom.com,
liuparalake@tom.com, jiaxin_c@163.com

Abstract. In picking manipulator location system, it is the actual challenge that the position of a bio-object is exactly determined in complex and uncertain environment. An accurate location method for a bio-object in virtual environment based on binocular stereo vision was studied. Firstly, an experiment platform for a bio-object and picking manipulator location based on binocular stereo vision was established, and the imaging of a bio-object and environment were gained by binocular stereo vision. Secondly, the causes of errors existing in the process of handling the bio-object and environment imaging were analyzed. An approach was point out that the vision errors were amended by means of building a neural network and a corresponding training system. Thirdly, considering there are many complex and uncertain factors existing in natural environment, fuzzy rule sets for ascertain the bio-object position were extracted and modeled. A fuzzy reasoning model which has multi-rules and multi-inputs is created based on experience of agricultural experts, which makes the location system more accurate and intelligent. Finally, a location simulation system was developed by VC++ and EON SDK, which can simulate a process of picking manipulator automatically locating to a bio-object.

Keywords: Virtual picking manipulator, Neural network, Fuzzy reasoning, Binocular stereo vision.

1 Introduction

Automatic picking operation is a changeable and complex process. One of key steps is to control picking manipulator, so that picking manipulator moves its terminal exciter to approach a bio-object, and then locates them as well as clamping them. Automatic picking operation is a changeable and complex process. One of key steps is to control picking manipulator, so that picking manipulator moves its terminal exciter to approach a bio-object, and then locates them as well as clamping them. Because some uncertainty problems always exist in natural environment and picking

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 1004–1012, 2009.

bio-objects randomly distribute in a fruit tree, which make a picking manipulator to locate at a random position for a variable bio-object each time in picking process and should control the clamping force appropriately so that it cannot damage fruits. Compared to a manipulator used in industrial area, an agricultural picking manipulator is characterized by difficult precision positioning, complex decision-making, and low success rate in picking.

A new technology for picking manipulator accurately locating at a picking bio-object by means of a binocular stereo vision is becoming popular. Naoshi Kondo in Japanese Okayama University established a vision system composed of color camera and image processing card, which can automatically select and identify mature fruits and implement precise positioning [1]. A cucumber picking robot developed in Japan measures target locations by using binocular stereo vision system, finds fruit stems by using fruit stems detector which is located on a terminal exciter, positions a cucumber by using mechanical positioning fingers [2]. Tiezhong Zhang and his colleagues regarded a position of strawberry center of gravity and picking point as key problems in robot picking operations. They used vision to study and do some correlated experiments [3]. Xinzhong Wang and his research team analyzed error reasons about detecting spatial position of tomato. The position error of a tomato was revised employing neural networks [4].

Automatically picking fruits operation is a complex process which includes mechanism, vision, biology and control technologies. The researches on agricultural picking manipulator mainly focus on single field simulation and modeling, fewer systems of virtual design and virtual simulation to simulate as well as evaluate whole comprehensive performance about picking manipulators. This paper explores a new location approach for bio-objects and picking manipulators in virtual environment. The vision errors caused by handling the imaging taken by a binocular stereo vision were amended by means of building a neural network and a corresponding training system. Fuzzy rule sets for determine a bio-object position in natural environment were extracted and modeled. A fuzzy reasoning model which has multi-rules and multi-inputs is created based on experience of agricultural experts, which makes the location system more accurate and intelligent.

2 The Experiment Platform for Bio-object Location

When using the combination methods of hardware and simulation to analyze location mechanism of a bio-object and a picking manipulator, the experiment platform should be constructed firstly, and then hardware test data about position information of a bio-object are collected and transmitted into virtual environment.

The experimental platform used in the paper is shown in Figure 1. It consists mainly of cameras, picking manipulator body, and background computer. Two cameras are used as binocular in binocular stereo vision system to acquire bio-objects and environmental image. The background compute is used to collect the uptake images by the two cameras, compute 3D coordinates of a bio-object in picking coordinates, and construct quickly the bio-object and picking scene in virtual environment. The accurate position of the bio-object can be realized by means of neural network and fuzzy reasoning algorithms.

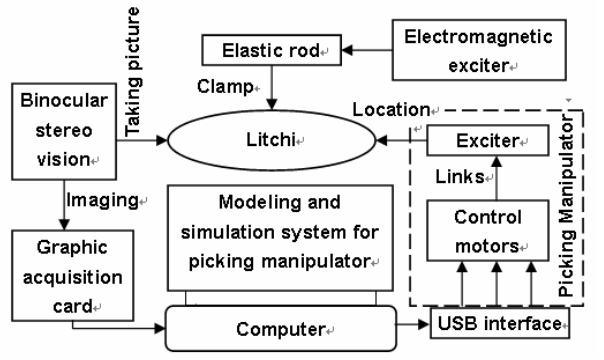


Fig. 1. The experiment platform

3 The Fusion Algorithm of Neural Network and Fuzzy Reasoning

The model structure of neural network and fuzzy reasoning is shown in Figure 2. It consists mainly of confidence estimation and weight fusion.

3.1 Location Errors Amended by Neural Network

There are three kinds of location errors in location system for a fruit bio-object based on binocular stereo vision system, which are position error of two cameras, demarcation error of two cameras, and mapping error coming from 2D imaging being converted to 3D data.

Position error of two cameras occurs when the two camera coordinates have some rotation angles, thus the position and shape of the same spatial point in world coordinate generated by the two cameras are different. It will bring position error compared to real position of a bio-object.

Camera demarcation is an inverse process of perspective imaging. By using modeling parameters of demarcation and 2D imaging coordinates of mapping process, the demarcation can get the 3D spatial point coordinates. It is a solution of uncertain problems, which has many demarcation errors. Because some variable factors exist in agricultural fruits environment (such as light condition, noise interference, occlusion and so on), images of the same fruit which were obtained by two different positional cameras are not the same. Imaging mapping error which was induced by extracting fruit edge feature is shown in Figure 3.

There are many uncertain error factors in the process of detecting the spatial of a fruit bio-object using binocular stereo vision system in natural environment. It is impossible that position errors can be amended by solving equations or once optimization searching. In fact, it don't need complication computer when a binocular stereo vision system simulating human eyes recognizes things. It can truly apperceive 3D information according to projection imaging in eyes by experiences. So, the three

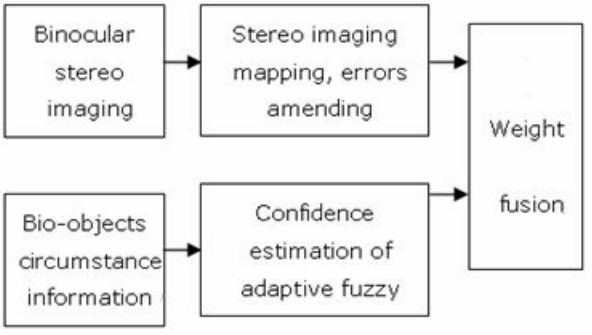


Fig. 2. The model structure of neural network and fuzzy reasoning

Fig. 3. The location errors caused by image mapping

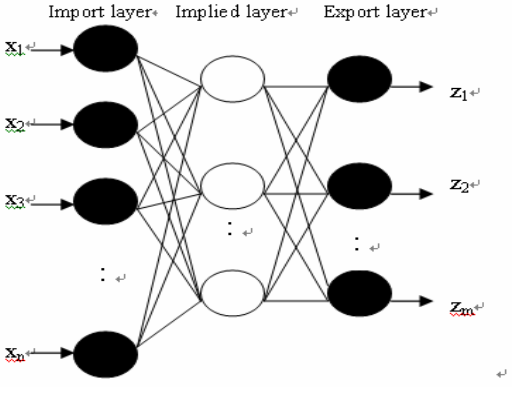


Fig. 4. The structure of BP neural network with three layers

kinds of location errors can be amended by establishing a three layer BP neural network, which is forward network with multi-layer and includes many uncertain factors. BP neural network is consisted of import layer, implied layer, and export layer, its structure was shown in Fig 4, whose import layer uses Sigmoid function, neural cells of export layer employ Purelin linear function.

3.2 Fuzzy Rules

Working Environment of picking manipulator is always uncertain as well as complex.

Some factors, such as weather condition, changing of wind direction and occlusion of leaves and its shadow, will affect the location accuracy, leading to the failure of picking operation. Therefore, in the process of picking operation simulation, the location of bio-object which was based simply on the mapping result of binocular stereo vision is inaccurate and the factors that effect picking pinpoint should be considered. Because of such uncertainty and fuzziness in the information of bio-object and agricultural environment, fuzzy rule sets of fruits positioning control should be constructed on the base of error correction with neural network. The using of fuzzy reasoning technology in picking location dramatically improves the location precision.

Before fuzzy reasoning, fuzzy language variables, which contain the picking decision related factors in the natural fruits growth environment, such as location coordinate error, weather condition, wind direction changing Leaves density and so on. Variables of fuzzy language with the fuzzed processing are show in table 1.

Table 1. Fuzzy lingual variables

Fuzzy variables	Fuzzy lingual		
Location errors	High	middle	low
Weather condition	good	middle	bad
Wind direction	quick	middle	slow
Leaf density	dense	general	sparseness

Table 2. Fuzzy rule sets

Fuzzy rules	Condition				Conclusion
	Location errors *A*	Weather condition *B*	Wind direction *C*	Leaf density *D*	Belief *R*
1	High	Good	Quick	Dense	r_1
2	High	Good	Quick	General	r_2
3	High	Good	Quick	Sparseness	r_3
…	…	…	…	…	…
63	Middle	Middle	Middle	General	r_{63}
64	Low	Bad	Slow	General	r_{64}

Integrated positioning of bio-object which was realized by using fuzzy neural network use fully the dizziness and uncertainty of agriculture biological information to realize location decision-making. The fuzzy rule sets which were based on the knowledge of agricultural expert are shown in Table 2.

In the topologic structure of fuzzy neural network, each layer has the same type node function. By using O_{ij} to express the output of node j in layer i, node function of node i in first layer can be expressed as:

$$O_{1,i} = \begin{cases} \mu_{Ai}(t) & i = 1,2,3 \\ \mu_{Bi}(u) & i = 4,5,6 \\ \mu_{Ci}(v) & i = 7,8,9 \\ \mu_{Di}(w) & i = 10,11,12 \end{cases} \tag{1}$$

In the formula, A_1, A_2 and A_3 are fuzzy sets of location error nodes i=1,2,3, corresponding separately to high, medium and low, B_1,B_2 and B_3 are fuzzy sets of weather nodes i=4,5,6, corresponding separately to good, medium and bad, C_1,C_2 and C_3 are fuzzy sets of wind direction changing nodes i=7,8,9, corresponding separately to violent, Medium and stationary, D_1,D_2 and D_3 are fuzzy sets of leaves density nodes i=10,11,12,corresponding separately to dense, general and sparse, and Q_{i1} is the fuzzy membership of a correspondent input.

The output of node i in second layer is product of all input and O_{2i} is the activity intensity of rule i.

The output of node i in third layer is the activity intensity of rule i divided by the sum of all rules activity intensity. It can be expressed as:

$$O_{3i} = \frac{O_{2i}}{\sum_{j=1}^{64} O_{2j}}, \quad j = 1,2,\cdots,64 \tag{2}$$

In this formula, Q_{3i} is the activity intensity of unified rules. It is equivalent to the using of w_i and $\overline{w_i}$ corresponding separately to the activity intensity of rule i and unified rules. Modified formula can be expressed as:

$$\overline{w_i} = O_{3i} = \frac{w_i}{\sum_{j=1}^{64} w_j}, \quad j = 1,2,\cdots,64 \tag{3}$$

The output of node i in fourth layer can be express as:

$$O_{4i} = \overline{w_i} \cdot r_i, \quad i = 1,2,\cdots,64 \tag{4}$$

in which r_i is conclusive parameter set.

There is only one node in fifth layer, whose output is a sum of all input. It can be expressed as:

$$O = O_{51} = \sum_{i=1}^{64} O_{4i} = \sum_{i=1}^{64} \overline{w_i} \cdot r_i \tag{5}$$

According to result by the training and study of neural network, a fuzzy reasoning model with multi-rule, multi-input and single output was obtained, which included many factors ,such as location coordinate, weather condition, wind direction changing and leaves density. The concrete structure can be described as follow:

$$\begin{aligned}
&\text{Rule 1 if } A_{11} \text{ and } A_{12} \ldots \text{and } A_{1n} \text{ then } B_1;\\
&\text{Rule 2 if } A_{21} \text{ and } A_{22} \ldots \text{and } A_{2n} \text{ then } B_2;\\
&\ldots\ldots\\
&\text{Rule m if } A_{m1} \text{ and } A_{m2} \ldots \text{and } A_{mn} \text{ then } B_m\\
&\text{Input } A_1^{'} \text{ and } A_2^{'} \text{ and } \cdots\cdots A_m^{'}\\
&\text{Result } B^{'}
\end{aligned} \tag{6}$$

Among these, A and A' are fuzzy sets of fruits creation environment on X, B and B' are fuzzy sets of

Location decision-making on Y, while $A_{i1}, A_{i2}, \cdots A_{in} \rightarrow B_i$ is fuzzy inference rule i, generating a multiple fuzzy relation $R_i(x_1, x_2, \cdots x_n, y)$.

The total fuzzy relation is $R = \bigcup R_i$

3.3 Fuzzy Reasoning

Fuzzy sets B' in the conclusion can be determined by inputting A':

$$B' = A'R = A'\bigcup R_i = A'\bigcup R_i = \bigcup A'R_i \tag{7}$$

Its fuzzy membership function is:

$$B'(y) = \vee\left[A'(x) \wedge (\vee A_i(x) \wedge B_i(y))\right] \tag{8}$$

In the agricultural knowledge reasoning process, the conception of similarity is introduced. If the similarity of A and A' in the rules are:

$$q(A', A) = \vee(A'(x) \wedge A(x)) \tag{9}$$

then

$$B(y) = \vee(q(A', A_i) \wedge B_i(y)) \tag{10}$$

By computing the similarity q (A', A) of A and A', the multi-rule fuzzy reasoning model compares this similarity with rule consequent, and integrates these compared results to get the fuzzy reasoning result.

4 The Virtual Simulation System

The location simulation system of a bio-object based on binocular stereo vision takes EON Studio 5.0 as the simulation platform of VE modeling and visualization, and is developed by using VC++.

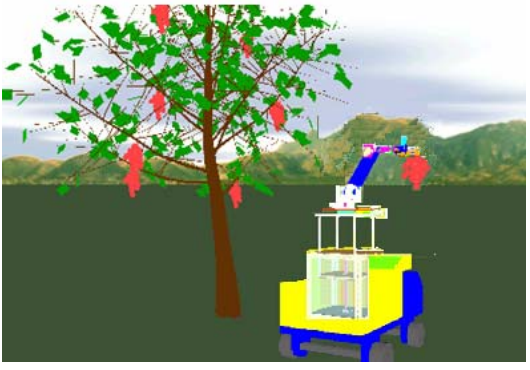


Fig. 5. Location simulation of picking manipulator in VE

Real-time positioning simulation of manipulator picking process in VE consists mainly of picking bio-objects, surrounding environment, and dynamic as well as static geometry modeling of a manipulator. Picking bio-objects and surrounding environment are constructed by use of EON Studio nodes function, while geometric properties value of those nodes are generated by 3D coordinate information which are acquired by binocular stereo vision. Determined space coordinate of picking bio-objects and obstacles by specific data calibration can provides analyzing data for motion planning. Attribute characteristics of nodes originate from segmentation processing on the images which are acquired by binocular stereo vision, and also from extracting information, such as color, gray and brightness.

Static and dynamic modeling is involved in a picking manipulator. In order to raise the system display speed and consider simultaneously that the simulation system is used to verify whether picking manipulator data is fused and whether motion planning and control algorithms are reasonable and correct, static modeling is replaced by simplified geometric model whose geometric dimension and scale are the same as real machines, and dynamic modeling is achieved by refreshing continuously apace coordinate value of static modeling with a group of intermediate data which is analyzed and calculated (by such dynamic modeling) between target location and current location by using inverse kinematics knowledge. Human-computer interaction (HCI) is achieved by using sensors, event-driven and routing (Route) mechanisms in VE, while behavior interaction is achieved by using transmission of information mechanism. Simulation result is shown in Figure 5.

Acknowledgments. This work was supported by the National Natural Science Foundation of China with Research Grant, 50775079, Natural Science Foundation of Guangdong, China, 9151064201000030, 925106420100009 07006692, 05006661, and Ph.D. Program Foundation of Ministry of Education of China, 200805640009.

References

1. Song, J., Zhang, T.Z.: Research Actuality and Prospect of Picking Robot for Fruits and Vegetables. Transactions of the Chinese Society for Agricultural Machinery, Tech. J. 37(5), 158–161 (2006)
2. Kawollek, M., Rath, T.: Machine Vision for Three-dimensional Modelling of Gerbera Jamesonii for Automated Robotic Harvesting. In: Int. Conf. Sustainable Greenhouse Systems – Greensys, pp. 15–20 (2004)
3. Zhang, T.Z., Chen, L.B., Song, J.: Study on strawberry harvesting robot: II. Images based identifications of strawberry barycenter and plucking position. Journal of China Agricultural University, Tech. J. (10), 48–51 (2005)
4. Zhong, Q.M., Han, P., Wang, X.Z.: Study on error amendment method for fruits position in natural environment. Machinery, Tech. J. (6), 25–26 (2006)

Planner9, a HTN Planner Distributed on Groups of Miniature Mobile Robots

Stéphane Magnenat, Martin Voelkle, and Francesco Mondada

LSRO laboratory, EPFL, 1015 Lausanne, Switzerland
stephane@magnenat.net
http://mobots.epfl.ch

Abstract. Autonomous mobile robots are promising tools for operations in environments that are difficult to access for humans. When these environments are dynamic and non-deterministic, like in collapsed buildings, the robots must coordinate their actions and the use of resources using planning. This paper presents Planner9, a hierarchical task network (HTN) planner that runs on groups of miniature mobile robots. These robots have limited computational power and memory, but are well connected through Wi-Fi. Planner9 takes advantage of this connectivity to distribute the planning over different robots. We have adapted the HTN algorithm to perform parallel search using A* and to limit the number of search nodes through lifting. We show that Planner9 scales well with the number of robots, even on non-linear tasks that involve recursions in their decompositions. We show that contrary to JSHOP2, Planner9 finds optimal plans.

1 Introduction

Autonomous mobile robots are promising tools for operations in environments that are difficult to access for humans. In particular, groups of miniature low cost robots are well suited for missions in hazardous environments; for instance to do search and rescue in partially destroyed structures. Indeed, small robots can pass through small holes to access most locations. Several robots can collaborate to free a path, for instance by moving objects or by extinguishing fires. Multiple physical robots provide redundancy against contingencies resulting from the instability of the environment. As the robots are inexpensive, the loss of some of them is acceptable. In this context of rescue operations, most of previous works focused on reactive controllers, see for instance [1,2]. However, real-world robotics tasks that involve manipulation of the environment cannot be performed by reactive controllers only, because the actions of the robots might modify the state of the world irreversibly into situations that prevent the completion of the mission. For instance in a rescue scenario, robots must never irreversibly fill the only passageway to humans trapped in a room, regardless of how this action might help completing other tasks. To alter the world the right way, the robots must reason about their course of action; the common way to do so is to use an automated planner.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 1013–1022, 2009.

This paper presents Planner9, a distributed hierarchical task network (HTN) planner [3, ch. 11] that runs on groups of miniature mobile robots. These robots are good models of expendable field robots because they are affordable and all-terrain [4]. They are built around a smartphone-level computer which provides one tenth of the power for one twentieth of the energy consumption of a laptop computer. This computer runs Linux, and provides a Wi-Fi connection to its peers. Planner9 takes advantage of this large bandwidth on CPU ratio to distribute the planning over different robots.

In the results section, we show that Planner9 scales well with the number of robots. We study how the planning durations vary when we change the complexity of the environment. We compare Planner9 to JSHOP2, a free and an often referenced HTN planner.

2 Related Work

A vast literature exists on planning with multiple agents. However, the choice of a particular algorithm is a trade-off between several factors such as planning expressiveness, distributivity, bandwidth consumption, and speed of execution [5]. In this section we focus on the approaches that are implementable on real robots with limited resources.

To distribute planning among different agents, one can run an independent planner on each agent. Works in simulated robotics soccer have used HTN planning that way [6]. However, in general this solution cannot improve the time nor reduce the memory required to solve a specific problem. In the direction of distributing the planning process, [7] has integrated the SHOP HTN planner within a distributed agent framework. The resulting A-SHOP planner uses the provided infrastructure to query the state of the world, evaluate preconditions, apply effects, and estimate potential states from remote agents. However, in this work the planning itself is still centralized. Theoretical works in multi-agent systems have shown that it is possible to integrate the distributed aspect in the planning algorithm [8,9]. For instance [10] has proposed a market-based approach where *and/or* trees of tasks are exchanged between agents. These authors acknowledge the interest of more expressive planning but address the issues of distributivity and scalability first ("Further research will also investigate further generalization of the tree structures and task constraints." [11, p. 25]). Recent works in HTN have proposed stratified planning where remote agents plan subtasks and report them to a master. In [12], finding the final plan is the result of the exchange of proposals and counter proposals between agents. In [13], the child agents are also responsible for execution, and interleave planning, execution, and replanning. These methods require a large number of synchronization messages. On the contrary, Planner9 considers the different robots as a computer cluster and distributes the planning of any task to any robot and thus takes advantage of all the available computational power using simple synchronization.

3 Materials and Methods

To measure the performances of Planner9, we have developed a small search and rescue scenario (Fig. 1). In this scenario, groups of robots are dropped into a damaged building and must bring a medical kit to humans trapped in a room. Multiple fires block the ways; robots can use fire extinguishers to put them out, but an extinguisher can put out only one fire. It is a typical situation where reactive controllers would fail because there are not enough extinguishers directly available to put out all fires. The robots must use the right extinguishers on the right fire in the right order and thus need a plan (Fig. 2). We believe that this scenario is representative of the complex tasks that would require a planner. Here we suppose that the robots know the locations of fires, extinguishers, objects, and people. To discover them in a real-world situation, we can imagine to deploy exploration robots that would map the environment.

We conceived Planner9 to run on our latest generation of miniature mobile robots, which comes down from the S-bot robot [4]. Only one prototype is currently available, yet dozens of robots are in production. We have compiled, run, and benchmarked Planner9 on this prototype. To simulate more robots, we then run multiple slave processes of the planning algorithm on a multi-core computer while another computer runs a master control program. We use cpulimit[1] to only allow as much processing power as available on the robot, which corresponds to 5 % of an Intel Core2 at 2.83 GHz. Thanks to ulimit[2], we limit the memory to 100 MB, which is the amount available on the robot. Finally, we divide the available Wi-Fi bandwidth (19 Mbps using IEEE 802.11g) by the number of slaves and limit the network bandwidth using trickle[3]. This way, our measurements are representative of the expected performances on the robots. This approach works as long as all slaves plan for a long duration. When we increase the number of slaves, the time to find a solution decreases which is a problem. Indeed, cpulimit and trickle impose limitations by pausing and restarting execution and data transmission. They interfere with our measurements when slaves find solutions in the same order of magnitude of time as these tools operate. We have observed such interferences starting from 8 slaves. Therefore, we limit our scalability measurements to 7 slaves.

We perform two experiments. First we measure the scalability of Planner9 by varying the number of planning slaves (simulated number of robots), as explained in the previous section. We perform 100 runs for each point, and show the average and the standard deviation. Second, we vary the difficulty of the planning by adding elements to the world that are useless to the robots, as show in Fig. 3. We first add two empty rooms, then we add extinguishers in these rooms and a medical kit in a3. These elements add more branches to the search tree without affecting the solution. For each case, we perform 100 runs using 1 and 7 slaves. We also solve the same problem using JSHOP2[4] [14], a free and an often referenced

[1] http://cpulimit.sourceforge.net/
[2] http://www.linuxhowtos.org/TipsandTricks/ulimit.htm
[3] http://monkey.org/~marius/pages/?page=trickle
[4] http://sourceforge.net/projects/shop

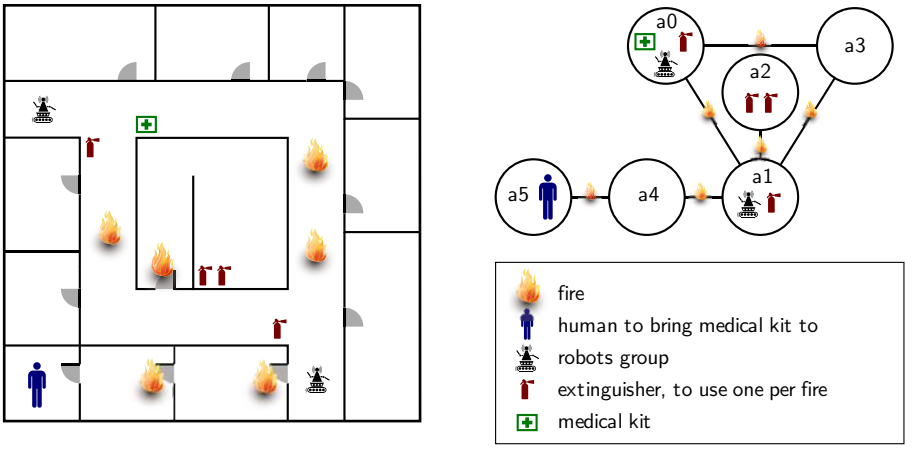


Fig. 1. Search and rescue scenario (left) with abstract representation (right). The robots are dropped into a damaged building. They must bring a medical kit to the humans trapped in the lower left room. To do so, they must extinguish the fires in the right order; otherwise they would fail the rescue operation as there are not enough extinguishers readily available to put out all fires.

```
take(r0, e0)                          a0, a1, a2, a3, a4, a5 : rooms
extinguish(a1, a0, r0, e0)
take(r1, e1)                          r0, r1 : groups of robots initially
extinguish(a2, a1, r1, e1)            in a0 and a1 respectively
move(a2, r0)
take(r0, e2a)                         e0, e1, e2a, e2b : extinguishers
move(a1, r0)                          initially in a0, a1, a2, and a2
extinguish(a4, a1, r0, e2a)           respectively
move(a2, r0)
take(r0, e2b)                         o0 : medical kit, initially in a0
move(a4, r0)
extinguish(a5, a4, r0, e2b)
move(a0, r1)
take(r1, o0)
move(a5, r1)
drop(r1, o0)
```

Fig. 2. Solution plan for the scenario presented in Fig. 1. `take` lets a group of robots pick up an object. `extinguish` puts out a fire between two rooms. `move` displaces a robot group to a room. `drop` puts down the object the robots hold.

HTN planner implemented in Java. As JSHOP2 is a depth-first search planner, we add bookkeeping actions that prevent infinite recursions. We do not count these actions when comparing plans sizes. We substract the startup time of JSHOP2 (time to decompose an empty task) to put it on an equal basis with Planner9.

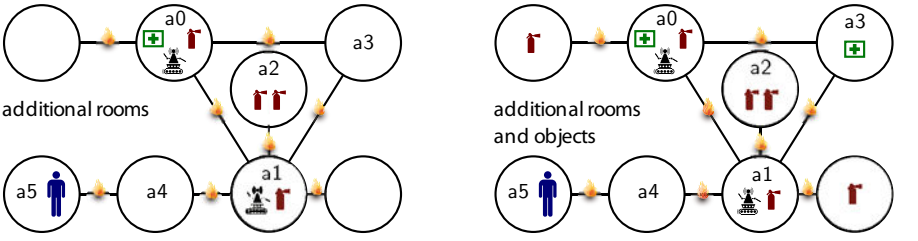


Fig. 3. Scenarios with additional elements useless to the robots. These elements add more branches to the search tree without affecting the solution.

4 Planner9

Planner9 is a HTN planner. A HTN planner decomposes a goal task into subtasks until it finds a sequence of actions that the robots can perform. The planner knows the available methods and their possible decompositions, like in Fig. 4. When given the goal task and the initial state of the world, the HTN planner seeks an admissible sequence of actions. The actions can affect the state of the world; so the planner records these alterations.

Planner9 plans partially-ordered graphs of tasks using forward decomposition, as in [3, p. 243]. It keeps track of each possible decomposition in a different search node. Planner9 starts planning with a single node containing the goal task and the initial state of the world. When visiting a node, Planner9 iterates through all tasks that have no predecessor. If the task is an action, it applies this action to the current state of the world and stores the action as part of the plan. Otherwise, Planner9 instantiates the different possible decompositions of the task. This process goes on until there is no more node left or until Planner9 has found a node with no more task to decompose.

The state of the world consists of a set of n-ary relations over a set of values. The application of an action affects these relations. The values represent things from the real world, like rooms or robots in Fig. 4. In the latter, the `move` action will update the `isIn` relation between the robots and the rooms. The planner

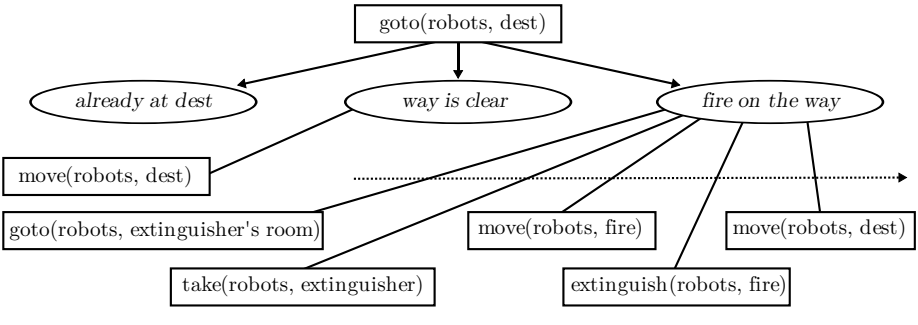


Fig. 4. Methods for going to a room by recursively putting out fires on the way

creates variables when decomposing tasks. For example in Fig. 4, the `goto` task can be decomposed using the *fire on the way* alternative. This decomposition introduces new variables: the extinguisher to use and the room where the extinguisher is located. The decomposition has preconditions over these variables that the state of the world must satisfy. For instance, the extinguisher must be located in an accessible room. When Planner9 decomposes a task, it performs *lifting*: it accumulates its preconditions for delayed check. Planner9 assigns a value to a variable only when an action changes a relation this variable appears in. For every variable assignations allowed by the accumulated preconditions, Planner9 updates the state and creates a new search node. To do so in an efficient way, it assigns values using DPLL [15]. The use of lifting is one of the improvements of Planner9 over the basic HTN algorithm. It results in fewer search nodes and more processing per node, which is desirable for parallelization.

Planner9 chooses the node to visit by selecting the least expensive one using A* [16]. As cost it adds the total cost of the decomposition so far (path-cost in A*) and the number of remaining tasks to be decomposed (heuristic-cost in A*). One advantage of A* with respect to a depth-first search is to allow free recursions in the definition of the planning domain. This is useful in robotics because real-world problems are often expressed in a recursive way, like in Fig. 4. Moreover, as each node is independent, Planner9 can distribute the A* search over the network using a master/slave architecture.

The slaves run the algorithm described in the previous paragraphs and report periodically the cost of their cheapest nodes to the master. When the cost differences between slaves are significant, the master requests nodes from the slave with the lowest cost. It then transfers them to the slave with the highest cost. As our results show, this load-balancing algorithm is simple yet efficient. Slave robots can appear on and disappear from the network dynamically, for instance when they boot or because their batteries are discharged. The master discovers available slaves dynamically using the Zeroconf protocol[5], which is based on broadcast announcing. Thanks to it, Planner9 does not require any central registry and can use any robot available on the local network.

Planner9 is implemented in C++, is open source[6], and runs on embedded Linux. It only depends on the C++ standard library, the boost library[7], and Qt Embedded[8], which are all open source. Planner9 is thus easy to embed in any robot running a modern Linux board.

5 Results

Fig. 5 shows the measures of the scalability of Planner9. The first plot shows that the performances scale nicely with the number of slaves. The speedup, which

[5] `http://www.zeroconf.org/`

[6] One can access the source tree using git; to retrieve it, type: `git clone git://gitorious.org/planner9/planner9.git`. We performed the experiments using revision `a501cfd704e4c759a0d83ea27dfcc85be53929ef`.

[7] `http://www.boost.org`

[8] `http://www.qtsoftware.com/products/`

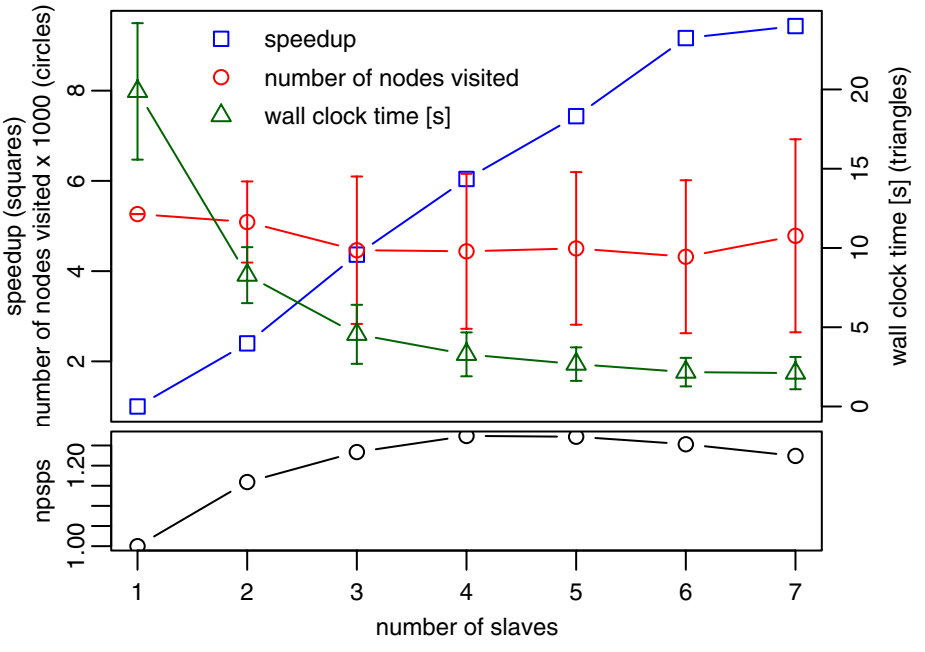


Fig. 5. The scalability of Planner9 with respect to the number of slaves. The scenario is shown in Fig. 1. The error bars show the standard deviation over 100 runs. In the second plot, *npsps* represents the number of nodes visited per second per slave, renormalized.

is the planning time using n slaves divided by the time using 1 slave, is even super linear. We analyse this using the second plot, which shows the number of nodes visited per second per slave, renormalized. When we increase the number of slaves, each can processes more nodes per second. The cause of the super linearity is a combination of the structure of the problem [17] and a memory cache effect. Indeed, by distributing the planning, each slave holds fewer search nodes in memory. Saving memory reduces the strain on the CPU cache and on the memory allocator, which results in a better performance. When the number of slaves grows further, the performance deteriorates. We attribute this effect to the load balancing, which increases the average time to find a solution because it moves low-cost nodes around.

Fig. 6 shows the measures of the ability of Planner9 and JSHOP2 to cope with environments of increasing difficulty. On the basic environment, Planner9 with 1 slave is faster than JSHOP2; using 7 slaves, it is one order of magnitude faster. Moreover, Planner9 uses both groups of robots while JSHOP2 extinguishes all fires with `r0`, which adds one more move action to the plan. Adding two empty rooms does not disturb Planner9 much, but JSHOP2 cannot cope: it exhausts its memory before finding any solution, even if we give it 2 GB instead of the 100 MB available on the robot. When we add useless objects in the empty rooms, the planning time of Planner9 increases but does not explode. In this environment,

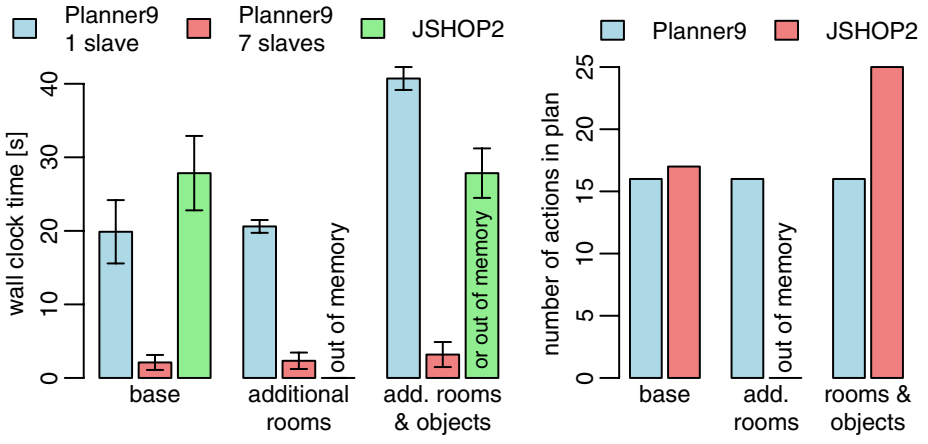


Fig. 6. The scalability of Planner9 with respect to the complexity of the environment. The scenario are shown in Fig. 3. The error bars show the standard deviation over 100 runs. JSHOP2 fails to find a plan when we add rooms, but sometimes succeeds when we further add objects, as we explain in the main text.

JSHOP2 either finds a solution quickly or exhausts its memory, depending on the order of appearance of the new elements in its initial state of the world. If we declare the new elements (rooms and objects) before the basic environment, JSHOP2 finds a solution. If we declare them after, it fails. Moreover, when it finds a solution it is a long plan with a lot of useless actions, such as using one extinguisher to access a room to get another one.

These results show that Planner9 scales well with the complexity of the environment, compared to JSHOP2. We attribute these excellent performances to the lifting which keeps different values assignments in a single node by abstracting them using a variable. Planner9 assigns a value later, when it has collected more constraints on this variable. On the contrary, JSHOP2 assigns values early and performs depth-first search. If by chance the first search branches lead to the solution, JSHOP2 finds it quickly. But otherwise, it must explore an exponentially large number of branches before finding a solution.

6 Future Work

The next step is of course to run Planner9 completely on our robots and confirm its scalability. Then, we will integrate Planner9 with the perception subsystem to create the initial state of the world and implement the execution of plans using the robots' actuators. We want to study the expectations and the limitations of using Planner9 to bring reasoning to collective robotics.

Planner9 is currently a basic HTN planner and does not provide any extension such as probabilistic planning or conditional actions. The rationale is that planning real-world scenarios is computationally intensive and we want it to run live

on the robots themselves. Moreover, we can make good use of the availability of multiple robots to reduce the uncertainty of the world perception through concurrent sensing from different locations. Merging these would produce a robust estimation of the state of the world for planning at symbolic level. Nevertheless, the parallelization capabilities of Planner9 would work with extensions as well so we can implement them should the need arise.

7 Conclusion

We have presented Planner9, a HTN planner that dynamically distributes its processing to multiple mobile robots by considering them as a computer cluster. To do so we have enhanced the HTN algorithm with an A* search and with the lifting of the tasks' preconditions. Previous works have explored how to distribute the decomposition of a particular subtask, but Planner9 distributes the processing of any task and thus benefits from all available computational power. Albeit simple, this mechanism is efficient as our results show. When compared to JSHOP2, a milestone in HTN research, Planner9 always finds optimals plans; while JSHOP2 only sometimes finds plans, and they are sub-optimal. Because Planner9 runs on inexpensive robots, we believe that it is a firm step towards bringing intelligence to autonomous mobile robots in non-industrial environments.

Acknowledgements

We thank Valentin Longchamp for compiling Planner9 on the prototype of the robot. We thank the Open Clip Art library[9] and Valessio Soares de Brito for the graphics elements used in Fig. 1 and Fig. 3. We thank Emmanuel Eckard and Basilio Noris for proofreading the manuscript.

References

1. Kumar, V., Rus, D., Singh, S.: Robot and sensor networks for first responders. IEEE Pervasive Computing 3(4), 24–33 (2004)
2. Stormont, D.: Autonomous rescue robot swarms for first responders. In: Proceedings of the 2005 IEEE International Conference on Computational Intelligence for Homeland Security and Personal Safety, pp. 151–157. IEEE Press, Los Alamitos (2005)
3. Ghallab, M., Nau, D., Traverso, P.: Automated Planning: theory and practice. Morgan Kaufmann, San Francisco (2004)
4. Mondada, F., Pettinaro, G.C., Guignard, A., Kwee, I., Floreano, D., Deneubourg, J.L., Nolfi, S., Gambardella, L., Dorigo, M.: SWARM-BOT: a New Distributed Robotic Concept. Autonomous Robots, special Issue on Swarm Robotics 17(2-3), 193–221 (2004)
5. Erol, K., Hendler, J.A., Nau, D.S.: HTN Planning: Complexity and Expressivity. In: AAAI, pp. 1123–1128. AAAI Press, Menlo Park (1994)

[9] `http://www.openclipart.org/`

6. Obst, O., Boedecker, J.: Flexible Coordination of Multiagent Team Behavior Using HTN Planning. In: Bredenfeld, A., Jacoff, A., Noda, I., Takahashi, Y. (eds.) RoboCup 2005. LNCS (LNAI), vol. 4020, pp. 521–528. Springer, Heidelberg (2006)
7. Dix, J., Munoz-Avila, H., Nau, D., Zhang, L.: IMPACTing SHOP: Putting an AI planner into a multi-agent environment. Annals of Mathematics and Artificial Intelligence 37(4), 381–407 (2003)
8. Durfee, E.: Distributed problem solving and planning. Multiagent systems: a modern approach to distributed artificial intelligence, 121–164 (1999)
9. DesJardins, M., Durfee, E., Ortiz, C., Wolverton, M.: A survey of research in distributed, continual planning. AI Magazine (1999)
10. Dias, M., Zlot, R., Kalra, N., Stentz, A.: Market-based multirobot coordination: A survey and analysis. Proceedings of the IEEE 94(7), 1257–1270 (2006)
11. Zlot, R., Stentz, A.: Market-based multirobot coordination for complex tasks. The International Journal of Robotics Research 25(1), 73–102 (2006)
12. Pellier, D., Fiorino, H.: A Unified Framework Based on HTN and POP Approaches for Multi-Agent Planning. In: IEEE/WIC/ACM International Conference on Intelligent Agent Technology, pp. 285–288. IEEE Press, Los Alamitos (2007)
13. Hayashi, H., Tokura, S., Ozaki, F.: Towards Real-World HTN Planning Agents. Knowledge Processing and Decision Making in Agent-based Systems 170, 13–41 (2009)
14. Nau, D., Au, T., Ilghami, O., Kuter, U., Murdock, W., Wu, D., Yaman, F.: SHOP2: An HTN planning system. Journal of Artificial Intelligence Research 20(1), 379–404 (2003)
15. Davis, M., Logemann, G., Loveland, D.: A machine program for theorem-proving. Commun. ACM 5(7), 394–397 (1962)
16. Hart, P., Nilsson, N., Raphael, B.: A formal basis for the heuristic determination of minimum cost paths. IEEE Transactions on Systems Science and Cybernetics 4, 100–107
17. Rao, V.N., Kumar, V.: Superlinear speedup in parallel state-space search. In: Proceedings of the Eighth Conference on Foundations of Software Technology and Theoretical Computer Science, London, UK, pp. 161–174. Springer, Heidelberg (1988)

An Adaptive Rolling Path Planning Method for Planet Rover in Uncertain Environment

Jinze Song, Bin Dai, Huihai Cui, Enzhong Shan, and Hangen He

College of Mechatronic Engineering and Automation,
National University of Defence Technology
410073 Changcha, Hunan, P.R. China
nwsac97@gmail.com

Abstract. Though path planning methods based on rolling windows have been successfully applied to planet rover in uncertain environment, the efficiency still remains unsatisfactory due to the single-step advancement strategy. In this paper, a novel path planning approach called Rapidly-exploring Random Tree (RRT) is introduced to efficiently find local feasible paths in the region of rolling windows. The planning step-size in each rolling window can be adaptive adjusted according to the local environment it's moving in, which is recorded by historical information perceived by the sensors. Combined with the goal-oriented heuristic strategy, the global collision-free solution path can be generated by successively connecting the local feasible paths. A number of infield experiments demonstrate the effectiveness of the proposed method.

Keywords: Adaptive path planning, RRT, uncertain environment.

1 Introduction

The path planning problem of an autonomous robot operating in a 2-dimensional surface with obstacles is a fundamentally issue to accomplish autonomous navigation. The objective of path planning is to find a collision free path from a given initial position to a predefined target point. According to the difference of environment information what robot known, there are two kinds of planning methods: global panning and local planning. Many authors have considered a model with complete information [1] [2] [3], where the robot has perfect knowledge about the obstacles.

In reality, path planning of a robot often cannot be based on complete priori knowledge of the environment. In this paper the laser scanner is used because laser scanner can provide not only distance of obstacles but also the angles between robot orientation and obstacles.

There are lots of research works on the local planning using various techniques, such as artificial potential field, neural networks, fuzzy logic etc. There is a method based on the rolling optimization principle adopted in predictive control [4] [5] [9]. This method makes more full use of the real-time local environment information by the on-line rolling optimization and feedback. The rolling path planning based on rolling windows is a powerful tool for the planning problems in uncertain environment. The

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 1023–1031, 2009.

rolling planning executes the local planning repeatedly by updating the newest local environment information. Though the rolling path planning method can solve the planning problem in an uncertain environment, the performance remains to be improved due to the single-step extension strategy.

Rapidly-exploring Random Tree (RRT) [6][7][8] is a very useful tool for designing single-queried path planners. The original RRT algorithm prefers to create a tree structure to incrementally explore unknown configuration space by Voronoi-biased extension. Urmson and Simmons [10] present a series of modified versions of the RRT algorithm that select tree nodes for expansion based on the cost of their current path from the initial node. Kuffner et al [11] present bidirectional RRT-Connect algorithm that grows two trees to explore and connect with each other rapidly during the process of tree expansions, resulting in substantial improvement of the performance.

RRT and its variants can also be considered as powerful tools for local path planning problems when intermediate configurations are inputted as planning queries. Therefore we propose a novel variant of RRT based on goal-oriented heuristic strategy to efficiently find a series of local feasible paths in the region of rolling window. The global collision-free solution path can be generated by successively connecting the local feasible paths. The method is implemented on the planet rover tested which designed by ourselves (as shown in Fig.1). A number of experiments on planet rover demonstrate the effectiveness of the method.

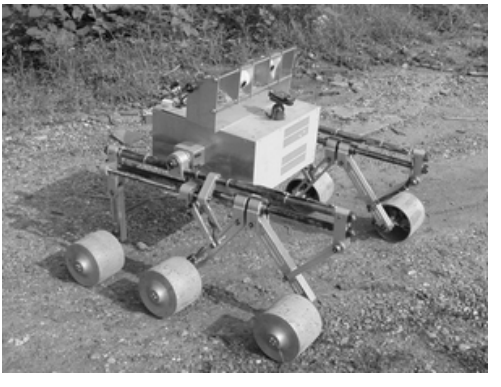


Fig. 1. Pototype of planet exploration rover

2 Path Planning Algorithm Based on Rolling Windows

2.1 Definition and Presentation of the Local Environment

The object of planning is finding a collision free path from start to end in the global environment. As shown in Fig.2, let W_s denote the unknown global environment, the origin of W_s coordinate frame locates at the start point and the x axis coincide with the orientation of the robot. And at arbitrary time, the robot can detect only the circular area W with radius r_d , ∂W denotes the boundary of detecting range. Let $\triangle t$

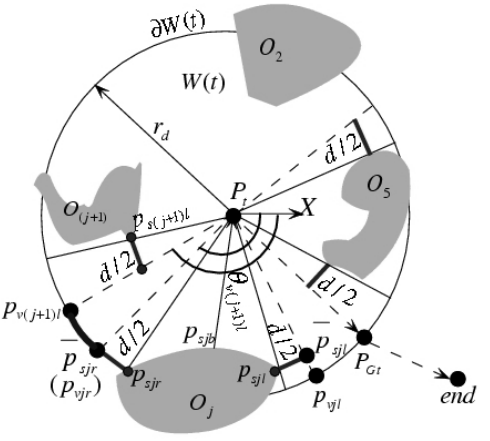


Fig. 2. The selection of local goal

denote the control period and ε denote the step length within the $\triangle t$, $0 < \varepsilon < r_d$. The coordinate in W_s of robot at instant t is $P_t(x_t, y_t)$, and the coordinate of local goal point is $P_{Gt}(x_g, y_g)$, and $\forall p \in W_s$ has determinate coordinate $p(x, y)$.

There are finite static obstacles $O_1, O_2, ..., O_n$ in W_s, and ∂O_i denotes the boundary of the *ith* obstacle. Let $d(p_i, p_j)$ denote the distance from $p_i(x_i, y_i)$ to $p_j(x_j, y_j)$.

$$d(p_i, p_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{1}$$

Definition 2.1: $\forall p \in W_s$, $p \notin O_i$,(i = 1,…,n), then p is called feasible point, and the set of all feasible point is called global feasible region, i.e. VD ; $\forall p \in O_i$,(i = 1,…,n), then p is called unfeasible point, and the set of all unfeasible point is called global unfeasible region, i.e. NVD then

$$VD = W_s \bigcap \overline{(U_{i=1}^{n} O_i)} \tag{2}$$

$$NVD = W_s \bigcap (U_{i=1}^{n} O_i) \tag{3}$$

Definition 2.2: $W(t)$ is called detecting region at P_t ,then, $W(t) = \{p \mid p \in W_s, d(p, P_t) \le r_d\}$, and $W(t)$ is also called rolling window at time t.

Definition 2.3: At time t, $\forall p \in W(t)$, $p \notin O_i$,(i = 1,…,m)m ≤ n, then p is called feasible point at time t, the set of all of p at time t is called feasible region at time t, i.e. $VD(t)$. Obviously $VD(t) \in VD$; $\forall p \in W(t)$, $p \in O_i$,(i = 1,…,m),then p is called unfeasible point at time t, the set of all of unfeasible point is called unfeasible region at time t, i.e. $NVD(t)$. Obviously $NVD(t) \in NVD$.

2.2 Obstacles Expanding and the Selection of Local Goals

The crucial task of path planning based on rolling windows is finding the mapping from the global goal end to the boundary of rolling windows $\partial W(t)$. We call it local goal P_{Gt} at time t, obviously $P_{Gt} \in \partial W(t)$ and $P_{Gt} \in VD(t)$ .Therefore, first of all we must expand the local obstacles detected.

As shown in Fig.2, at time t we got a series of unfeasible point when robot had scanned the jth obstacle, i.e. $p_{sj} = (p_{sj1}, p_{sj2}, ..., p_{sjq})$. Obviously, $p_{sj} \in \partial O_j$ and $p_{sj} \in \partial NVD(t)$. p_{sj1} was the first unfeasible point of O_j when robot was detecting the O_j in a counterclockwise manner, i.e. p_{sjl}, we called it left boundary point of O_j . Similarly, p_{sjq} i.e. p_{sjr} called the right boundary point of O_j . And let $p_{sjb} = \{p_s \mid \min d(p_s, P_t), p_s \in \partial p_s\}$ denote the point of O_j with minimum distance to robot.

The detected O_j must be expanded to make the selected P_{Gt} be feasible to get. Making two segments pass the left boundary point p_{sjl} and right boundary point p_{sjr} and vertical with detecting radial line respectively. The length of segments is $d/2$, d is the diameter of robot. And the end-points of segments are $\overline{p}_{sjl}$ and $\overline{p}_{sjr}$ respectively. Connect the point P_t and $\overline{p}_{sjl}$ cross the boundary $\partial W(t)$ at p_{vjl}, and connect the point P_t and $\overline{p}_{sjr}$ cross the boundary $\partial W(t)$ at p_{vjr}.

Definition 2.4: For the adjacent O_j and $O_{(j+1)}$,if $\theta_{vjr} < \theta_{v(j+1)l}$, Where the θ_{vjr} and $\theta_{v(j+1)l}$ denote respectively the angle from line $\overrightarrow{P_t p_{vjr}}$ and $\overrightarrow{P_t p_{v(j+1)l}}$ to X axis,then arc $\overset{\frown}{p_{vjr} p_{v(j+1)l}}$ is the feasible selection range for P_{Gt} between O_j and the adjacent $O_{(j+1)}$. Obviously $\overset{\frown}{p_{vjr} p_{v(j+1)l}} \in \partial W(t)$. $(\bigcup_{j=1}^{m-1} \overset{\frown}{p_{vjr} p_{v(j+1)l}})$ denote the feasible selection region of P_{Gt} at time t, i.e. P_v, obviously $P_v \subset (VD(t) \bigcap \partial W(t))$.

$$p_v = \left\{ p \mid p \in (\bigcup_{j=1}^{m-1} \overset{\frown}{p_{vjr} p_{v(j+1)l}}), \theta_{vjr} < \theta_{v(j+1)l} \right\}$$

Next step is the selection of P_{Gt}. If the $end \in W(t)$ at t, then $P_{Gt} = end$; Otherwise, we use heuristic function $f(p) = g(p) + h(p)$ to choose the $p \in p_v$ as P_{Gt} which can minimize $f(p)$, i.e.

$$\min_p f(p) = g(p) + h(p) \tag{4}$$

$$\text{s.t.} \qquad p \in p_v$$

Where $g(p)$ is the cost function from present location to p, and $h(p)$ is cost from p to end. Commonly, we only select $d(p, end)$ as $h(p)$ to estimate $f(p)$, since the information beyond the boundary is unknown. And that, the calculation of $g(p)$ is expressed as below:

$$g(p)=\begin{cases}0, & p\in p_v\\ \infty, & p\notin p_v\end{cases} \tag{5}$$

And then, the selection of P_{Gt} reduces to the optimal problem as below.

$$\min\ J\ =\min\ d(p,end) \tag{6}$$

$$\text{s.t.}\qquad p\in p_v$$

Algorithm 1: Path planning method based on rolling windows

step 1 Initialize the start point $(start)$, the end point (end),the detecting radius r_d and the step distance ε;

step 2 Terminate the plan if (end) arrived;

step 3 Refresh the information of rolling window $\partial W(t)$;

step 4 Select P_{Gt} as local goal, and radius line $\overrightarrow{P_tP_{Gt}}$ as local path of time t;

step 5 Move toward P_{Gt} along $\overrightarrow{P_tP_{Gt}}$ with ε and arrive the position at time $t+1$;

step 6 Return to step2;

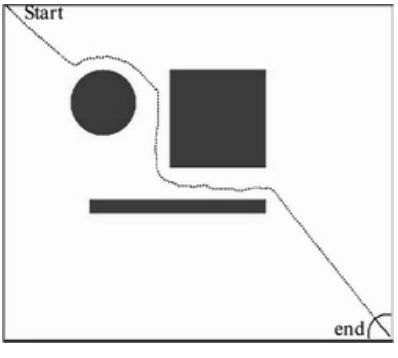


Fig. 3. Simulation of Algorithm 1 of path planning

The simulation of algorithm 1 is carried out to study the effect of this method. We know that it is effective for the simple map from the simulation result, as shown in Fig.3. However, the computation cost will be increased dramatically when more obstacles are contained in the environment, as the algorithm adopts one-step strategy in each rolling planning period. Therefore we attempt to improve above algorithm by using multi-steps strategy based on local planning of RRT. The detailed algorithm is described in the next subsection.

3 Adaptive Rolling Path Planning Methods

RRT can be regarded as a powerful tool for local path planning when intermediate configurations are inputted as planning queries. Different from above one-step planning

strategy, multi-steps solution path can be returned by the RRT planner, resulting in high performance in finding a global feasible solution path.

Since local goal configuration is not specified in the rolling window in advance, we introduce goal-oriented heuristic strategy to generate a collision-free path through tree expansion, and the algorithm is described as follows:

Algorithm 2: adaptive local path planning algorithm:

step 1: initialize a tree structure T rooted at the current position of planet rover: q_{init}.

step 2: sample a random configuration in the configuration space, denoted by q_{rand};

step 3: find the nearest neighboring of q_{rand} in T: q_{near};

step 4: extend one step from q_{near} to q_{rand}, arrive at a new configuration q_{new};

step 5: if q_{new} is collision-free, and it lies in the maximum permitted region specified by a circle with variable-radius r, then q_{new} is added as a new node to T;

step 6: Repeating step 2-step 5 till the maximum permitted number of iteration has been reached;

step 7: search for a node in T that is closest to the *end* position, denoted by q_{goal};

step 8: find a collision-free path from q_{init} to q_{goal} by traversing the expanded tree T;

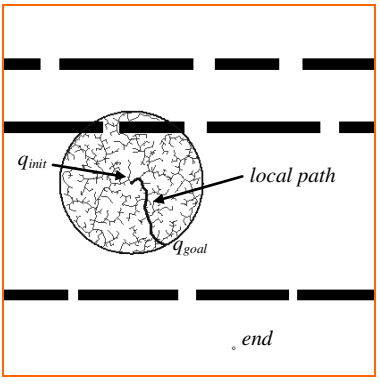


Fig. 4. Process of RRT-based local path planning

The process of RRT-based local path planning algorithm is illustrated in Fig. 4. It's observed that RRT structure can effectively capture the connectivity of local region. As a result, multi-steps solution path can be returned in a required computation time.

When the planet rover arrives at a new position q_{init}, the update of the radius r of the maximum permitted exploring region is designed as below:

Firstly we assigned an n-dimensional Gaussian vector v_k with zero-mean vector and covariance matrix $\sum_k$ to the planner. As the environment to be explored is not known to the planner, it's non-trivial to fix the value of $\sum_k$. As a result, the algorithm starts with an arbitrary covariance matrix $\sum_0$. In each time step t, $\sum_k(t)$ is updated based on the evolution of the historical radius of the exploring regions h-step ahead, that is,

$$\sum_k(t) \leftarrow update(r(t\text{-}h), r(t\text{-}h\text{+}1), \ldots, r(t\text{-}2), r(t\text{-}1)) \quad (7)$$

If the algorithm can not find a feasible solution path in the restricted circle region, implying that the planet rover lies in the extremely narrow passages. We then scale the radius r by a fixed factor δ which is set to the range of (1, 2]:

$$r(t) \leftarrow r(t)/\delta \quad (8)$$

The above feedback process iterates until the algorithm returns a feasible local path.

Once the local path is returned successfully by the RRT-base path planner, global path planning algorithm can be easily constructed. The detailed algorithm is described as follows:

step 1 Initialize the start point (*start*), the end point (*end*);
step 2 Terminate the procedure if *end* is arrived;
step 3 Refresh the information of rolling window;
step 4 call RRT-based path planning algorithm to find a local path in the region restricted by the rolling window;
step 5 Move to the goal position q_{goal} found by RRT-based path planning algorithm;
step 6 Return to step2;

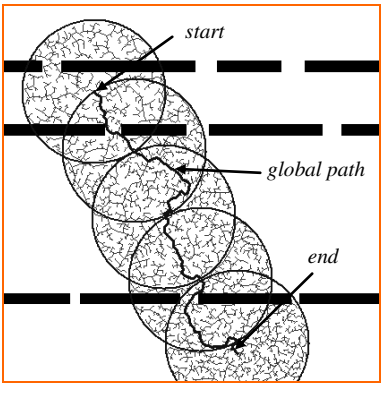


Fig. 5. Process of global path planning

The process of global path planning algorithm is illustrated in Fig. 5. It's observed that global feasible path is successfully returned by successively connecting local paths found by RRT-based local path planning algorithm. The number of rolling windows is five in this experiment, far less than one-step algorithm. In this simulation test, the radiuses of the explored local regions are not changed in each time step, as no restricted regions are not perceived all through the running procedure. Equipped with the effective local path planner, the planet rover doesn't need to perceive the environment by sensors frequently, resulting in significant improvement of the performance.

4 Experimental Results

We have implemented the proposed adaptive rolling path planning algorithm based on RRT exploration, and validated the algorithm on the platform of our planet rover with double-eyed vision sensors and six-wheeled activators. The testing environment and mechanical structure of the robot is illustrated in Fig. 6.

Fig. 6. Experiments on planet rover

It's observed from Fig. 6 that the planet rover can successfully avoid the obstacles during the local planning process while keep the validity of the global path. The whole exploring area in the experiment is restricted in a square with the dimension 15*15m^2, robot speed is 1km/h, and the comparison of computation time between original rolling planning algorithm and our algorithm is shown in table 1. Our algorithm gives about 3 times improvement in the running time over the original algorithm.

Table 1. Comparison of experiments time

	Time(sec)
original algorithm based on Rolling Windows	194.4
improved algorithm based on the goal-oriented heuristic RRT	76.2

5 Conclusion

We have presented a novel adaptive path planning algorithm for the planet rover in uncertain environment based on rolling window and RRT algorithm. Firstly we presented an effective representation of the local environment, which is suitable for planning procedure. Next, the rolling windows methodology has been introduced which is able to plan an avoidance path in simple environment. A more efficient local path planning method based on the goal-oriented heuristic RRT algorithm is applied to rapidly find a local feasible path in the rolling window, resulting in substantial improvement of the performance. The global collision-free path is generated by successively connecting local feasible paths when continuously pushing the rolling window toward the *end* position. The experimental results demonstrate the validity of the proposed method. The maximum permitted regions of the local path planning can be adaptive adjusted by the historical explored information, which forms the unique feature of our planner.

Acknowledgement

Supported by National Natural Science Foundation of China (No: 60774076).

References

1. Latombe, J.C.: Robot Motion Planning. Kluwer Academic Publishers, New York (1990)
2. Canny, J.F.: The Complexity of Robot Motion Planning. MAT Press, Cambridge (1988)
3. Diguez, A.R., Sanz, R., Fernndez, J.L.: A global motion planner that learns from experience for autonomous mobile robots. J. Robotics and Computer-Integrated Manufacturing 23(5), 544–552 (2007)
4. Xi, Y.G.: Predictive control. National Defense Industry Press, Beijing (1993)
5. Zhang, C.G., Xi, Y.G.: Robot path planning in globally unknown environment based on rolling windows. Science in China 44(2), 131–139 (2001)
6. LaValle, S.M.: Rapidly-exploring Random Trees: A New Tool for Path Planning, in TR 98-11, Computer Science Dept., Iowa State University (1998)
7. Kim, J., Ostrowski, J., Motion, C.: planning of aerial robots using Rapidly-exploring Random Trees with dynamic constraints. In: Proceedings of the IEEE International Conference on Robotics and Automation (2003)
8. LaValle, S.M.: Planning Algorithms. Cambridge University Press, Cambridge (2006)
9. Zhang, C.G., Xi, Y.G., Robot, J.: rolling path planning based on locally detected information. Acta Automatica Sinica 29(1), 38–44 (2003)
10. Urmson, C., Simmons, R., Approaches, C.: for heuristically biasing RRT growth. In: Proceedings of the IEEE International Conference on Intelligent Robots and Systems (2003)
11. Kuffner, J.J., LaValle, S.M.: RRT-connect: An Efficient Approach to Single-query Path Planning. In: IEEE International Conference on Robotics and Automation, pp. 995–1001 (2000)

Planning and Control of Biped Walking along Curved Paths on Unknown and Uneven Terrain

GuoQing Zhang, Ming Xie, Hang Yin, Lei Wang, and HeJin Yang

School of Mechanical and Aerospace Engineering, Nanyang Technological University, 50 Nanyang Drive, Research TechnoPlaza, Level 4, BorderX Block, Singapore 637553
{zhanggq,mmxie,hyin,wanglei,hjyang}@ntu.edu.sg

Abstract. This paper investigates the planning and control of biped walking along curved paths on unknown and uneven terrain. For widespread use of biped robots, the capability of walking on unknown and uneven terrain is essential. The description of uneven terrain, as the basis of discussion, is presented in terms of the spatial relationship between the world frame and the robot local frames. Then the principle and implementation of walking pattern planning are given phase by phase. Feedback controllers, including the phase switching controller, the stabilizing controller and the foot landing controller, are also designed to guarantee stable and agile walking. The motion planning, path following, and controller design are discussed within the identical framework, such that various walking behaviors can be generated with few walking parameters modified. Some results of simulation and experiments performed on the LOCH robotic platform are given to show the effectiveness of the proposed approach.

1 Introduction

Biped robots have been the focus of both academic and industry community for decades. The biped, compared with robots in other forms, features the potential to walk stably and agilely like human beings, which comes with prominent advantages in autonomous exploration and environment adaption [1]. Great efforts have been made to turn the potential into reality, and people have seen many encouraging progresses, such as the ASIMO [2], HRP [3] and HUBO [4].

At the same time, in terms of biped walking, plenty of efforts have been made toward human-level motion capability. Related subjects include biped walking stability [5,6], walking pattern generation [7,8,9,10], navigation and autonomous exploration[11], and biped walking on uneven terrain [12,13,14,15]. These researches independently investigate the biped walking from different point of view, and the proposed approaches are effective for particular occasions. However, successful application of biped robots in real world simultaneously requires the capability of stable walking, path following and environment adaption. The objective of this paper is, therefore, to discuss these problems within an identical framework.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 1032–1043, 2009.

Recently Nanyang Technological University is developing a low cost humanoid robot, LOCH. It is built up with the intention of developing an adult-sized humanoid robot, which is skillful, knowledgeable and affordable. The robot measures 1.8 meters in height, and weighs 80 kilograms. It is capable of dexterous manipulation and complex biped locomotion, and it also features effective vision system and linguistic interaction ability. Issues about the system design have been presented in[16]. This paper aims to give a theoretical approach that will enable the LOCH robot to accomplish complex locomotion on uneven terrain.

Walking on unknown and uneven terrain remains a challenging problem, in which the environment adaptability is the major concern. In the proposed work, the description of uneven terrain is presented first. The posture of the robot pelvis is determined from sensory feedback. Then the unevenness of the terrain is described in terms of the spatial relationship of the foot frame and the world frame. On the basis of this description, the walking pattern generation and compliant control are designed to be adaptive to terrain conditions.

Approaches of path planning fall into three catalogues: offline trajectory planning, offline trajectory planning with online feedback compensation and online trajectory planning. The offline planning causes the robot fail to duly respond to external events. In the proposed framework, an online planning scheme is designed in such a way that motion trajectories are generated phase by phase. Thus the walking pattern can be flexibly modified to accommodate changes in the environment.

The remainder of the paper is organized as follows. Sect. 2 presents the setup of coordinate frames and the description of uneven terrain. Sect. 3 and Sect. 4 respectively describe the principle of trajectory planning and controller design. Sect 5 gives the simulation and experiment results, and Sect. 7 is the conclusion.

2 Preliminaries

2.1 Setup of Coordinate Frames

In this paper we consider a 7-link robot that is shown in Fig 1. Each hip has three degrees of freedom (DOF), each knee has one DOF, and each ankle has two DOF. Such a distribution of DOF is employed by the majority of existing humanoid robots. For the purpose of describing the walking behavior, we set up coordinate frames for the world and for the pelvis and dual feet.

$\{XYZ\}$ is the world frame with the $\{XY\}$ plane parallel to the horizontal plane. $\{\bar{x}\bar{y}\bar{z}\}$ moves together the pelvis with $\bar{y}$ pointing to the front and $\bar{x}$ pointing to the left. $\{\tilde{x}\tilde{y}\tilde{z}\}$ locates at the left ankle. $\tilde{y}$ points to the front along the sagittal middle line of the foot and $\tilde{x}$ points to the left. Similarly $\{\hat{x}\hat{y}\hat{z}\}$ is at the right ankle. For simplicity the frames $\{XYZ\}$, $\{\bar{x}\bar{y}\bar{z}\}$, $\{\tilde{x}\tilde{y}\tilde{z}\}$ and $\{\hat{x}\hat{y}\hat{z}\}$ are respectively denoted by $\{W\}$, $\{PV\}$, $\{LF\}$ and $\{RF\}$.

2.2 Walking Cycle and Phases

Biped walking is a periodic phenomenon, and every walking cycle is composed of four phases (see Fig 2). Phase 1 and Phase 2 comprise the first step, and Phase

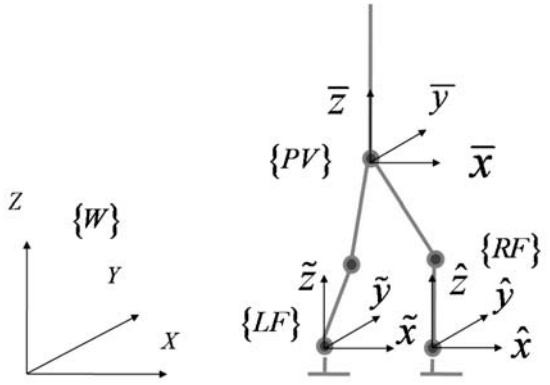


Fig. 1. Setup of coordinate frames

3 and Phase 4 comprise the following. The roles of dual legs in every walking cycle are described in Table 1. The walking planning is done phase by phase. That is, the planning for every single phase is performed only when the walking progress has reached the end of the previous one. By dong this, the robot is able to change walking parameters in adaption to external conditions at the end of each phase.

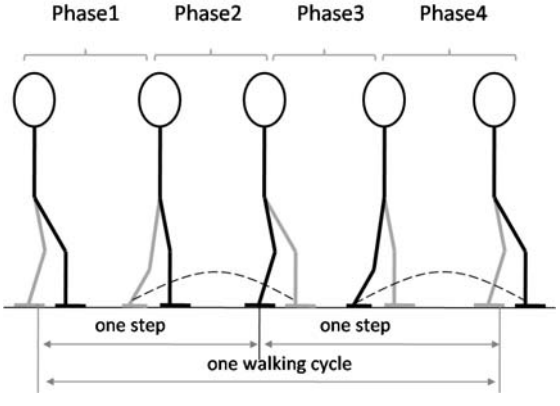


Fig. 2. Phases in a walking cycle

2.3 Description of the Uneven Terrain

When walking on unknown and uneven ground, the robot adopting walking patterns that are planned for even terrain is probable to overturn. The walking pattern must be generated according to the terrain conditions to guarantee walking stability. Then we are facing the problem of how to describe the uneven terrain. Throughout a stable walking process, at least one foot is on the ground

Table 1. Roles of dual legs in a walking cycle

	Left Leg (grey)	Right Leg (black)
Phase 1	Push	Support
Phase 2	Swing	Carry
Phase 3	Support	Push
Phase 3	Carry	Swing

and only the area beneath this foot affects the walking behavior. In view of this fact, it's reasonable to use the corresponding foot frame as the indicator of terrain unevenness. First we give some notations below.

1. Relative to frame $\{W\}$, the frame $\{LF\}$ is described by the homogeneous transform ${}^{W}_{LF}T$, which has the form

$$ {}^{W}_{LF}T = \begin{bmatrix} {}^{W}_{LF}R & {}^{W}_{LF}P \\ 0 \quad 0 \quad 0 & 1 \end{bmatrix} \tag{1} $$

   ${}^{W}_{LF}R$ is the rotation matrix and ${}^{W}_{LF}P$ is the origin coordinates of $\{LF\}$ relative to $\{W\}$. Extending the notations to frames $\{RF\}$ and $\{PV\}$, we obtain ${}^{W}_{RF}T$, ${}^{W}_{PV}T$, ${}^{RF}_{W}T$, ${}^{PV}_{W}T$, etc, and the rotation matrices and origin coordinates thereof.
2. Given a rotation matrix R, the corresponding Euler angles are denoted by $(\gamma(R), \beta(R), \alpha(R))$.
3. Given a group of Euler angles (γ, β, α), the corresponding rotation matrix is denoted by $R(\gamma, \beta, \alpha)$.

Assume ${}^{W}_{PV}R$ is detectable with sensory feedback. This assumption can be satisfied by the installation of a three-axis gyro or an inclination sensor on the robot trunk. The origin coordinates of the pelvis frame, ${}^{W}_{PV}P$, can be calculated using a certain navigation technique. Then we obtain ${}^{W}_{PV}T$. At the same time, with forward kinematics, ${}^{PV}_{LF}T$ and ${}^{PV}_{RF}T$ are also available at any instant by reading joint angles. Then we have

$$ \begin{pmatrix} {}^{W}_{LF}T \\ {}^{W}_{RF}T \end{pmatrix} = \begin{pmatrix} {}^{W}_{PV}T \, {}^{PV}_{LF}T \\ {}^{W}_{PV}T \, {}^{PV}_{RF}T \end{pmatrix} \tag{2} $$

During the walking process, as soon as either robot foot has established firm contact with ground, the homogeneous transform ${}^{W}_{LF}T$ or ${}^{W}_{RF}T$, whichever corresponds to the foot on ground, is known and will remain unchanged till the next swing action. Utilizing this property, we present the following proposition.

Proposition 1. *The inclination of terrain is described by the homogeneous transform that relates the frame of the foot on ground to the world frame, as computed by (2).*

3 Walking Pattern Planning

3.1 Dynamic Planning of the Curved Path

Path planning is performed at the beginning of Phase 2 and Phase 4. Plenty of methods can be used to generate paths for different tasks, and here we assume the path has been generated and denoted by $P(\rho) = (X_P(\rho)\ Y_P(\rho))$ where $\rho \in [0\ \rho_{end}]$. The path is described within the $\{XY\}$ plane of the frame $\{W\}$, and satisfies Assumption 1 and Assumption 2. Assumption 1 means the path starts from the current position and ends at the goal position, and Assumption 2 means the velocity is continuous along the path.

Assumption 1. $P(0)$ corresponds to the initial location of the robot, and $P(\rho_{end})$ corresponds to the final location.

Assumption 2. $P(\rho)$ is first order differentiable with respect to ρ. $P'(0)$ has the same direction with the initial orientation of the robot.

3.2 Dynamic Planning of the Pelvis for Phase 1

In Phase 1 both feet of the robot are on the ground. The left leg is the push leg and the right leg is the support leg. According to the inverted arm model, we select the frame $\{RF\}$ as the base frame and $\{PV\}$ as the tool frame. That is, the motion of the pelvis is described relative to $\{RF\}$. As shown in Fig 3, the position and orientation of the pelvis are denoted by a 6×1 vector $p(t)$. The first three rows of $p(t)$ are the Cartesian coordinates and the last three rows are the Euler angles.

From the figure we can find that $p(0)$ and $p(t_m)$ are right above the inner borderline of the supporting polygon. Their positions are determined in connection with the walking speed. If the robot walks at full speed, $p(0)$ is at the

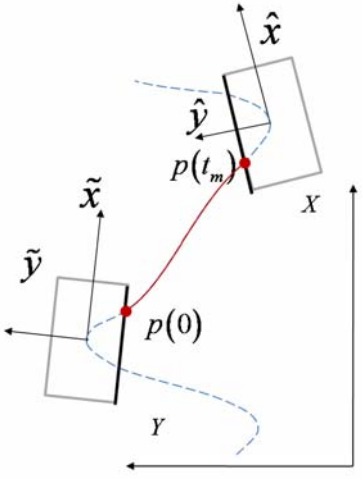


Fig. 3. Pelvis trajectory for Phase 1

right upper corner of the left supporting polygon, and $p(t_m)$ is at the left lower corner of the right supporting polygon. If the robot marches in place, $p(0)$ and $p(t_m)$ are both in the middle of the corresponding borderline. In this context the supporting polygon is not necessarily the foot outline. It can be smaller than the foot outline to achieve static walking with some stability margin, and can be smaller to achieve dynamic walking.

The clamped 3^{rd} order polynomial interpolation is used to generate trajectory. $p(0)$ can be calculated by forward kinematics, and $\dot{p}(0)$ is known from the previous walking cycle. The other boundary conditions are given in (3)–(5).

$$p(t_m) = \begin{pmatrix} {}^{RF}_{W}T \left({}^{W}_{RF}T \begin{pmatrix} \left(\frac{1}{2} - q\right) l_f \\ \frac{w_f}{2} \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ h_p \end{pmatrix} \right) \\ \gamma\left(R_p(t_m)\right) \\ \beta\left(R_p(t_m)\right) \\ \alpha\left(R_p(t_m)\right) \end{pmatrix} \tag{3}$$

$$\dot{p}(t_m) = \begin{bmatrix} I_3 & 0 \\ 0 & 0 \end{bmatrix} \frac{1}{t_m} \left(p(t_m) - p(0)\right) \tag{4}$$

$$q = \frac{l_s}{2l_s^{max}} + \frac{1}{2} \tag{5}$$

In (3)–(5), l_f and w_f are respectively the length and width of the physical supporting polygon. h_p is the desired hip height. l_s is the step length and l_s^{max} is the maximal step length allowed. $R_p(t_m)$ is the orientation matrix of $p(t)$ at instant $t = t_m$, which satisfies

$$R_p(t_m) = {}^{RF}_{W}R \quad {}^{W}R\left(0 \;\; 0 \;\; \alpha\left({}^{W}_{RF}R\right)\right) \tag{6}$$

3.3 Dynamic Planning of the Foot Landing Position

In Phase 2 the left leg swings from the current pose to a new one, thus we must plan the landing position before the swing starts. Suppose $\{LF\}$ and $\{RF\}$, relative to the $X - Y$ plane, are described respectively by $\{(X_l, Y_l, \theta_l)\}$ and $\{X_r, Y_r, \theta_r\}$, which can be obtained by extracting the x-coordinate, the y-coordinate and the roll angle from ${}^{W}_{LF}T$ and ${}^{W}_{RF}T$. Then we give the following proposition.

Proposition 2. *Relative to the $\{XY\}$ plane, the spatial description of the robot at the jth step is denoted by (7) or (8), whichever is leading along the walking direction and corresponds to the foot on ground.*

$$P_j : \begin{pmatrix} X_j \\ Y_j \\ \theta_j \end{pmatrix} = \begin{pmatrix} X_l + w_s \sin\theta_l \\ Y_l - w_s \cos\theta_l \\ \theta_l \end{pmatrix} \tag{7}$$

$$P_j : \begin{pmatrix} X_j \\ Y_j \\ \theta_j \end{pmatrix} = \begin{pmatrix} X_r - w_s \sin\theta_r \\ Y_r + w_s \cos\theta_r \\ \theta_r \end{pmatrix} \tag{8}$$

With Proposition 2, we present the following algorithm to search for landing positions of the swing foot. Suppose the current robot location P_j corresponds to ρ_j in the path.

1. Set the desired step length l_s^d and the maximal turning angle allowed for every single step δ_θ^{max}. Let $\rho_t = \rho_j$.
2. $\rho_t = \rho_t + \delta_\rho$.
3. $|P(\rho) - P_j| < l_s^d$? If yes, goto 4, otherwise goto 5.
4. $\angle(P(\rho) - P_j) < \delta_\theta^{max}$? If yes, goto 2, otherwise goto 5.
5. P_{j+1}=$P(\rho_t - \delta_\rho)$, $\theta_{j+1} = \angle(P(\rho) - P_j)$, $l_s = |P_{j+1} - P_j|$.

After we obtain P_j from this algorithm, the landing position can be accordingly computed using (7) or (8), which depends on the swing foot being the left foot or the right one.

3.4 Dynamic Planning of the Pelvis for Phase 2

In Phase 2, the left foot is the swing foot and the right foot is the carry foot. According to the inverted arm model, we select the right foot as the base link, and the right hip as the end effector. The planning is described relative to $\{RF\}$.

As shown in Fig 4, $p(t_m)$, $p(t_{mid})$ and $p(T/2)$ are critical points. The boundary conditions can be determined using geometric relationship in the similar way as in Phase 1.

3.5 Dynamic Planning of the Swing Foot

Because of the unevenness of terrain, the landing process must be handled carefully to avoid foot-ground collision. If the foot hits ground earlier than expected, the ground reacting force will be a serious threat to stability. To the contrast, if the swing foot cannot reach ground in the end, the robot is probable to overturn in the following phase. In view of these facts, we partition Phase 2 into two subphases: Subphase 2A and Subphase 2B.

Subphase 2A is a free swing phase, in which the swing foot moves until it reaches its nominal final pose, or hits the ground earlier than expected. The description of trajectory is relative to the frame $\{PV\}$. As shown in Fig 4, three critical points are involved in the planning. The boundary conditions can be determined from geometric relationship. If the swing foot hits the ground earlier than expected, Subphase 2A is terminated immediately and Subphase 2B will accordingly be started. Physically several on/off sensors are installed on the foot outline for collision detection.

In Subphase 2B, the motion of the swing foot is controlled by the landing controller. In the second half of Phase 2, the pelvis moves down along the Z-axis

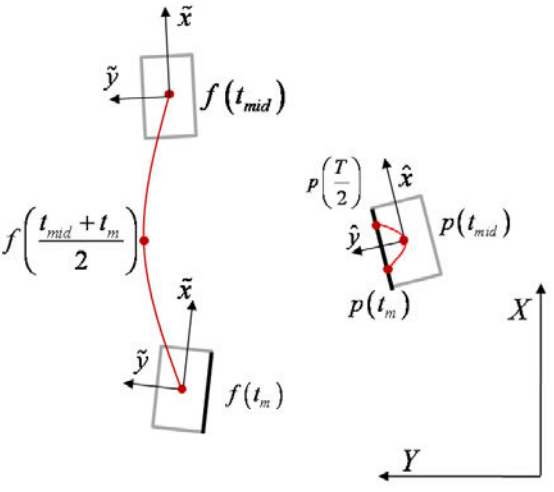


Fig. 4. Planning of swing trajectory for Phase 2

direction. By dong this, the pelvis exerts a downward push force on the swing foot. At the same time the landing controller provides the compliant landing capability, thus the posture of the swing foot is forced to be compliant with the terrain. Once firm ground contact is reported by pressure sensors on the foot sole, the swing foot will keep its posture till the next phase starts. If the swing foot cannot reach ground at the end of Subphse 2B, the walking process will be redirected to the safe pose.

4 Control System Design

4.1 Phase Switching Controller

In a walking process, the phase switching is distinguished into normal switching and emergency switching. The flowchart of phase switching and the critical instants are shown in Fig 5.

In normal switching, the phases are assembled in the sequence of Phase 1 $\rightarrow$ Phase 2 $\rightarrow$ Phase 3 $\rightarrow$ Phase 4 $\rightarrow$ Phase 1. The switching instants in every walking cycle are respectively t_m, $T/2$, $T/2+t_m$ and T. In emergency switching, which is used to handle emergency stop requests, the switching controller will redirect the walking procedure towards the safe pose by setting $l_s^d = 0$ and $h_s = \varepsilon$ with $\varepsilon \rightarrow 0$. Then the robot will first switch into the mode of marching in place to conquer the inertial force and then stop walking almost immediately.

4.2 Stabilizing Controller

Stability control is intended to overcome unexpected external disturbances and adapt to changes in walking conditions, and thus guarantee the walking stability.

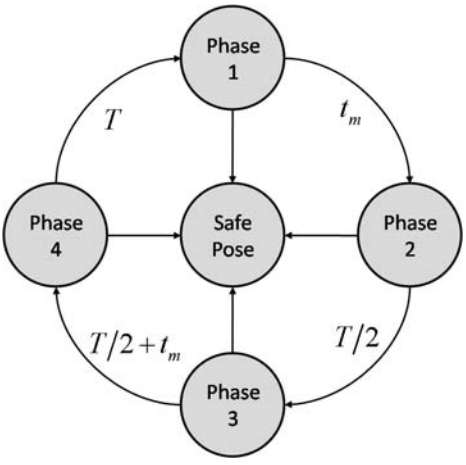


Fig. 5. Flowchart of phase switching

According to the inverted arm model and the leg stability concept, the principle of stabilizing control is illustrated in Fig 6. If the inverted arm outputs an active force $-F_a$ at the robot hip, the reacting force F_a can be used as the control input for the stability control of lower body. The computation of F_a can be based on monitoring the real-time ZMP, the upperbody inclination or other stability criterion.

4.3 Foot Landing Controller

Foot landing controller is intended for compliant landing for Phase 2 and Phase 4. Suppose the effect of gravity has been compensated, the sensor reading relative

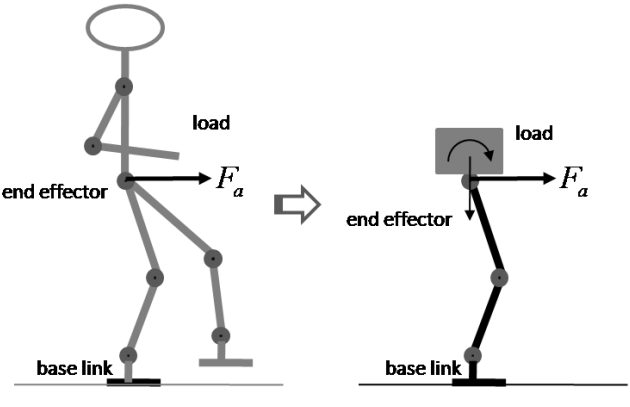


Fig. 6. Inverted arm and the stabilizing control

to the foot frame is denoted by $(f_x, f_y, f_z, \tau_x, \tau_y, \tau_z)$, and output of the landing controller is $(\Delta_z, \Delta_y, \Delta_x, \Delta_\gamma, \Delta_\beta)$. Taking the x-coordinate of the left foot in Phase 2 for example, the output of the left foot landing controller satisfies

$$c_x \dot{\Delta}_x + k_x \Delta_x = {}^{PV} f_x \tag{9}$$

where ${}^{PV} f_x$ is the x-component of the detected force, relative to a frame that is fixed on the left ankle and has the same orientation as $\{PV\}$. c is the damping coefficient, and k is the stiffness coefficient. The output of the landing controller is used to control the landing foot directly.

4.4 Inclination Detector

Once the swing foot has landed firmly, the inclination detector is activated to establish the homogeneous transform description of the swing foot relative to $\{W\}$. The firm landing is reported by the pressure sensors mounted on the sole. At this critical instant, the output of the foot landing controller $(\Delta_z, \Delta_y, \Delta_x, \Delta_\gamma, \Delta_\beta)$ will be locked. Then the terrain inclination can be detected using Proposition 1.

5 Simulation and Experiments

First we examine the proposed method for path following. Given a circular path and a sinusoidal path, we obtain Fig 7 that shows the planned steps along the paths. It can be found from these steps that the robot position (marked by cross) moves forward exactly along the given paths. At the same time, the robot orientation changes gradually by modifying the orientation of the swing foot. It's clear that the landing position planning method is feasible in dealing with different situations.

As even terrain is the special case of uneven terrain, in the following, we test the planning and control scheme through biped walking on even terrain first. For

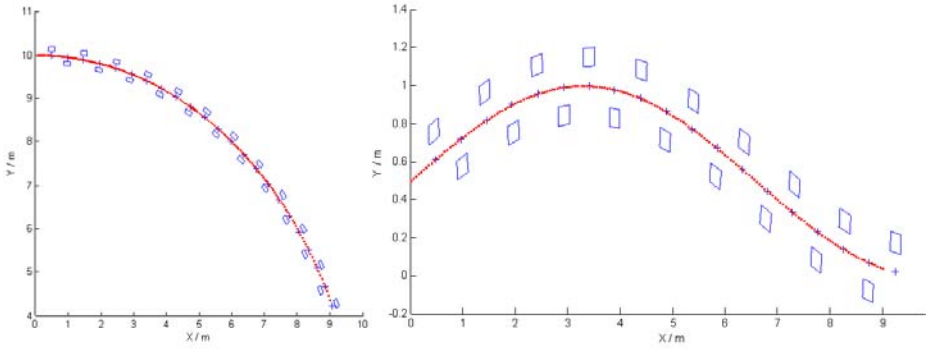


Fig. 7. Planning of the landing position of swing feet

this case, the landing time of the swing foot is predictable, and the Subphase 2B and Subphse4B are bypassed. Both simulations and experiments on the LOCH robot platform [16] have been carried out. The videos of the experiments can be found on http://picasaweb.google.com/lochrobot/Desktop, and thus will not be given here. From the above results, we can conclude that the presented scheme is effective for biped walking and path following on even terrain.

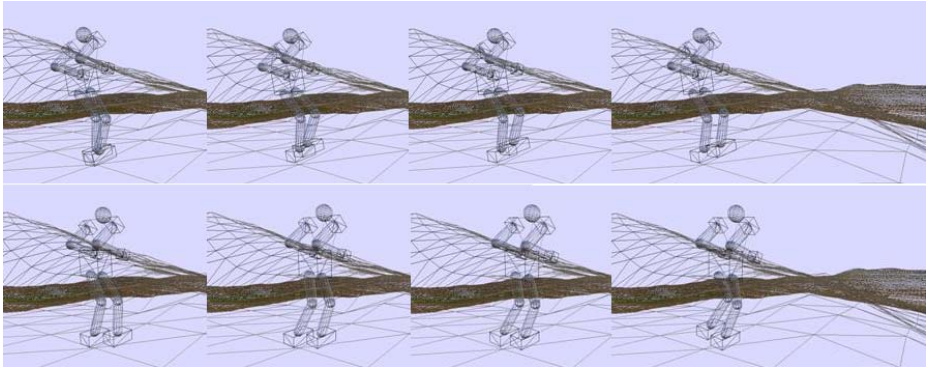


Fig. 8. Snapshots of the simulation for biped locomotion on uneven terrain

For the case of uneven terrain, simulations have been performed and the snapshots are given in Fig 8. From the snapshots it can be seen that the robot is able to walk stably, and thus we come to the conclusion that the proposed scheme is effective for biped locomotion on uneven terrain. Further experiments for uneven terrain are still in progress, and the results will be given in future publications.

6 Conclusion

This paper studies the planning and control of biped walking for following curved paths. Principle and implementation of path following, walking pattern generation and control system design are discussed in the frame of curved path following on uneven ground. The proposed scheme is able to generate various walking behaviors with different parameters specified. Simulation and some experiment results validate the effectiveness of the proposed scheme.

References

1. Xie, M., Dubowsky, S., Fontaine, J.G., et al.: Advances in climbing and walking robots. World Scientific, Singapore (2007)
2. Sakagami, Y., Watanabe, R., Aoyama, R., Matsunaga, C., et al.: The Intelligent ASIMO: System Overview and Integration. In: Proc. IEEE Int. Conf. on Intelligent Robots and Systems, pp. 2478–2483 (2002)

3. Kaneko, K., Kanehiro, F., Kajita, S., Hirukawa, H., et al.: Humanoid robot HRP-2. In: Proc. IEEE Int. Conf. Robotics and Automation, pp. 1080–1090 (2004)
4. Park, I., Kim, J., Lee, J., Oh, J.: Mechanical design of humanoid robot platform KHR-3 (KAIST humanoid robot -3: HUBO). In: Proc. IEEE Int. Conf. on Humanoid Robots, pp. 321–326 (2005)
5. Vukobratovic, M., Borovac, B.: ZMP - Thirty five years of its life. International Journal of Humanoid Robotics 1, 157–173 (2004)
6. Hurmuzlu, Y., Genot, F., Brogliato, B.: Modeling, stability and control of biped robots - a general framework. Automatica 40(10), 1647–1664 (2004)
7. Nagasaka, K., Inaba, M., Inoue, H.: Dynamic walking pattern generation for a humanoid robot based on optimal gradient method. In: Proc. of IEEE Int. Conf. on Systems, Man, and Cybernetics, vol. 6, pp. 908–913 (1999)
8. Yamane, K., Nakamura, Y.: Dynamics filter - concept and implementation of online motion generator for human figures. In: Proc. IEEE Int. Conf. on Robotics & Automation, pp. 688–695 (2000)
9. Nakaoka, S., Nakazawa, A., Yokoi, K., et al.: Generating whole body motions for a biped humanoid robot from captured human dances. In: Proc. of IEEE Int. Conf. on Robotics & Automation, pp. 3905–3910 (2003)
10. Sugihara, T., Takano, W., Yamamoto, K., et al.: Online dynamical retouch of motion patterns towards animatronic humanoid robots. In: Proc. IEEE-RAS Int. Conf. on Humanoid Robots, pp. 117–122 (2005)
11. Chestnutt, J., Lau, M., Cheung, G., et al.: Footstep planning for the Honda ASIMO humanoid. In: Proc. IEEE Int. Conf. on Robotics and Automation, pp. 629–634 (2005)
12. Shimizu, H., Wakazuki, Y., Pan, Y., Furuta, K.: Biped walking robot using a stick on uneven ground. In: Proc. SICE Annual Conference, pp. 83–88 (2007)
13. Huang, W., Chew, C., Zheng, Y., Hong, G.: Pattern generation for bipedal walking on slopes and stairs. In: Proc. IEEE Int. Conf. on Humanoid Robots, pp. 205–210 (2008)
14. Hashimoto, K., Sugahara, Y., Kawase, M., Ohta, A., et al.: Landing pattern modification method with predictive attitude and compliance control to deal with uneven terrain. In: Proc. IEEE Int. Conf. on Intelligent Robots & Systems, pp. 1755–1760 (2006)
15. Nishiwaki, K., Kagami, S.: Walking Control on Uneven Terrain with Short Cycle Pattern Generation. In: Proc. IEEE Int. Conf. on Humanoid Robots, pp. 447–453 (2007)
16. Xie, M., Zhong, Z.W., Zhang, L., Xian, L.B., et al.: A deterministic way of planning and controlling biped walking of LOCH humanoid robot. Industrial Robot - An International Journal 36(4), 314–325 (2009)

Experimental Study on Alpine Skiing Turn Using Passive Skiing Robot

Norihiko Saga and Kengo Kono

Kwansei Gakuin University, 2-1 Gakuen, Sanda,
669-1337 Hyogo, Japan
{saga,blo24449}@kwansei.ac.jp

Abstract. In recently, study on skiing turn is researched from various viewpoints. And the mechanism of the skiing turn using the skiing robot is clarified now. However, those mechanical models derived by using the approximate expressions don't match to an alpine skiing turn. Therefore, to provide further details of theoretical consideration, the passive type skiing robot is developed. The influence on the skiing turn such as the position of gravitational center and the shape of skiing are examined by using this robot. In this paper, it reports on their experimental results.

Keywords: Ski, Passive skiing turn robot, Human-Robot interaction, Mechanism.

1 Introduction

At present, many studies of turning [1]-[10] while skiing are being conducted on various viewpoints such as coupled problems that might result from combinations of skis and skiers, simulation of various modes of turning using robots, consideration of mechanical characteristics of skis, and possible loads burdening the skier's body when a ski turn is made. Specifically in studies using a ski robot, the body positions taken when a turn is made are reproduced using a robot based on the corresponding human body positions taken in the same situation. Such studies are undertaken to clarify the dynamics of different turns. However, because every mechanical model of turning thus derived from these studies is solving nonlinear differential equations with many approximations made simultaneously, the results do not fully represent the turns that an actual skier would make, nor do they explain anything of the velocity of turning. Furthermore, such mechanical models are so complex that, practically speaking, it is difficult for skiers to use the results to improve their ability as athletes in any Alpine event.

Under these circumstances, we considered the actual turn positions that skiers would make during Alpine skiing. We then analyzed the dynamics of every turn made during an Alpine skiing giant slalom competition and produced a passive-turn type of ski robot with a simple structure. Using the robot, we checked on how each turn might be affected by the skier's centre of gravity, by the timing of the skier's weight shift, and by the ski shape. Consequently, we examined the dynamics of ski turning to help skiers improve their abilities as athletes.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 1044–1050, 2009.

2 Alpine Ski Turning

Mainly, Alpine ski turning is a carving turn that is made with the skier's upper body bent forward, creating some angle between the snow surface and the ski with the skis bent simultaneously. To make that turn possible, the ski assumes a form by which its middle part is curved or hollowed, so to speak, which is called the side curve. Then the turn is made with the skis standing sideways against the snow surface, thereby using the side curve to the greatest degree possible (hereinafter, the angle taken between the snow surface and the ski is called the edge angle). The turn uses the centrifugal force resulting from the ski sliding speed while not losing any speed. The turn finally uses the arc-shaped trajectory that the ski itself creates. In the past, using straight skis, turns used to be made by skidding turns, which enable skidding of the skis created by a revolving motion of the skier's legs with the outside leg directed inward during the turn.

In carving turns made during Alpine ski competitions, in which the speed is crucial for competition, a sharper turn than that of a side curve can be achieved. However, in skidding turns, which necessitate a revolving motion of the skier's legs, the left and right legs are not aligned in the same direction. If they were aligned as parallel, they would cause a skid that would suppress the turn speed.

Fig. 1 shows carving turn behaviour during an actual competition. Fig. 1a is of a high-school-aged girl who is the number-one ranked skier in Akita prefecture, Japan. Fig. 1b portrays a fourth grade elementary school pupil who started competition less than a year ago. Defining a start point at which the centre of a skier's waist part is shifted from the right side of the ski to the left side, a top point at which the skis are directed straight downward, and a finishing point at which the centre of the skier's waist part is shifted from the right side of the ski to the left side, it is readily apparent that the athlete depicted in Fig. 1a reaches her top point just beside the flag gate, whereas the athlete portrayed in Fig. 1b reaches the top point after having passed the flag gate, causing the finishing point to be delayed. Looking at the respective body positions of the athletes during the turn, it is apparent that the athlete in Fig. 1a is making a turn keeping some angle between the snow surface and the skis while bending her skis.

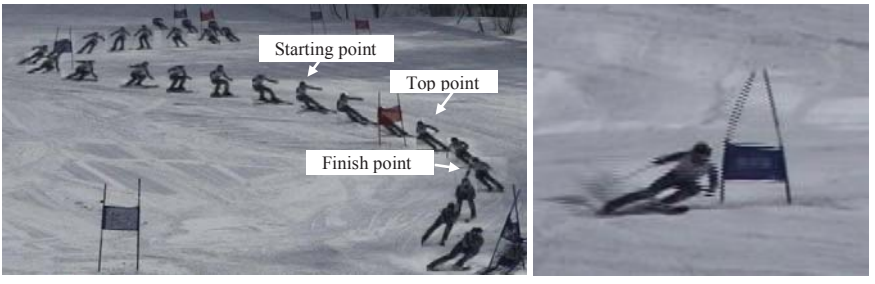
3 Passive-Turn Type of Ski Robot

3.1 Composition of Ski Robot

Next we examine a carving turn. To do so, we have produced a passive-turn type of ski robot that can use some angle between the snow surface and the skis by shifting its centre of gravity. Using it, we will verify experimentally how much impact is attributable to ski turning by the ski shape difference or by the centre of gravity position difference. The passive-turn type of ski robot is presented in Fig. 2.

The robot is made entirely of ABS resin. It weighs 50 g with skis attached. The robot is structured such that its legs and the body are connected by hinges, with its left and right legs separated stably using a support bar attached at the body part. Because the connections are made using hinges, the body part can move freely sideways. Then

a) Senior skier



b) Rookie skier

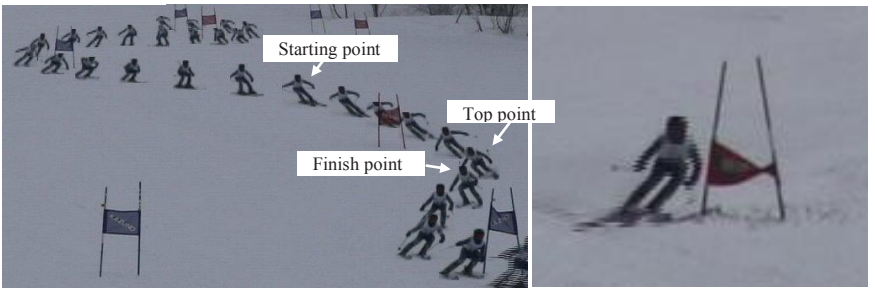


Fig. 1. Turn of alpine skiing

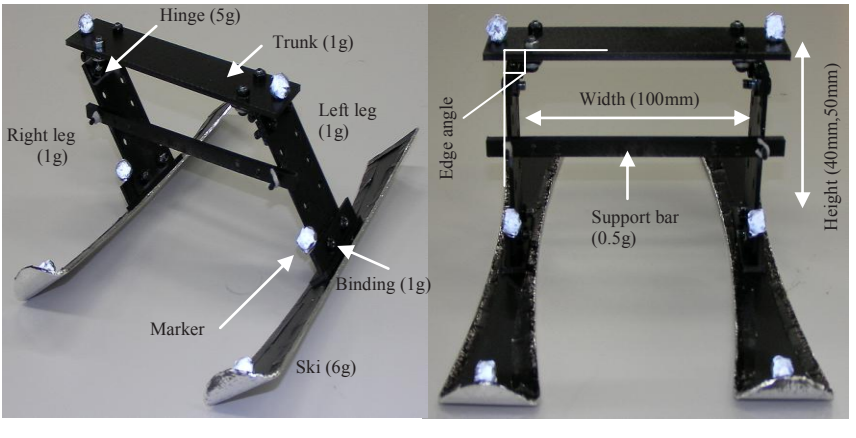


Fig. 2. Structure of passive turn type ski robot (50g)

we set the gap separating the left and right skis using the support bar such that the edge angle can be as great as ±35 deg. Each robot leg is connected firmly with a ski so that each leg is always perpendicular to each associated ski surface in its width direction, enabling determination of a ski edge angle as a result. Six white markers are attached to the robot, each at a different location; they are used for image analyses.

3.2 Form of Ski

Fig. 3 shows two pairs of skis, each of which is made of ABS resin. They are 1 mm thick, 170 mm long, and 35 mm wide at the front end and back end, with respective side curve radii *R* of 400 mm and 800 mm.

Because the ski robot we use for this experiment cannot bend the skis as a real skier can, we created and used two pairs of skis, each with a different side curve radius *R*, by which we were able to assume that the skis were bent.

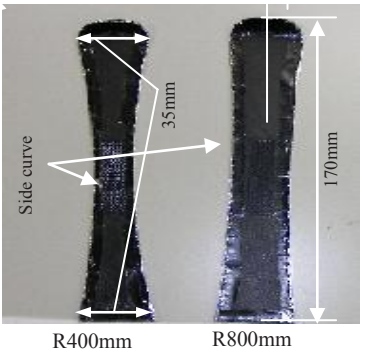


Fig. 3. Design of ski

4 Experiments

4.1 Method of Measurement

In the experiment using the passive-turn type of ski robot, we made measurements using three different CCD cameras and performed three-dimensional motion analyses. Fig. 4 shows the experiment. The X-, Y-, and Z-axes are presented also in Fig. 4. We placed one camera at the front part of the ski slope; the other two were placed respectively at 45 deg left and right from the front part. The slope angle was set to 25 deg.

We chose to use a carpet (0.9 m × 1.8 m) as a ski slope, with its fluff height sufficient for the skis to retain contact when a ski edge was angled. Stainless tape was placed on the lower surfaces of the skis to facilitate sliding. The ski direction ω was set on the slope along the Y-axis.

During the experiment, we observed the ski trajectory, the ski speed, and the timing of the centre-of-gravity shift. We verified it by making the ski robot turn one direction after another continuously while trying to use its body height of between 40 mm and 50 mm and its skis with side curve radii *R* of 400 mm and 800 mm. Initially, the skis were oriented to ω 0°.

4.2 Experimental Results and Examination

Fig. 5 presents results of the experiment in which the sliding trajectories of the ski robot were compared for waist heights of 40 mm and 50 mm, with the ski side curve radius *R* of 400 mm. Fig. 6 shows a similar comparison for radius *R* of 800 mm.

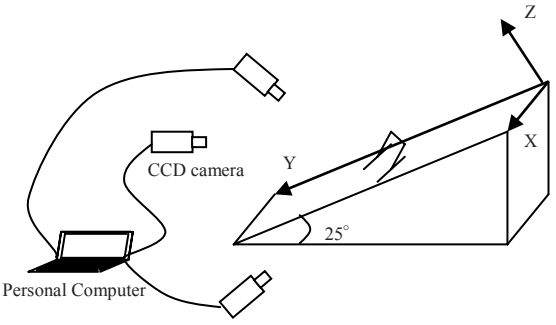


Fig. 4. Experimental setup

The right leg sliding start point was at the origin 0. The right leg trajectory coordinate points are expressed numerically with the X-coordinate (positive) on the right-hand side and Y-coordinate (positive) on the downward side, as viewed from the bottom of the slope.

Changing the side curve from R=400 mm to R=800 mm, it was recognized that the turn radius achieved by the robot increased accordingly along the curve connecting three points: the starting point of the turn; the top point, at which the skis were directed straight downward; and the end point of the turn. Using the robot's waist height of 50 mm, the turn radius was doubled from 280 mm to 560 mm, which was proportional to the doubled side curve radius R.

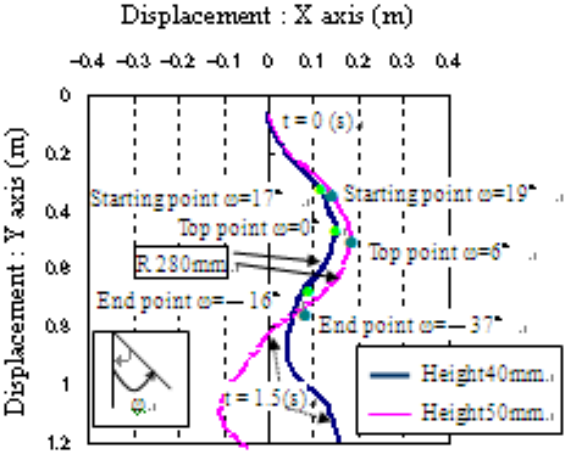


Fig. 5. Experimental results (Side Curve : R400mm)

Furthermore, changing the robot's waist height from 40 mm to 50 mm to find a possible impact caused by changing the centre of gravity in that way, we found that the turn cycle increased. However, although we found a difference when we used the side curve radius R=800 mm, we found no difference from the turn radius of R=400 mm.

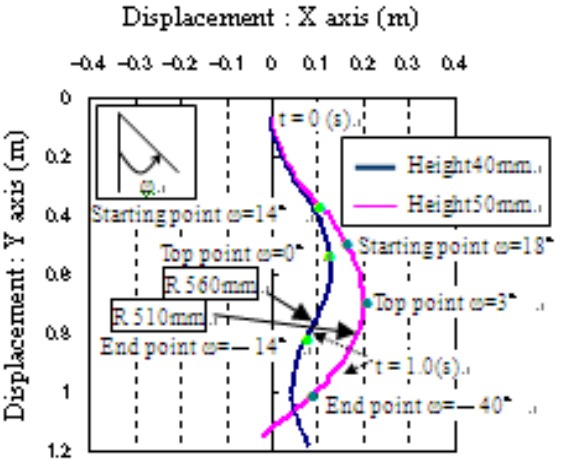


Fig. 6. Experimental results (Side Curve : R800mm)

Therefore, we conclude that the turn radius was determined based on the ski shape as long as the turn was made with no ski skidding. Next we examine the fact that the turn radius varies from a constant R=800 mm according to the waist height difference. First looking at the ski direction ω at each of the start point, top point, and finish point, it is apparent that the higher the waist part is, the larger ω we have. That phenomenon is explainable: the skis tend to skid because of failure to catch the slope steadily as the centre of gravity position increases, even when we have a fixed edge angle and a fixed amount of centrifugal force.

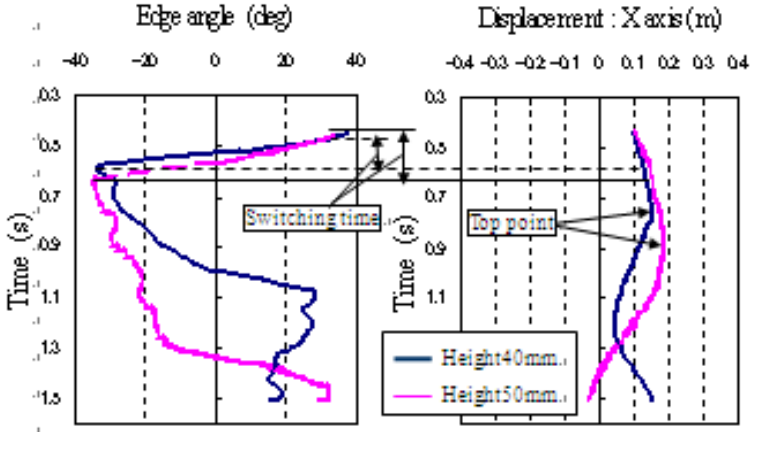


Fig. 7. Edge angle and switching time (Side Curve : R400mm)

Fig. 7 shows the relationship of the edge angle with the ski position over time for cases in which the side curve radius R=400 mm and the waist heights are 40 mm and 50 mm. Considering the time duration in which the turn finishes completely from one direction to another, we recognize that it is shorter and that the turn cycle is smaller when we have a lower waist part.

5 Conclusions

In conclusion, to elucidate the dynamics of ski turning, we produced a passive type of turn robot that was able to make turns continuously, merely by shifting its centre of gravity position. Using it, we examined the possible impact on the turn of different positions of centres of gravity and different side curves. The salient results are the following.

The smaller the side curve is, the smaller the turn radius. The position of the centre of gravity has little relevance.

A higher centre of gravity increases the turn cycle, causing the skis to skid more easily, increasing the turn radius, and lengthening the necessary duration for turning.

Based on these results, for a carving turn in actual Alpine skiing, we verified that the skier would be able to make a quicker turn by reducing the side curve radius R by stepping down on the skis harder during the turn, and by lowering the centre of gravity.

References

1. Kagawa, H., Yoneyama, T., Okamoto, A., Komatsu, H.: Development of a Measuring System for Joint Angles of a Skier and Applied Forces during Skiing. JSME International Journal Series C 41(2), 214–219 (1998)
2. Sakata, T., Tsukiyama, M.: Effects of Position of Shoe Center on Ski Turn. JSME International Journal Series C 42(4), 922–929 (1999)
3. Kawai, S., Otani, H., Sakata, T.: Coupled Motion of Ski and Elastic Foundation Under Ski Control. JSME International Journal Series C 46(2), 614–621 (2003)
4. Kawai, S., Yamaguchi, K., Sakata, T.: Ski Control Model for Parallel Turn Using Multibody System. JSME International Journal Series C 47(4), 1095–1100 (2004)
5. Sahashi, T., Ichino, S.: Coefficient of Kinetic Friction of Snow Skis during Turning Descents. Japanese Journal of Applied Physics, part 1 37(2), 720–727 (1998)
6. Hosokawa, K., Kawai, S., Sakata, T.: Improvement of damping property of skis. Sports Engineering, 107–112 (2002)
7. Sakata, T., Ito, T.: Simulation of Ski Turn. In: Proceedings of the 2nd International Conference the Engineering of Sports, pp. 361–368 (1998)
8. Morawski, J.M.: Control System Approach to a Ski-turn Analysis. J. Biomechanics 6, 267–279 (1973)
9. Tada, N., Hirano, Y.: In Search of the Mechanics of a Turnig Alpine Ski Using Snow Cutting Force Measurements. Sports Engineering 5(1), 15–22 (2002)
10. Sakata, T., Nakashima, K.: A Simulation on Ski Turn (Effects of Skier's Edging Action). In: Proceedings of D & D 1996 Conference (in Japanese), vol. (96-5), pp. 481–484 (1996)

Design and Application of High-Sensitivity Hexapod Robot

P.S. Pa* and C.M. Wu

Department of Digital Content Design, Graduate School of Toy and Game Design
National Taipei University of Education
No.134, Sec. 2, Heping E. Rd., Taipei City 106, Taiwan
myhow@seed.net.tw

Abstract. In recent years, with the advancement of electronic and control technologies, robots are being designed not only to perform dangerous or automated tasks, but also to serve in other fields such as education, entertainment, cleaning, security, tour guiding, and environmental exploration. Among the various types of robots, walking robots are less stable than wheeled ones. Moreover, it is also known that controllers, when required to multiplex and generate PWM signals for controlling servo actuation, may fail to handle multi-axis control and other external tasks simultaneously. Therefore, robots intended for both sensing and communication purposes are typically designed as wheeled robots rather than walking robots. This study aims to develop a low-cost walking robot that is capable of exploring the environment in a walking manner and transmitting environmental information to the computer end through a Bluetooth module. In this study, a hexapod robot is designed as a test carrier and is integrated with a single chip and a variety of sensing devices for environmental detection. The single chip is coupled with a CPLD, which controls the actuation of servos and thereby enables locomotion of the hexapod robot. More particularly, the single chip is coupled with ultrasonic sensors, infrared sensors, a biaxial accelerometer, an electronic compass, a temperature sensor, an infrared human-body sensor, and a Bluetooth module so as to realize a moving device capable of walking and high-sensitivity sensing. A digital man-machine interface is also designed in this study for transmitting information sensed by the single chip to the computer end, thus allowing a user to apply the information as needed.

Keywords: Robots, Servo, CPLD, PWM, MCU.

1 Introduction

The locomotion of a walking robot composed of a plurality of servos is characterized by the number of legs and is therefore identified as biped, tripod, quadruped, hexapod, or multi-legged. Due to the plurality of servos, a walking robot has more degrees of freedom and is more adaptive to the surroundings than a wheeled robot. Now that the existing environment is not suitable for wheeled robots to move around, it is the walking robots that will have the most effective use in human society in the long run.

* Corresponding author.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 1051–1064, 2009.

Furthermore, while a wheeled robot advances without any anticipatory actions, a walking robot is designed to move in imitation of living creatures. Therefore, the movements of a walking robot can be anticipated according to our own (human) experience and are less likely to evoke a sense of threat to us [1]. Hexapod and multi-legged robots are mechanical walking devices having six and more than six legs, respectively. These robots exhibit high stability, can move on bumpy ground, and are therefore more studied. The very first walking robot equipped with a complex control system is the "Hexapod", which was designed by McGhee et al. of Ohio State University in 1978 and further developed by Waldron and Song in 1984 into the hexapod Adaptive Suspension Vehicle (ASV), whose assembly was completed in 1985. One year later, after the installation of control software in 1986, the first hexapod robot with a complex control system was born. In 1988, Umesh analyzed and designed an isotropic leg mechanism for ASV, wherein the items for analysis include degree of freedom, position, velocity, and acceleration. Although the resulting robot was not so agile as those developed by others, the relatively inexpensive mechanical components and the simple control device used in Umesh's robot started a trend for low-cost robots [2].

Nowadays, several research institutions are dedicated to the development of walking robots, and it is believed that significant advances will some day be achieved that allow walking robots to work not only for but with humans. The research of the biped robot combines different disciplines such as mechanism, mechanics of machinery, electronic engineering, control engineering, biological engineering, and robotics. The major research content includes: the design of the leg mechanism, gait planning, walking track, and balance & control theories. In 1977, Gollidary & Hemani adopted Langrangian's dynamics theory to deduce and linearize the mathematic model of the kneeless biped robot to analyze its stability, manipulability, and observability. In 1980, Miyazaki & Arimoto applied the Singular Perturbation method to categorize the dynamic behaviors of the biped robot into the fast mode and slow mode, and designed the controller based on the method. In 1986, Railbert published his book The Balance of the Robot with feet , which is regarded as the most pioneering and contributive research for all oil-pressure-driven robot with one, two, or four feet [3-4]. The robot has no body but with a high pelvis that connects the arms. It is a biped robot with "loaded" 12 DOF, including: 3 DOF in the hip, 1 DOF in the knee, and 2 DOF in the ankle. There are 5 linkages with 4 DOF from the front view, and 7 linkages with 6 DOF from the side view. The weight of two feet is only around 5 ~ 10% of the total weight. If the load of the arms has little impact on the balance, and the proportion of the weight of the two arms to the total weight is small, the robot will be able to up and down the stairs, walk on a slope with less than 10% degree, turn, and step back [5]. Grishin et al. (1994) designed a biped robot with 2 degrees of freedom (DOF) of full load. The robot is composed of a pelvis and two stretched feet. Its mechanic system has 4 DOF, 2 spinning and 2 moving, and minus 2 restraints. The total length of the two feet is a constant. The pelvisis kept at the center between two feet. To prevent from toppling over, the leg of the robothas been installed a feet vertical to the ground. From the side view, it liked a 3 linkage movement. Therefore, the biped robot is able to walk straight on the floor. The first generation of the Japanese Honda robot was developed by Honda R & D Center (Hirai, 1997) [6]. A walking track for robots to follow is necessary before successfully driving the robot to walk or move something from one place to another place. There are several ways to generate the walking track.

One is through observing the qait of real human [7] another by instant calculation. In 1970, Vukobratovic et al. adopted the numerical methods to calculate the dynamic walking track of the biped robot. Kato applied the same method to generate the dynamic walking track for his biped robot, WL-10RD [8-9].

However, it takes a long time to calculate the reference track when the robot is walking, and it is hard to adjust to different surface. Unless the CPU can calculate faster or the algorithm is simplized, the numerical method still has the problem of slow calculation. Other methods to generate the walking track include: inputting the least power [10], using actuator of the neural network and using the genetic algorithm [11]. The DOF of the robot's arm depends on the type of robot and flexibility like human arms can be achieved. There are 6 DOF in human arm-and-wrist mition. The upper arm around the shoulder juncture has two DOF for rotating and swinging. The elbow juncture has the 3rd DOF. The 4th DOF is for spinning the wrist. The 5th DOF is for the up and down movement of the wrist. The right and left movement of the wrist is the 6th DOF. Operations of the fingers and thumb are other additional DOF. The robot arm kinematics is about the movement operation of the robot arm during a period of time relative to a fixed coordinate system. In traditional analysis, the robot's base is regarded as a reference point, and other movements must be based on such reference point. When we know about all positions of junctures and linkage on the robot's arm, then we can calculate the exact position of the arm's terminal in the space. This is a question about the direct kinematics. The reversed question is to decide the target position of the juncture and the linkage to enable the robot's arm to move to the specific positions and directions in the space. The reversed question is more difficult and may have more than one answer [12]. Modern commercial robots equip with the hybrid spin and slip junctures, connected by the arm linkage or the wrist parts of the robot's arm. The spin juncture controls the movement of two linkages in a specific degree, and the slip juncture controls only the line movement between the two linkages. Theoretically, other juncture relationship is possible. However, in reality, only the two junctures are adopted. The serial combination of the juncture and the linkage is called a chain. The chain is open or closed. Every linkage end of the chain only connects to a juncture point. Open chain refers to a chain that does not connect to linkages that are closer to the base. A chain that connects a linkage to the prior juncture point is called the closed chain. The main type in modern industrial robots is the open chain [13]. The analysis and control of the robot's arm requires the development of the analysis and control method. An arm with multiple juncture points is impacted by many force interactions and external environment, and it requires a more complex analysis. Researches of Paul about the homogenous transformation matrix method and the coordinate transformation in the analysis of the robot's kinematics are useful references [14].

As an embodiment of modern science and technology, a walking robot involves such disciplines as electronics, electrical engineering, information technology, artificial intelligence, control engineering, mechanics, robotic kinematics, and so on. However, despite considerable progress in the past few decades in robotic research and the gradually maturing conditions for the development of robots, it is still extremely difficult to control the locomotion of a walking robot, which is not supported by a stationary base and requires highly non-linear control. In order to achieve higher stability, it is necessary to use a more complicated control system in conjunction with the theories of kinematics

and various sensors, so as to precisely control the locomotion of a walking robot. As there are presently a good number of research institutions dedicated to the development of walking robots, it is believed that significant advances will be made in the near future that allow walking robots to work not only for but also with humans.

2 Objectives of Study

It is an objective of this study to design a hexapod robot capable of walking as well as performing multiple sensing and communication functions. In addition, the hexapod robot is provided with a computer-based man-machine interface for transmitting various sensing results to a computer at the user end. In this study, a multi-axis servo control device based on a CPLD is designed so as to relieve the workload of the main control chip and thus allow the main control chip to carry out other tasks such as communication and sensing. Using the CPLD for servo control achieves effective division of labor for controlling the robot. Hence, not only the requirements but also the costs for research and development of the robot are lowered.

Servo robots have a lot of potential, and it is highly desirable for future robots to have various sensing abilities and intelligent functions that further broaden the applicability of robots. In this study, a walking servo robot is coupled with different types of sensors and configured to perform a variety of pre-programmed actions, thus helping humans with certain vexing, boring, and dangerous tasks. Also, the integration of wireless transmission with the man-machine interface of this study contributes to accelerating the development of, and exploring new applications for, robots.

3 Experimental Setup

1. Control structure

The control structure proposed in this study includes a CPLD as the controller of servomotors and a single chip designed to work with the CPLD. The single chip is in charge of external communication with a computer, a plurality of sensors, a memory, and a biaxial accelerometer, such that the entire system can be effectively controlled even when different tasks are executed at the same time. Referring to Fig. 1 for a structural diagram of the control system of this study, an MCU serves to communicate with the CPLD and external devices while the CPLD controls the motion of servomotors. In this study, a dedicated servo interface device is designed with VHDL (Very High Speed Integrated Circuit Hardware Description Language) so as to provide synchronous control over the 32-axis servos. Fig. 2 is an analysis diagram of the 32-axis PWM pulse signals generated in this study. As shown in Fig. 2, the PWM signals for all axes are synchronous and have a fixed period of 20ms.

2. Design of Tilt Sensor

The hexapod robot of this study uses Memsic MX-2125 for sensing the tilt of the test carrier. The sensor outputs a pulse signal (in PWM format). According to the drawing, the output period of T2 is fixed at 10ms at 25°C, and the output period of T1 is 5ms when horizontal. As the Memsic MX-2125 sensor detects a tilt, the width of T1 changes accordingly, but the width of T2 remains. Therefore, the tilt angles of the two axes can be obtained by measuring the width of T1 signal.

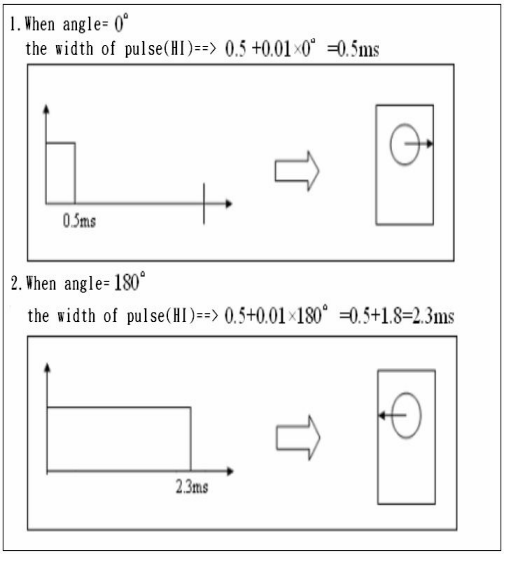


Fig. 1. Structural diagram of control system

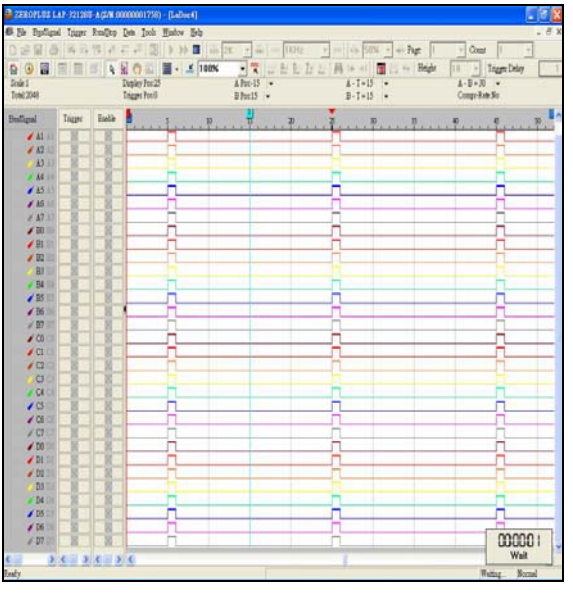


Fig. 2. Analysis diagram of PWM signals

3. Integration of Servos and Components

The servomotors for use in this study are GWS S03T manufactured by GWS of Taiwan. The torsion of the motor is as high as 7.2kg/cm when the operating voltage is 4.8V, and 8kg/cm when the operating voltage is 6V. On the other hand, the rotation

speed of the motor reaches 0.33sec / 60deg when the operating voltage is 4.8V, and 0.27sec / 60deg when the operating voltage is 6V. The test carrier, i.e., the hexapod robot, of this study is made essentially of 12 servomotors, a specially designed aluminum plate, and several screws and copper posts, as shown in Fig. 3.

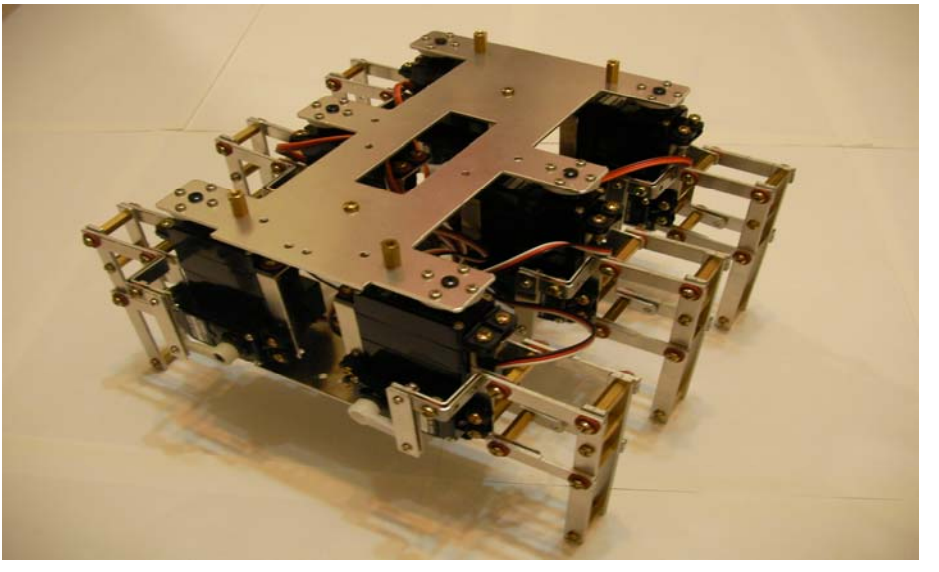


Fig. 3. Physical embodiment of the robotic mechanism

4 Results and Discussion

3-1 Principle of Motion and Gait Design of the Hexapod Robot
The hexapod robot of this study uses 12 servomotors and is enabled to walk by power output from the servos in conjunction with connecting rods in the robotic mechanism. While the focus of this research is on electric control, the hexapod robot of this study must have the ability to move forward. Hence, we gathered and analyzed information on gaits of existing hexapod robots and obtained a relatively simple gait sequence, as shown in Fig. 4, which is the sequence used in this study to simplify the gaits of the designed hexapod robot.

6. Electromechanical Integration of the Hexapod Robot
After completing the design and manufacture of the required mechanisms and circuits, all the components, modules, and motors are put together to complete the hexapod robot of this study, as shown in Fig. 5. The hexapod robot is capable of sensing obstacles (via the ultrasonic and infrared obstacle sensing circuits), temperature, human body (via the infrared sensors), tilt (via the biaxial accelerometer), and

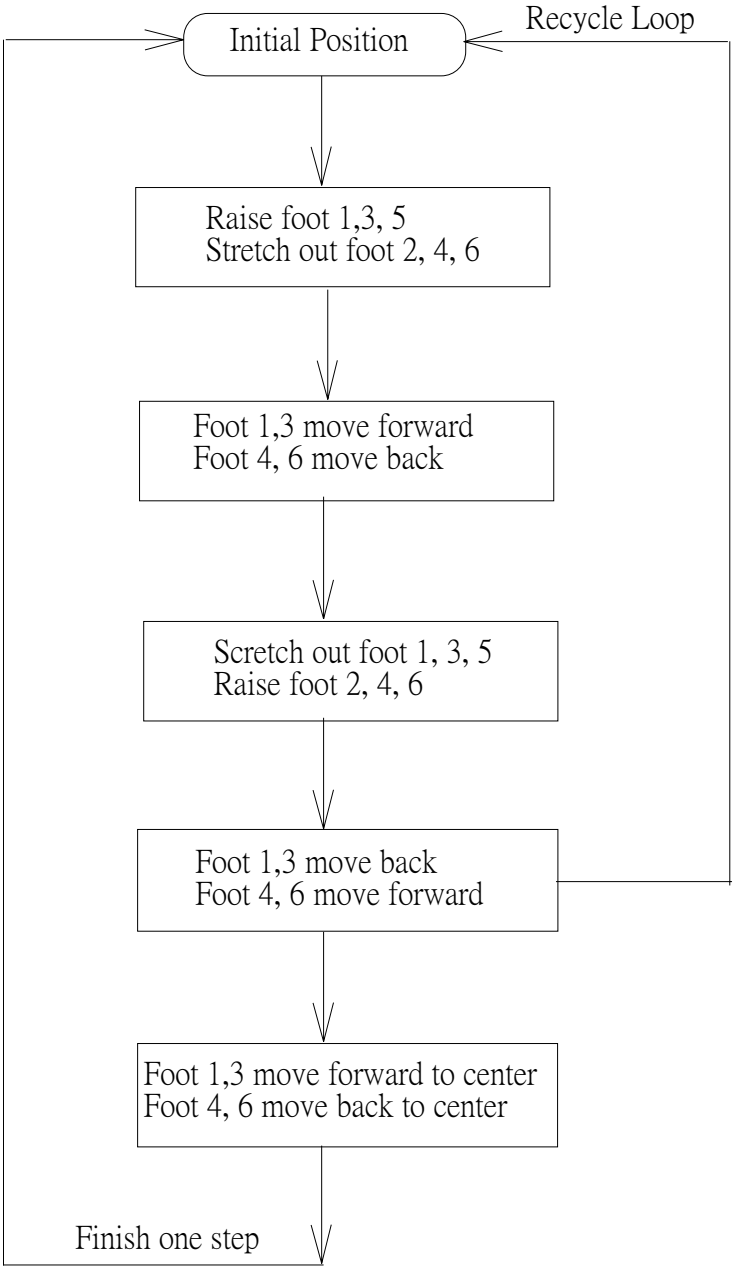


Fig. 4. Gait sequence of the hexapod robot

azimuth (via the electronic compass), as well as wireless remote control (via a 433MHz wireless receiver module). Additionally, a wireless Bluetooth module transmits all the sensing results to the man-machine interface at the computer end to facilitate information integration. The man-machine interface of this study also enables the plotting of simple topographic maps.

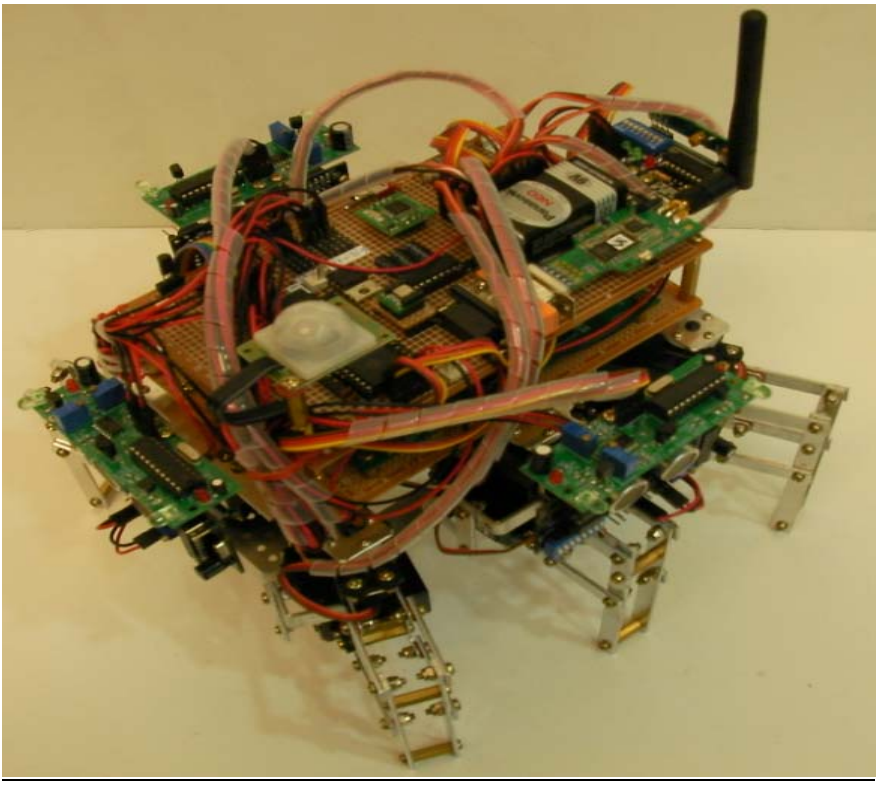


Fig. 5. The completed hexapod robot

Fig. 6 and Fig. 7 illustrate the relative positions of the sensors and the CPLD. On the upper circuit board, the infrared human-body sensing module is at the lower right, the electronic compass is at the left, and the biaxial accelerometer is at the center. In addition, the 433MHz RF wireless receiver module is coupled to the upper circuit board from above, and each of the front, left, and right ends of the upper circuit board is provided with an ultrasonic sensor and an infrared sensor. Moreover, a DS-1821 temperature sensor is at the lower left of the upper circuit board while the wireless Bluetooth module is at the upper right. The lower circuit board is the CPLD of this study. The CPLD is coupled with the controlled servos (test carrier) after the aforementioned sensors are connected to the single chip, thus completing the design of this study.

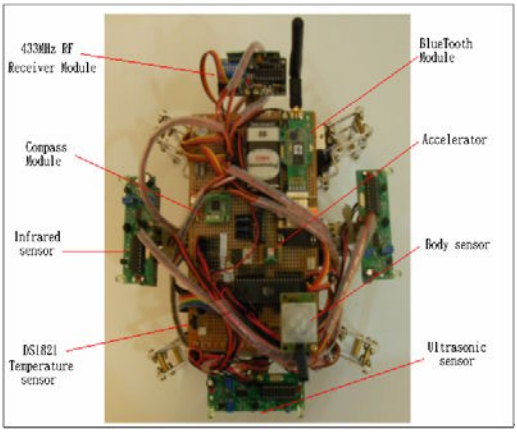


Fig. 6. Sensors of the hexapod robot

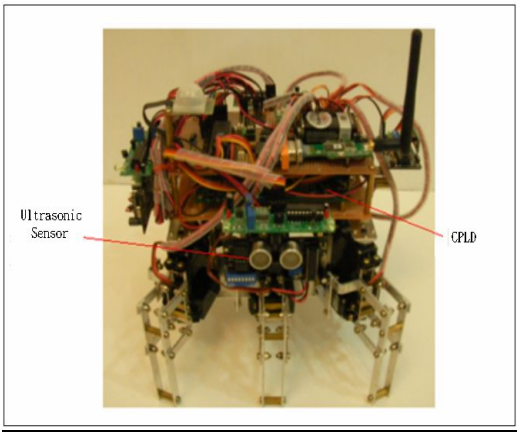


Fig. 7. Location of CPLD of the hexapod robot

7. Integration of the Map-Making Man-Machine Interface of the hexapod robot

In this study, the man-machine interface at the computer end and related programs are developed and written with Visual Basic. The programs allow the computer to communicate with the hexapod robot and control the locomotion thereof. On the other hand, the robot transmits the sensing results to the computer via the wireless Bluetooth module. The sensing results to be collected by the computer include the ultrasonic and infrared sensing results, ambient temperature, tilt angles, and data from the electronic compass. These sensing results are then used to plot a virtual planar space. Referring to Fig. 8 for the homepage of the man-machine interface designed in this study, a user is prompted to select the appropriate COM PORT and then press “Start Connection” to enter the function interfaces listed in the menu on the top. The three

major function interfaces are an environment-sensing interface, a map-making interface, and a motion-editing interface.

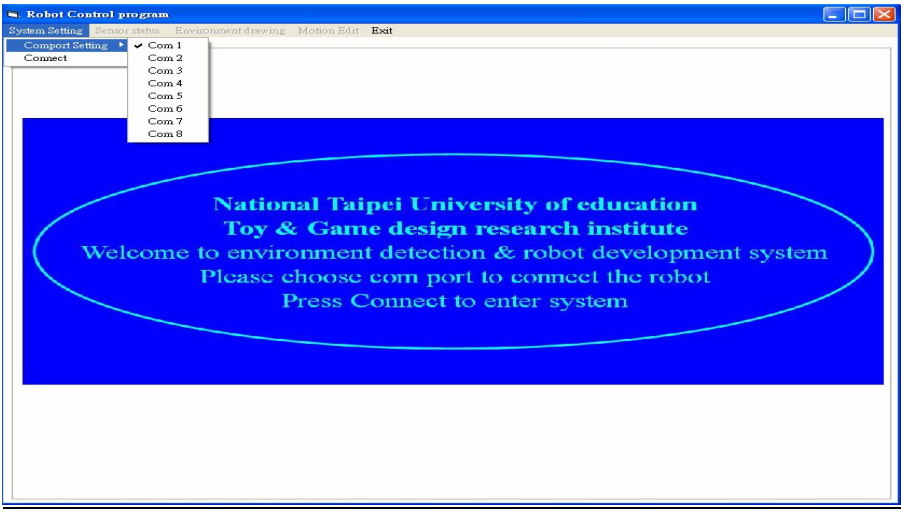


Fig. 8. Homepage of the man-machine interface of the hexapod robot

Fig. 9 illustrates the environment-sensing interface of the hexapod robot. Once this interface is activated, the hexapod robot is connected to the computer and transmits thereto the various information obtained by the sensors. The sensing results are presented by patterns of different colors so as to indicate whether there are obstacles in front of or to the left or right of the hexapod robot. Tilt angles of the hexapod robot in the X- and Y-directions are also shown, in addition to the temperature around the robot. A description of all the sensing statuses is provided on the right-hand side of the man-machine interface.

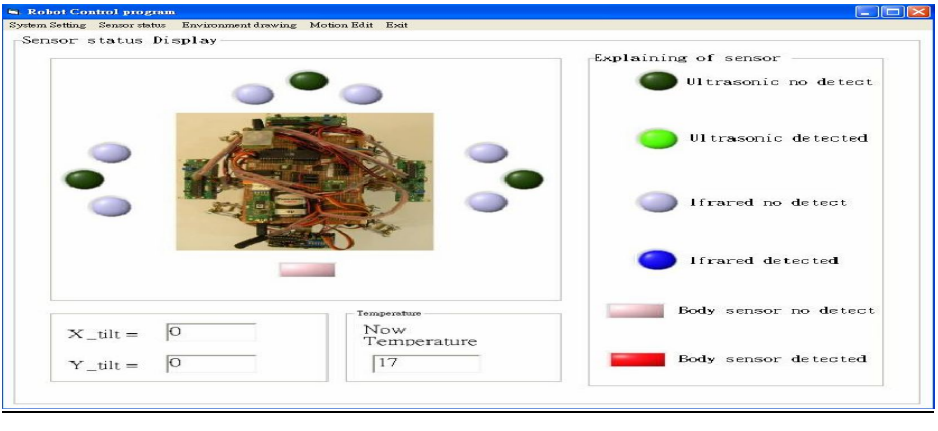


Fig. 9. Environment-sensing interface of the hexapod robot

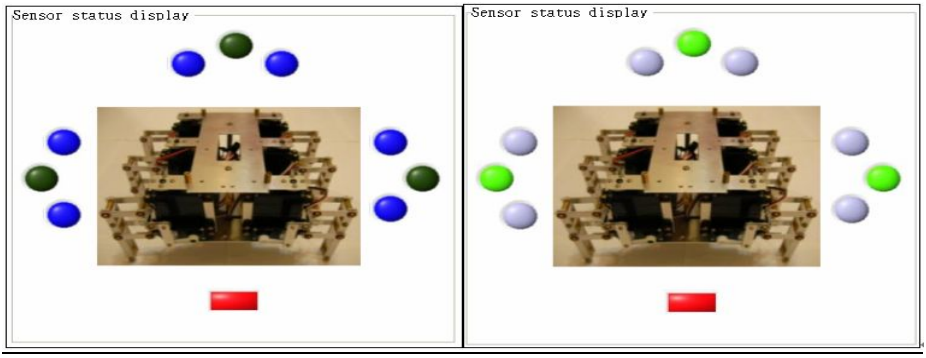


Fig. 10. Ultrasonic and infrared sensing diagrams of the hexapod robot

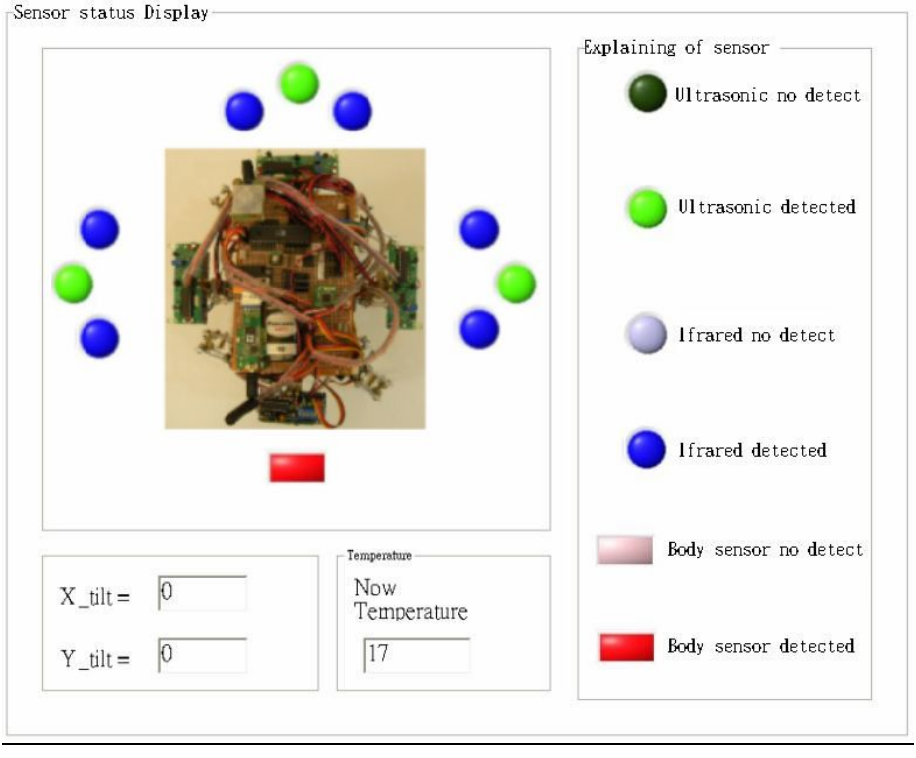


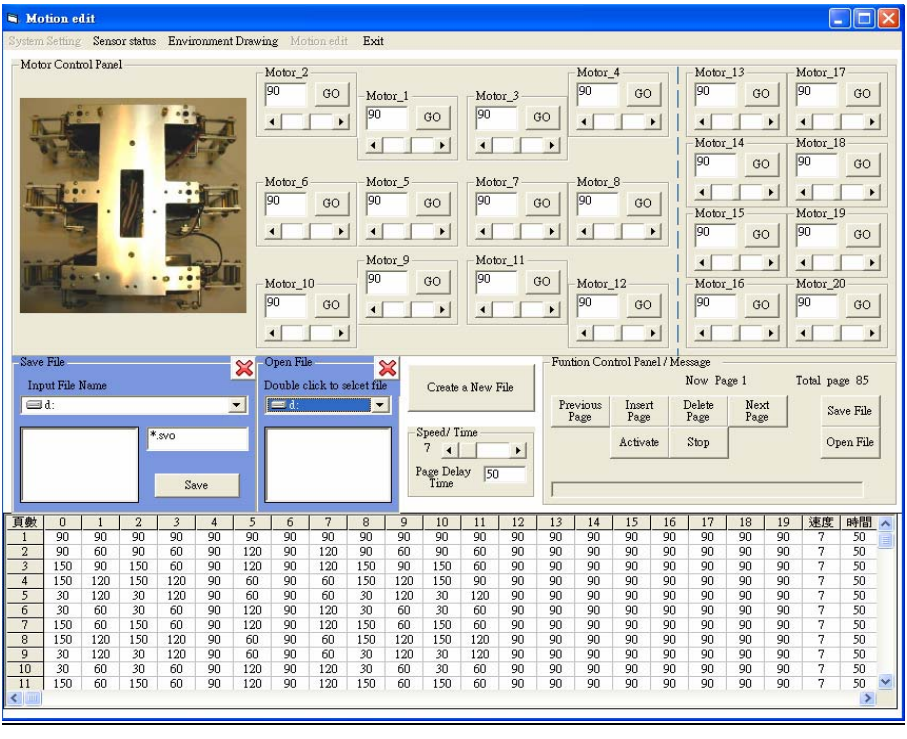
Fig. 11. Environment-sensing statuses of the hexapod robot

Fig. 10 provides ultrasonic and infrared sensing diagrams of the hexapod robot. The diagram on the left shows that all the infrared obstacle sensors at the front, left, and right ends of the hexapod robot have positive detection results. In other words, there are objects in all these three directions. The diagram on the right shows that the

ultrasonic obstacle sensors at the front, left, and right ends of the hexapod robot also have positive detection results, indicating the existence of objects in the three directions. Besides, the infrared human-body sensing result is shown at the lower part of each diagram with a red color, meaning that a human body is detected.

The patterns in Fig. 11 indicate that all the sensors have positive detection results. More specifically, the bright green color indicates the positive detection results of the ultrasonic sensors, the blue color indicates the positive detection results of the infrared sensors, and the red color indicates the positive detection result of the infrared human-body sensor. In addition, ambient temperature and tilt angles of the hexapod robot are shown at the lower part of the diagram.

Referring to Fig. 12 for the motion-editing interface of the hexapod robot, a photograph of the test carrier, i.e., the hexapod robot, designed in this study is displayed at the left of the interface. As the photograph corresponds to control items of the central motor, the interface serves to control the same motor. A user may input the desired moving angle of the motor or directly pull the bar below the motor to actuate the corresponding motor of the hexapod robot in real time. The simple operation interface not only provides a user-friendly motion-editing environment, but also stores data of the edited motion so as to build a motion database of the hexapod robot for further editing.



頁數	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	速度	時間
1	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	7	50
2	90	60	90	60	90	120	90	120	90	60	90	60	90	90	90	90	90	90	90	90	7	50
3	150	90	150	60	90	120	90	120	150	90	150	60	90	90	90	90	90	90	90	90	7	50
4	150	120	150	120	90	60	90	60	150	120	150	90	90	90	90	90	90	90	90	90	7	50
5	30	120	30	120	90	60	90	60	30	120	30	120	90	90	90	90	90	90	90	90	7	50
6	30	60	30	60	90	120	90	120	30	60	30	60	90	90	90	90	90	90	90	90	7	50
7	150	60	150	60	90	120	90	120	150	60	150	60	90	90	90	90	90	90	90	90	7	50
8	150	120	150	120	90	60	90	60	150	120	150	120	90	90	90	90	90	90	90	90	7	50
9	30	120	30	120	90	60	90	60	30	120	30	120	90	90	90	90	90	90	90	90	7	50
10	30	60	30	60	90	120	90	120	30	60	30	60	90	90	90	90	90	90	90	90	7	50
11	150	60	150	60	90	120	90	120	150	60	150	60	90	90	90	90	90	90	90	90	7	50

Fig. 12. Motion-editing interface of the hexapod robot

5 Conclusions

A servomotor control interface based on a CPLD is designed in this study. The device, which is directly integrated with a single chip, can precisely generate the desired pulse width for PWM and thereby control the servos of the hexapod robot of this study. A man-machine interface program at the computer end is also designed in this study for detecting the environment, plotting simple topographic maps, and performing servo motion control. The control method proposed in this study significantly reduces the workload of an MCU and thus allows the MCU to communicate with external devices. In addition, the control method of this study enhances synchronism between servos and is hopefully instrumental in further studies and relevant verifications. Commercially available robots capable of making topographic maps and avoiding obstacles are mostly wheeled robots, which fail to function properly on non-planar topography. This study, however, uses a hexapod robot for environmental exploration. The hexapod robot not only adapts to non-planar topography, but also detects and transmits the sloping conditions thereof to a computer so as to store the detected data and plot simple topographic maps accordingly. Nevertheless, the experiments of this study were conducted on specific, but not all, topographic conditions. For example, tests were not carried out on bumpy ground. Therefore, this study may be further extended by performing tests on various ground conditions.

Besides, the servomotor conversion device of this study simplifies servo control and reduces the power otherwise required by the single chip for servo control, thus allowing the single chip to work with more external sensors and accomplish more communication tasks. While the sensors employed in this study include ultrasonic sensors, infrared sensors, a biaxial accelerometer, and an electronic compass, it is desirable for subsequent studies to incorporate more sensing functions such as the measurement of humidity and altitude, or to integrate with mechanical sighting functions. It is hoped that this study will find applications in the exploration of extra-terrestrial planets or unknown environments and make considerable contributions to mankind.

References

1. Shih, C.L., Chi, C.T., Lee, W.Y.: Walking Experiment of a Biped Robot with Individual DC Motor Current and Position Control. In: The Fourth International Power Electronics Conference, Tokyo, Japan, pp. 1249–1254 (2000)
2. Jones, B.A.: Practical Kinematics for Real-Time Implementation of Continuum Robots. IEEE Transactions on Robotics 22(6), 1087–1099 (2006)
3. Golliday, C.L., Hemami, H.: An Approach Analyzing Biped Locomotion Dynamics and Designing Robot Locomotion Control. IEEE Transactions on Automation Control 22(6), 963–972 (1997)
4. Miyazaki, F., Arimoto, S.A.: Control Theoretic Study on Dynamical Biped Locomotion. ASME Journal of Dynamic Systems Measurement and Control 102, 233–239 (1980)
5. Sadain, P., Rostami, M., Thomas, E., Bessonnet, G.: Biped Robots: Correlations between Technological Design and Dynamic Behavior. Control Engineering Practice 7, 401–411 (1999)

6. Ambarish, G.: Postural Stability of Biped Robots and the Foot-Rotation Indicator (FRI) Point. The International Journal of Robotics Research 18, 523–533 (1999)
7. Hira, K.: Current and Future Perspective of Honda Humanoid Robot. In: Proceedings of 1997 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 500–508 (1997)
8. Frank, A.A.: An Approach to the Dynamic Analysis and Synthesis of Biped Locomotion Machines. Med. Biol. Eng. 8, 465–476 (1970)
9. McGeer, T.: Passive Walking with Knees. IEEE Intgernational Conference on Robotics and Automation 3, 1640–1645 (1998)
10. Shih, G.L., Zhu, Y., Gruver, W.A.: Optimization of the biped robot trajectory. In: IEEE International Conference on Decision Aiding for Complex Systems, Conference Proceedings, vol. 2, pp. 899–903 (1991)
11. Meifen, C., Kawamura, A.: An Evolutionary Design Scheme of Neural Oscillatory Network for Generation of Biped Walking Patterns. In: 5th International Workshop on Advanced Motion Control, pp. 666–671 (1998)
12. Denavit, J., Hartenberg, R.S., Kinematic, A.: Notation for Lower-Pair Mechanisms Based on Matrices. ASME Journal of Applied Mechanics, 215–221 (1995)
13. Shih, C.L., Chi, C.T., Lee, Y.W.: Walking Experiment of a Biped Robot with Individual DC Motor Current and Position Control. In: The Fourth International Power Electronics Conference, Tokyo, Japan, pp. 1249–1254 (2000)
14. Tai, J.C., Liao, H.T., Chen, C.T.: Electric and Mechanical Integration, vol. 307. Kao Lee Books, Taipei (2003)
15. Jones, B.A., Walker, I.D.: Practical Kinematics for Real-Time Implementation of Continuum Robots. IEEE Transactions on Robotics 22(6), 1087–1099 (2006)

Study on Mine Rescue Robot System

Juan Wei and Hong-wei Ma

College of Mechanical Engineering Xi'an University of Science and Technology
Xi'an, Shaanxi Province, China
weij@xust.edu.cn, mahw@xust.edu.cn

Abstract. Mine rescue robot system consists of the carrier robot and the exploration robot. After the incident, the rescue robot system can instead of rescuers to enter the mine, collect environmental information of the scene of accident. This paper introduces the structure and components of the carrier robot and the exploration robot, and analyse separately their kinematics, find out the motion control parameters. Accordance rescue robot with the workplace, put forward the autonomous navigation method based on coal mine geographic information system, and finally to a simple example of the path searching process.

Keywords: Rescue robot, kinematics analysis, autonomous navigation.

1 Introduction

In recent years, coal mine accidents have frequently happened in the world, the level of mine rescue is relatively lagged, and varieties of the rescue works are still mainly manual. Many units launched a variety of coal mine accident rescue robots, and some of them have developed prototypes of mine rescue robots. Xi'an Science and Technology University is one of them. In 2006, Xi'an Science and Technology University has launched the based research on mine rescue robot autonomous navigation and intelligent control in the National Natural Science Foundation's support, now have developed the mine rescue robot system, which can enter the mine tunnels instead of human after a explosion of methane or dust etc happened in coal mines, to obtain the concentrations of methane, carbon monoxide, oxygen and other gas and temperature, humidity and other environmental information of the scene of accident. Using the information collected in the scene of accident from robot systems, the first-hand data of the accident scene can be provide to ground rescuers timely and accurate. It can provide reliable information for effectively mine for emergency relief, to reduce casualties and property losses, but also for the scientific analysis of gas explosion during the fire disaster of the spread of mixed-gas law, under the conditions of a multi-hazard disaster coupled into the incident and expansion of the conditions and factors to provide the raw data of the study.

2 The Mechanical Subsystem

According to the status that mine accidents usually occur in the working face away from the mine entrance. The coal mine rescue robot system adopts two-stage robot

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 1065–1072, 2009.

structure. It is shown in fig.1. The first-stage robot is a carrier robot that adopts the wheel-type mobile structure. The second-stage robot is exploration robot that adopts a complex-crawler robot. After a mine accident, the first-stage robot carries the second-stage robot along the underground transport orbit moving fast to the nearest place where the accident happened. If the first-stage robot is obstruction by obstacles or the orbit is damaged. It can not continue to move forward, then open the gate, and the second-stage robot leaves the first-stage robot from the appropriate position, moves on with the features of itself navigation, autonomous mobile etc, till near the scene of accident, and collects the environmental information of the underground.

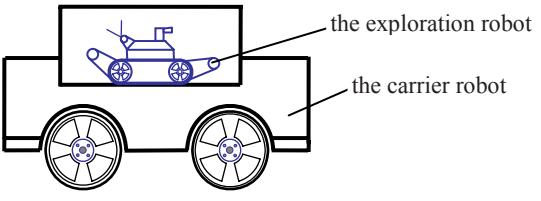


Fig. 1. The rescue robot system

2.1 The Structure Design of the Mechanical Subsystem

The carrier robot mechanical structure is made up of running gear 1, removing obstacle device 2, tank 3, sensors 4 and landing gear components 5 etc, as shown in fig.2. The running gear is composed of four wheels, and is driven by a explosion-proof motor, and its bound to run the track by orbit. There is only one degree of freedom for the one-dimensional space. The removing obstacle device can clear some obstacles on the orbits. The tank is separated into two parts from top to bottom. The lower part is placed for a driven motor, a reducer, a power, control subsystem, drive components etc, and the upper part is placed for the exploration robot and a variety of sensors. The landing gear is mainly used for the exploration robot to leave and back the carrier robot.

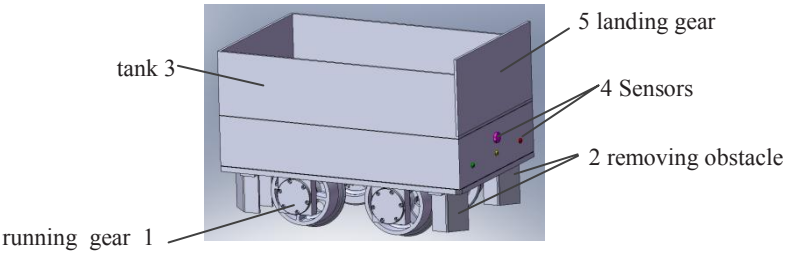


Fig. 2. The carrier robot

By modular design method, the exploration robot was designed a symmetrical modular structure [1]. It is shown in Fig.3, among them, (a) shows the structure of the exploration robot, (b) and (c) are shown its three-dimensional model of difference attitude. The robot includes left and right main track modules 1 and 2, a front-right 3,

a front-left 4, a back-right 5, a back-left 6 six swinging arms track modules and a box 7. The middle box can be load a variety of detection instruments and equipments for completing the mission, which is an exploration robot carriage platform.

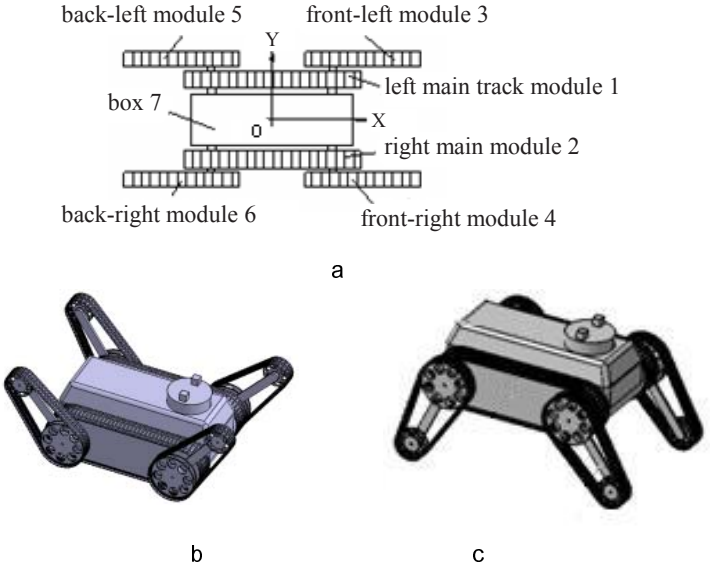


Fig. 3. The exploration robot

Each track module includes a motor, a decelerator, a track and some assist devices. Each main track is independently driven by a main motor, a reducer and a pair of bevel gear installed in the middle box, and is shown in Fig.4. Two main track modules commonly realized translational and rotation of the exploration robot. Each arm track is independently driven by an arm motor, a reducer and a pair of bevel gear to achieve rotation to the 360° around itself axis. Removing the driver-side structure of the main drive motor and reducer drive, encoder, meshing bevel gears, and it is the robot driven end structure [2].

The exploration robot has six degrees of freedom, namely 2-dof translational and 4-dof swing. The two main motors that was fitted in the middle box driver the six tracks come true the translational motion in any direction and rotation around the Z axis, the other four motors within the box respectively drive the four swinging arm tracks, the arm tracks accomplish main act of walk, such as grade climbing, obstacle crossing, fording, stepping upstairs with the main track etc.

2.2 The Kinematic Analysis of the Mechanical Subsystem

The mine rescue robot system is including of the carrier robot and the exploration robot two-stage robot structure. When explosive accidents happen in a mine, the robot system is sent in the mine, they advance in the underground tunnel by its own characteristics. In mine entry of away from the working face, the damage usually is smaller.

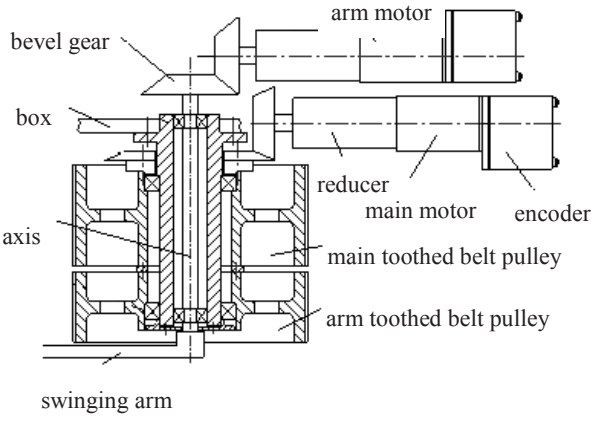


Fig. 4. The schematic diagram of the track module

The carrier robot can be high-speed operation along the orbit, which carriage the exploration robot, and clear the smaller obstacles on the orbit using its removing obstacle device. So on, until the orbit is damaged, or is obstructed from greater obstacles, or the track direction can not fill demand. The carrier robot will stop and expand the landing gear, and the exploration robot away the carrier robot from the landing gear.

The exploration robot is crawler robot, with the exception of the ground with walk, turn any angle, but also in the common of the body and the arms to achieve climbing stairs, obstacle crossing, fording, crossing groove trenches, stand by itself and other functions.

1. The kinematic analysis of the carrier robot

The carrier robot bears the task that long-rang fast-moving and carry the exploration robot. Sine the carrier robot can only move along orbits, its move is hold in by the orbit and the motor. If the motor speed is $n(t)$, and the exterior size of the carrier robot is $a \times b$, the distance between left and right wheel is l, the wheel radius is r. The angle between the orbit and the horizontal axis is $\theta(i)$, and between the vertical plane (Z axis) angle is $\beta(i)$, the position and the posture of the carrier robot in the mine coordinate system is:

$$P_I = [x \ y \ z \ \theta]^T = f(l, r, \theta, \beta, t). \tag{1}$$

discretization is:

$$\hat{p}_k = \begin{bmatrix} x \\ y \\ z \\ \theta \end{bmatrix} = p_{k+1} + \Delta p = \begin{bmatrix} x_{k-1} \\ y_{k-1} \\ z_{k-1} \\ \theta_{k-1} \end{bmatrix} + \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \\ \Delta\theta \end{bmatrix}. \tag{2}$$

If the mapping is from the mine coordinate to the robot coordinate system:

$$R(\theta) = \begin{bmatrix} cos\,\theta & sin\,\theta & 0 & 0 \\ -sin\,\theta & cos\,\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{3}$$

There is $P_R = R(\theta)P_I$, where P_R is the position and the posture of the robot in the robot coordinate system, P_I is the position and the posture of the robot in the mine coordinate system. The forward kinematic model of the carrier robot is:

$$\dot{P}_R = R(\theta)\begin{bmatrix} \dot{x} & \dot{y} & \dot{z} & \dot{\theta} \end{bmatrix}^T = R(\theta)\begin{bmatrix} \frac{2\pi rn}{i} & 0 & 0 & \dot{\theta} \end{bmatrix}. \tag{4}$$

where θ is hold in by the orbit.

2. The kinematic analysis of the exploration robot

The exploration robot adopt crawler configuration. Its body is driven by around the two main motors. Although there are four pulleys, but only two main motors drive, so the Kinematic analysis is the same as the two differential driving wheels which has two additional contact points [3].

The two differential driving wheels curved line of the exploration robot as Y_R-axis of the robot coordinate. The line that through the center of the robot perpendicular to Y_R-axis as X_R-axis of the robot coordinate, and let the point R of the robot as the reference point for the robot position, as shown in Fig.5. The position and the posture of the exploration robot in the mine coordinate system may be expressed as:

$$P_I = (x \quad y \quad \theta)^T. \tag{5}$$

If the drive wheel radius of the robot is the equal of *r*, the distance between the two wheels is *l*, the speed of two driving wheels are n_1 and n_2. According to the principle of differential velocity, the linear velocity v_R and the angular velocity ω_R in the robot coordinate system as follows:

$$\begin{cases} \mathrm{v_R} = \pi \mathrm{r} \mathrm{n}_2 + \pi \mathrm{r} \mathrm{n}_1 \\ \omega_\mathrm{R} = \frac{2\pi \mathrm{r} \mathrm{n}_2}{\mathrm{l}} - \frac{2\pi \mathrm{r} \mathrm{n}_1}{\mathrm{l}}. \end{cases} \tag{6}$$

According to the principle of rigid body translational, the robot movement can be seen as rotation that the body around instantaneous center *P*, the rotation radius *R* is:

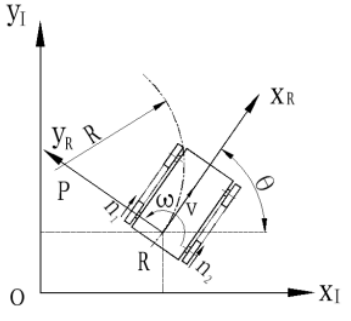


Fig. 5. The mine coordinate system and the robot coordinate system

$$R = \frac{v}{\omega} = \frac{l(n_2 - n_1)}{2(n_2 - n_1)}. \tag{7}$$

its ranges is : [0 ∞ .

So the robot equations of motion can be expressed as [4]:

$$V = \begin{bmatrix} \dot{x}_I \\ \dot{y}_I \\ \dot{\theta}_I \end{bmatrix} = \begin{bmatrix} cos\,\theta & 0 \\ sin\,\theta & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} v \\ \omega \end{bmatrix} = \begin{bmatrix} cos\,\theta & 0 \\ sin\,\theta & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \pi r & \pi r \\ \frac{2\pi r}{l} & -\frac{2\pi r}{l} \end{bmatrix}$$
$$= \begin{bmatrix} \pi r\,cos\,\theta & \pi r\,cos\,\theta \\ \pi r\,sin\,\theta & \pi r\,sin\,\theta \\ \frac{2\pi r}{l} & -\frac{2\pi r}{l} \end{bmatrix} \begin{bmatrix} n_1 \\ n_2 \end{bmatrix}. \tag{8}$$

It can be shown, the differential-driven robot system is the overall control system, via controlling the two drive wheels speed n_1 and n_2, can indirectly control the robot linear velocity v and the angular velocity ω, and achieved the motion of the robot in at any position and attitude. At the same time, the differential drive robot system is non-linear time-invariant systems too. The constraint conditions is $\dot{x}\sin\theta - \dot{y}\cos\theta = 0$, so assumptions in the analysis to it is pure rolling between the main track of the exploration robot and the ground, and no lateral sliding.

3 The Navigation Way of the Mine Rescue Robot System

Navigation of mobile robot is often used to describe the environment and self-perception via sensors, and realize the voluntary movement of their own goal-oriented in obstacle environment. Navigation includes environmental modeling, self-localization and path planning of the content [5].

Environmental model is the basis for robot navigation. Mine rescue robot works under the ground, and the geographic information of the mine is known, but there are lots of change after occurred accident, in which the original rules tunnels, there may occur the larger the deformation, become impassable and even the original access road will be added a number of obstacles ...

Mine rescue robot maximum speed of not less than 2Km/h. Considering that the damage of the tunnel shape and surface is limited to certain regions by accidents, and these changes can be only consider basis of the original tunnel. Therefore, to ensure that the speed of robot, the mine rescue robot navigation method adopt navigation based on geographic information system (GIS) of the mine.

The navigation based GIS should be completed in the computer before the robot system enter into the mine. Its main ideas is using the graph theory express the mine tunnels, and using the network analysis method seeking the shortest distance and the best path from the wellhead to the scene of accident. The process is shown in Fig.6.

Assuming that there is a simple map of the mine tunnel, and is shown in Fig.7. The vertex set V of the graph is contain the endpoints and intersection points of a tunnels, the edge set E of the graph is contain edge each tunnel, so a coal mine tunnel can be expressed by graph $G = (V, E)$ to indicate, then the adjacency matrix of the graph G can be expressed as: [6]

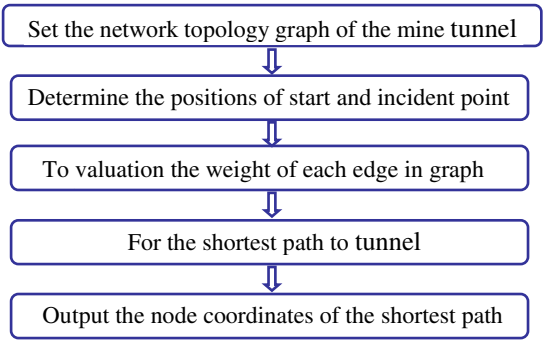


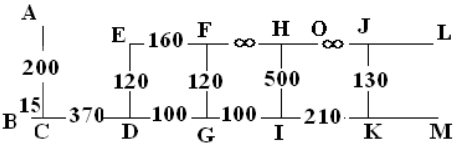
Fig. 6. The process of the navigation based GIS

$$a_{ij} = \begin{cases} x_{ij} & \text{if } v_i v_j \in E(G) \\ \propto, & \text{other} \end{cases} . \tag{9}$$

where x_{ij} is the weighted value considering the length, tilt, currency of tunnel and other factors prevailing factor. The Fig.7 (a) is a schematic diagram of a coal mine tunnel, and (b) is the tunnel graph with weighted value, point A is the entrance and point O is the scene of accident. According the above method, its adjacency matrix with weighted value is:

$$a_{ij} = \begin{Bmatrix} \infty & \infty & 200 & \infty & \infty & \infty & \infty & \infty & \infty & \infty & \infty \\ \infty & \infty & 15 & \infty & \infty & \infty & \infty & \infty & \infty & \infty & \infty \\ 200 & 15 & \infty & 370 & \infty & \infty & \infty & \infty & \infty & \infty & \infty \\ \infty & \infty & 370 & \infty & 120 & \infty & 100 & \infty & \infty & \infty & \infty \\ \infty & \infty & \infty & 120 & \infty & 160 & \infty & \infty & \infty & \infty & \infty \\ \infty & \infty & \infty & \infty & 160 & \infty & 120 & \infty & \infty & \infty & \infty \\ \infty & \infty & \infty & 100 & \infty & 120 & \infty & \infty & 100 & \infty & \infty \\ \infty & \infty & \infty & \infty & \infty & \infty & \infty & \infty & 500 & \infty & \infty \\ \infty & \infty & \infty & \infty & \infty & \infty & 100 & 500 & \infty & \infty & 210 \\ \infty & \infty & \infty & \infty & \infty & \infty & \infty & \infty & \infty & \infty & 130 \\ \infty & \infty & \infty & \infty & \infty & \infty & \infty & \infty & 210 & 130 & \infty \end{Bmatrix} . \tag{10}$$

(a)



(b)

Fig. 7. A schematic diagram of coal mine tunnel

From the entrance A to the other vertex of the shortest path computation process of change in vector D as follows:

$$D_1=\begin{bmatrix}\infty\\ \infty\\ \boxed{200}\\ \infty\\ \infty\\ \infty\\ \infty\\ \infty\\ \infty\\ \infty\end{bmatrix}\quad D_2=\begin{bmatrix}\infty\\ 215\\ \rightarrow\\ \boxed{370}\\ \infty\\ \infty\\ \infty\\ \infty\\ \infty\\ \infty\end{bmatrix}\quad D_3=\begin{bmatrix}\infty\\ \infty\\ \rightarrow\\ \rightarrow\\ \infty\\ 1570\\ \infty\\ \boxed{100}\\ \infty\\ \infty\end{bmatrix}\quad D_4=\begin{bmatrix}\infty\\ \infty\\ \rightarrow\\ \rightarrow\\ \infty\\ 1270\\ \infty\\ \rightarrow\\ \boxed{100}\\ \infty\end{bmatrix}\quad D_5=\begin{bmatrix}\infty\\ \infty\\ \rightarrow\\ \rightarrow\\ \infty\\ 1110\\ \infty\\ \rightarrow\\ \rightarrow\\ \infty\\ \boxed{210}\end{bmatrix}\quad D_6=\begin{bmatrix}\infty\\ \infty\\ \rightarrow\\ \rightarrow\\ \infty\\ \boxed{1110}\\ \infty\\ \rightarrow\\ \rightarrow\\ \infty\\ \rightarrow\end{bmatrix}\quad D_6=\begin{bmatrix}\infty\\ \infty\\ \rightarrow\\ \rightarrow\\ \infty\\ \rightarrow\\ \infty\\ \rightarrow\\ \rightarrow\\ \infty\\ \rightarrow\end{bmatrix}. \tag{11}$$

where the weighted value in the box express the shortest path to the starting point of the A, the position of weighted value express the corresponding vertex. the → express has been found that the vertex of the shortest path. So this can be derived from point A to point H of the shortest path as follows: A-C-D-G-I-K-J, the Path weighted value is 1110.

4 Conclusion

Mine rescue robot using two-stage robot system can effectively improve the efficiency of rescue and shorten rescue time. The autonomous navigation method GIS-based can plan out the best path close to a scene of accident before the rescue robot system go down the mine. Through coordination control the motor speed of the robot, complete the motions of the rescue robot in coal mine tunnel. With the sensors in the robot, the exploration robot can realize obstacle avoidance, obstacle crossing, climbing, fording, crossing groove and such as sports in barrier-free environment. At present, the prototypes of the carrier robot and the exploration robot have developed, and laboratory testing under way.

References

1. Juan, W., Guangli, J., Hongwei, M.: Simulation Analysis of the Virtual Prototyping of a Mine Rescue Robot. Mechanical Science and Technology for Aerospace Engineering 27(11), 1369–1373 (2008)
2. Guangli, J., Juan, W., Hongwei, M.: The Movement Analysis and Configuration Design of a kind of Six Pedrails Detecting Robot. Machine Tool & Hydraulics 36(11), 8–17 (2008)
3. Siegwart, R., Nourbakhsh, I.R.: Introduction to Autonomous Mobile Robots. Xi'an Jiaotong University Press, Shanxi Xi'an (2006)
4. Juan, W.: A Way of Path Control on Differential Drive Robot. Manufacturing Automation (10), 80–82 (2008)
5. Murphy, R.R.: Introduction to AI Robotics. Publishing House of Electronics Industry, Beijing (2004)
6. Chartrand, G., Zhang, P.: Introduction to Graph Theory. Posts & Telecom Press, Beijing (2007)

A Hybrid FES Rehabilitation System Based on CPG and BCI Technology for Locomotion: A Preliminary Study

Dingguo Zhang[1,⋆], Guangquan Liu[1], Gan Huan[1], Jianrong Liu[2], and Xiangyang Zhu[1]

[1] Institute of Robotics, School of Mechanical Engineering
[2] Institute of Neurology, Ruijin Hospital
Shanghai Jiao Tong University, China
dgzhang@sjtu.edu.cn

Abstract. A hybrid functional electrical stimulation (FES) rehabilitation system for locomotion is proposed in this paper. It has a hierarchical structure. In upper level, brain-computer technology (BCI) is used to acquire the subject's intention. In middle level, central pattern generator (CPG) is designed to generate appropriate rhythmic motor patterns. CPG is triggered by BCI command, and send control patterns to lower level. In lower level, FES technique is used to activate the muscles, and drive the lower limbs to achieve the expected movements (i.e. locomotion). The whole system is developed according to the general nervous structure of human being. The hybrid system aims at developing a neuroprosthetic bridge for the impaired nervous system of paralyzed patients. Some preliminary results on BCI are given.

Keywords: Functional electrical stimulation, brain-computer interface, central pattern generator, rehabilitation system, locomotion.

1 Introduction

Functional electrical stimulation (FES) is a popular technique used widely in rehabilitation engineering. It uses electrical current of low level to excitable tissue to supplement or replace function that is lost in neurologically impaired individuals [10]. This is a broad definition, and the targeted excitable tissue is skeletal muscle in our work.

Generally, a complete FES system contains four parts: controller, stimulator, musculoskeletal system (patients), and sensors. The kernel is the controller, whose role looks like the brain in human nervous system. The controller can regulate the stimulation pattern for stimulator with the feedback information. At present, various artificial control techniques have been applied into FES system such as PID control, model based control, sliding-mode control, fuzzy logic control, artificial neural network control, optimal control and so on [18]. However,

⋆ Corresponding author.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 1073–1084, 2009.

nearly all the controller design is derived from the classic traditional algorithms, which are merely developed by the researchers. The patient's own intention is not included. Therefore, the "functionality" is limited in some sense.

The ideal and ultimate objective of a neuroprosthetic or rehabilitation system should construct a bridge in such a routine "brain $\Longrightarrow$ spinal cord $\Longrightarrow$ muscle". Although this is a long-term dream for the researchers, it at least inspires us to provide a rudimental design using such an idea.

The promising technology of brain-computer interface (BCI) and central pattern generator (CPG) may give us possible solutions. BCI can acquire the intention from the human brain, and CPG can mimic the nervous function of spinal cord. The stream line is given as follows: CPG is triggered by BCI command; FES is controlled by CPG; FES activates the muscles and generates movements in the end. The general idea is illustrated in Fig. 1.

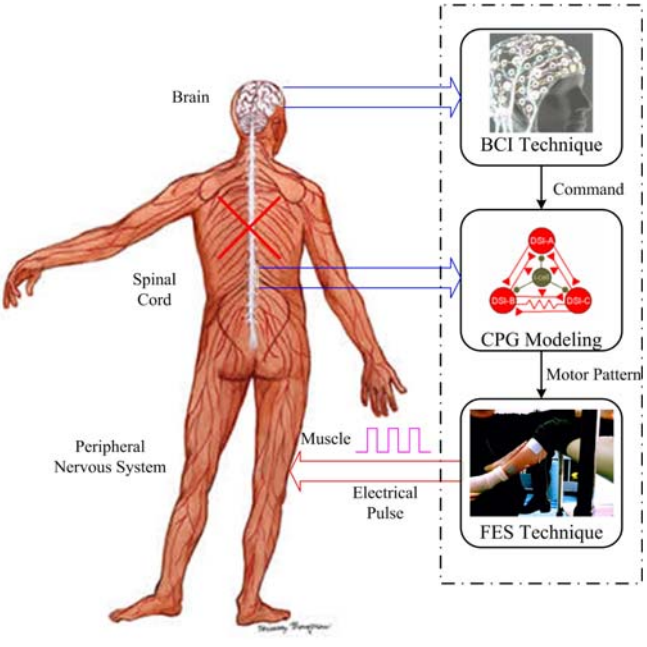


Fig. 1. General idea of a hybrid FES rehabilitation system for locomotion. It has a hierarchical structure. Upper level: BCI decodes the EEG signal and get the command. Middle level: CPG is triggered by BCI command, and generates the stimulation patterns for FES. Lower level: FES activates the muscles concerned.

A brief review on BCI and CPG is given in the following subsections. The methodology of each component of the whole system is discussed in Section 2, especially the algorithm for EEG-based BCI is given in detail. Preliminary result on EEG signal processing is given in Section 3. In the end, we make a conclusion and provide our future work in Section 4.

1.1 BCI Technology

BCI is developing rapidly in recent years. It is a communication system between human brain and external environment [15]. Researchers analyze the signals acquired from the brain using some specific algorithms. Some useful information may be distilled, for example, predicting the motion intention. BCI is successfully applied to control robotic arm using information from monkey brain [9], [14]. It becomes the leading-edge research topic since 2000. An upsurge arises using BCI to control external devices, such as communication system, wheelchairs, prostheses and so on [2], [15].

BCI in nature is a kind of technique, which contains magnetoencephalography (MEG), positron emission tomography (PET), function magnetic resonance imaging (fMRI), invasive electrophysiological method (e.g. ECoG), optical imaging, and electroencephalography (EEG) [15]. EEG is an important type of BCI, it is adopted by many researchers because of the noninvasive, economic, convenient and easy-to-use properties. Our work also adopts EEG in the BCI design.

In order to acquire the expected information from raw EEG signals. Many EEG signal processing methods are developed [7]. For feature selection, there are common space pattern (CSP), autoregressive (AR) / adaptive autoregressive (AAR) parameters, time-frequency features, and inverse model-based features. For pattern recognition, there are linear discriminant analysis (LDA), support vector machine (SVM), artificial neural network (ANN), nonlinear bayesian classifiers (NBC), and nearest neighbor classifiers (NNC) etc.

Is it possible to apply BCI technique into FES system? The answer is positive and the pioneer work was accomplished by group of Prof. Pfurtscheller at Graz University, Austria [11], [12]. Later, some researchers are devoted to developing BCI-FES system [1], [3]. In Asia, there are two groups working at this research topic at Hongkong Polytechnic University [8] and Nanyang Technological University [13] respectively.

Since the study of BCI-FES is still at the initial stage at present, there exist some big problems. First, the accuracy of pattern recognition for EEG signals needs improvement, and the movement dimensions that can be classified by EEG are still low (at most 2 dimensions with 4 directions). Second, BCI can only give the discrete simple commands that is very difficult to realize the continuous movements using FES. Third, the FES part in previous work is relatively too simple, a feedback loop is needed for better movement control. Fourth, a complete evaluation especially clinical evaluation of the BCI-FES system are still lacking.

1.2 CPG Technology

Neuroscience has already found that CPG is a type of neuronal circuit existing in the spinal cord of vertebrate [4]. CPG can generate the motor patterns for rhythmic movements (walking, flying, swimming etc.) independently without the continuous support from brain. CPG can also work without sensory feedback, but sensory feedback can enhance the adaptive ability of CPG.

In general, CPG may be classified into two types: biological CPG and engineering CPG. In neuroscience, researchers aims at exploring CPG at a microscopic level based on experiments in vivo, the different electrochemical ionic channels in neurons are studied. This kind of CPG has complex structure and numerous parameters, for example, the lamprey CPG [4]. Another direction targets the mechanism of CPG at a macroscopic level, i.e. the key function of CPG is mimicked while the structure is simplified. This kind of CPG can be analyzed mathematically and easy for the practical application. There are several classic models of engineering CPG, such as Matsuoka oscillator [17], Van der Pol oscillator, Hindmarsh-Rose model, Morris Lecar model and so on [6]. It should be noted that Hodgkin-Huxley model is the foundation for all these models. Engineering CPG is used to simulate the rhythmic movements for animal models such as fish, lamprey, lizard, cockroach etc; engineering CPG is also applied widely for robotic control such as biped, multi-leg robot, snake-like robot, and swimming robot [5].

Based on the previous research, some interesting properties of CPG are found. It has the strong robust stability properties of limit cycle behavior. The smooth online modulation of trajectories can be acquired by tuning the parameters of the dynamical system. The entrainment phenomenon is a useful feature when CPG is coupled with a mechanical system. Much less amount of calculation is required for movement control because of the coordination of physical parts induced by the rhythmic movements. During recent years, adaptive mechanism offers CPG with more powerful functions.

Although CPG has some advantages in purpose of control, it is rarely used in FES system before. Our previous work has provided the rudimental exploration on designing FES control system based on CPG [19]. We find CPG is specially suitable for FES control. (1) FES is a type of neuroprosthetic technology. CPG just "copy" the function from neuronal circuit in spinal cord. It is in accordance with the nature of neuroprostheses. (2) Our project focuses on studying FES waking control. Walking is a typical rhythmic movement that is controlled by CPG in spinal cord of the healthy person. (3) FES application inherently has significant time delay. For realtime control, predictive and adaptive functions are desired for controller design. CPG can serve as a feedforward controller with adaptive mechanism, which can meet the FES demand exactly. (4) CPG network is very suitable for the movement control involved with multi-muscle and multi-joint. It is a good alternative to extend the current FES application from single joint to multiple joints using CPG network.

2 Methodology

All the previous research regarding BCI-FES system focuses on the upper limb movements [1], [3], [8], [11], [12], [13]. Their controlled objective is different from ours. Our work targets the lower limb movement (i.e. walking). Different from upper limb movements that are totally controlled by brain, spinal cord circuit plays a main role in lower limb movements (i.e rhythmic movements) in

healthy persons. If we wish to achieve an ideal movement for upper limb such as lifting, reaching and grasping for paralyzed patients using BCI-FES system, the continuous control signal from brain must be achieved. Obviously, this is an impossible mission especially for EEG-based BCI at present. If we target the control of lower limb movements, the task becomes possible.

2.1 General Scheme

The feasibility of the hybrid system is discussed at first. According to the survey in previous section, the current research level of EEG-based BCI is still low, and only simple movement intentions in 2-dimension can be classified. But this will not affect our general idea. Since CPG can work independently without continuous supraspinal input, a simple command (e.g. stop or start) from the brain is enough to trigger it. CPG can deal with the continuous control for walking in the subsequent process.

The main work will be conducted at five levels as shown in Fig. 2 that includes EEG signal processing, CPG design, simulation study (based on musculoskeletal model), feedback information processing, and FES experiment. The EEG processing has four steps: (1) EEG data acquisition. The appropriate channels should be selected according to the type of signals wanted. (2) EEG feature selection. (3) EEG feature extraction and preprocessing. (4) EEG classification. For CPG design, there are three problems should be resolved: (1) CPG network design, (2) CPG parameter setting (3) CPG adaptive mechanism.

In simulation study, a musculoskeletal model of lower limb will be developed. It serves as a controlled plant, and test the performance of the system proposed. Feedback information is important for CPG design. The feedback information could be EMG, joint angle (angular velocity) or other motion signals. The experimental work is a hard and tedious task. Hardware and software platform should be developed. Healthy subjects will take part in the early evaluation of the hybrid FES system. Patients are also expected to be recruited, and the clinical test will be performed under help of doctors.

2.2 BCI Algorithm

The previous research has shown that *mu* or *beta* rhythms could be good signal features for EEG-based BCI communication [15]. Movement or preparation of movement will make a decrease in *mu* and *beta* rhythm that is named as "event-related desynchronization" (ERD). On the contrary, there is a rhythm increase after movement and with relaxation named as "event-related synchronization" (ERS). ERD and ERS do not need actual movement, they can occur with imaginary movement or movement intention. These features are very suitable for the paralyzed patients.

Common space pattern (CSP) is proved to be suitable for the ERD detection. It is a type of supervised spatial filtering methods for two-class discrimination problems, which finds directions that maximize variance for one class and at the same time minimize variance for the other class. In this work, CSP is adopted.

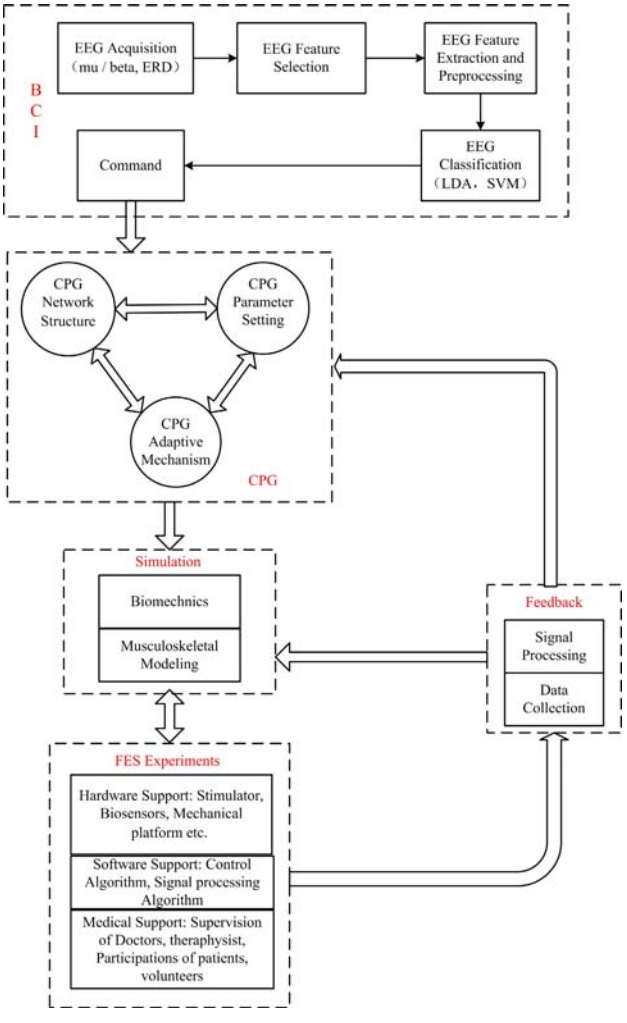


Fig. 2. General plan of the proposed FES system. It contains five parts: BCI processing, CPG design, simulation study, FES experiment, and feedback system.

The CSP algorithm is briefly introduced as follows. Let X_d^i, $d \in 1, 2$ denote the zero-mean EEG recordings of trial i, class d, and its dimension is $N \times M$, with N number of channels and M number of samples in time. The normalized spatial covariance of this trial can be written as

$$R_d^i = \frac{X_d^i X_d^{i*}}{trace(X_d^i X_d^{i*})} \tag{1}$$

where * indicates the conjugate transpose of a matrix (transpose for real matrix) and $trace(\cdot)$ means the sum of elements on the diagonal of a matrix. For each class, the normalized covariance matrices are averaged over trials,

$$R_d =< R_d^i >, d \in 1, 2 \tag{2}$$

The sum of the two averaged normalized covariances is

$$R_c = R_1 + R_2 \tag{3}$$

R_c can be decomposed as

$$R_c = U_c \lambda_c U_c^* \tag{4}$$

where U_c is the matrix of eigenvectors and λ_c is the diagonal matrix of eigenvalues. The whitening transformation matrix is given as

$$P = \sqrt{\lambda_c^{-1}} U_c^* \tag{5}$$

R_c can be whitened by P as

$$I = P R_c P^* \tag{6}$$

where I denotes the identity matrix. Let's transform R_1 and R_2 as

$$S_d = P R_d P^*, d \in 1, 2 \tag{7}$$

then S_1 and S_2 share common eigenvectors, and the corresponding eigenvalues for the two matrices sum up to 1. If

$$S_1 = B \lambda_1 B^* \tag{8}$$

then

$$S_2 = B \lambda_2 B^*, \lambda_1 + \lambda_2 = I \tag{9}$$

It means that the eigenvector with largest eigenvalue for S_1 has the smallest eigenvalue for S_2 and vice versa. The rows of W are called spatial filters, and the columns of W^{-1} are called spatial patterns. EEG recordings of trial i can be decomposed with W as

$$Z^i = W X^i \tag{10}$$

After the decomposition, the components most suitable for discrimination are the first and last few rows of Z. The normalized log-variances of these components are used as features. Linear classifiers can be used to classify these features.

In this paper, Fisher's LDA classifier is used. The brief description about the algorithm is given. If μ_1 and μ_2 are the means of two classes, and their corresponding covariance matrices are $\sum_1$ and $\sum_2$. The weight of LDA classifier is

$$w = (\sum\nolimits_1 + \sum\nolimits_2)^{-1} \cdot (\mu_1 - \mu_2) \tag{11}$$

and the bias is

$$b = -w^T \cdot (\frac{\mu_1 + \mu_2}{2}) \tag{12}$$

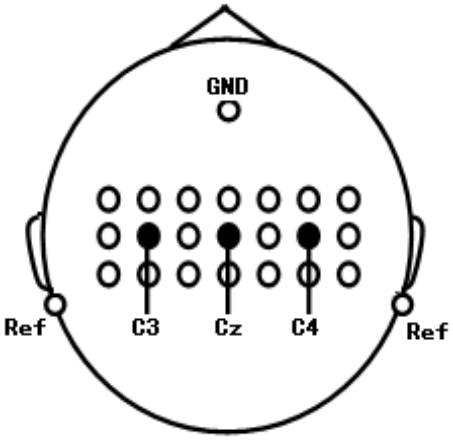


Fig. 3. Electrode positions for raw EEG collection. 'GND' means ground electrode, and 'Ref' means reference electrodes. The best three channels are from C_3, C_4 and C_z.

where w^T means the transpose of vector w. For each feature vector x, the classifier output is and the bias is

$$y = w^T \cdot x + b \tag{13}$$

If $y > 0$, then x is classified into class 1, other class 2.

The parameter set is defined as $\Theta = \{\mu_1, \mu_2, \sum_1, \sum_2\}$. In order to update the classifier, Θ should be updated. In a supervised way, where each trial before the current time is given a label. If $x(t)$ is the latest feature vector and its label is known, Θ can be updated as

$$\mu_i(t) = \mu_i(t-1) \cdot (1 - UC) + x(t) \cdot UC \tag{14}$$

$$\sum\nolimits_i (t) = \sum\nolimits_i (t-1) \cdot (1 - UC) + (x(t) - \mu_i(t)) \cdot (x(t) - \mu_i(t))^T \cdot UC \tag{15}$$

where i is the true label of $x(t)$, and UC is the update coefficient.

3 Experimental Work

Some preliminary results on EEG signal processing are given in this section.

3.1 EEG Data Collection

EEG signals were recorded using a SynAmps system (Neuroscan, USA). Signals from 21 channels over the related motor areas were recorded. The grounding electrode was mounted on the forehead and reference electrodes on the left and right mastoids. The electrodes were placed according to the extended 10/20-system (see Fig. 3). Horizontal and vertical EOGs were recorded for the purpose

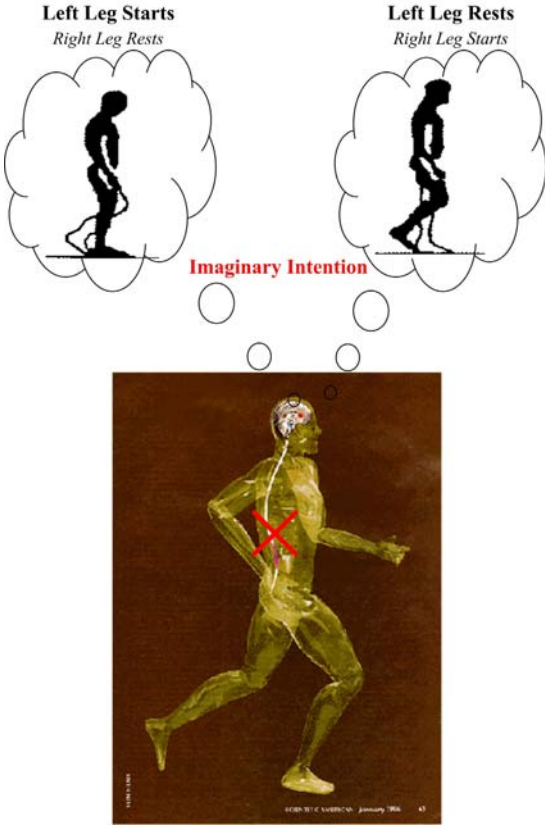


Fig. 4. Subjects imagine the left leg (or right leg) to move. This intention may be detected from EEG, and trigger CPG as a command "start to walk".

of artifact detection, and were not used for classification. The EEGs were first filtered by the recording system in a 5-30Hz frequency band, and the sampling rate was 1000Hz. Before feature extraction and classification, the signals were down-sampled to 200Hz and refiltered by a 50-order FIR filter in 8-30Hz frequency band. By the high-pass filtering, low-frequency EOG artifacts were also removed.

There are 2 subjects taking parts in the data collection. The subjects will imagine to initiate right or left leg to walk as shown in Fig. 4. The experimental procedure is given as follows. The subjects were seated in a comfortable armchair about $2m$ in front of a computer monitor. They were instructed to keep still and avoid blinking during a trial. At the beginning of each trial, the screen was black. After 1 second, a fixation cross appeared in the center of the screen. Another second later, an arrow pointing to either left or right was added to the cross indicating the imagination of left or right leg movements. The arrowed cross was showed until the end of the 5th second. During this time period (from the 2nd

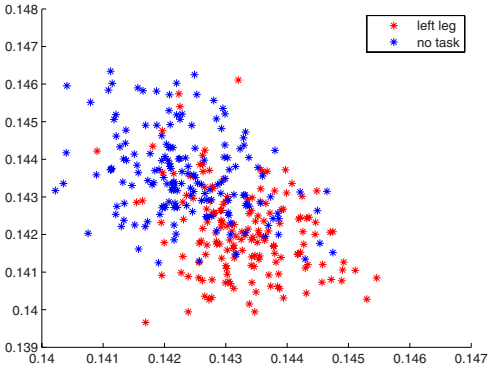


Fig. 5. Features based on CSP about left leg movement. Red stars indicate the features of "left leg movement", blue stars indicate the features of "no movement".

second to the 4th second), the subject had to imagine left or right leg to move. After an random interval varying from 1.5 to 2.5 seconds, the next trial began. The sequence of left and right trials was randomized and the chance for each class was flat. There were five runs in each session and three sessions for each subject. All data were saved for later analysis.

3.2 Results

Subject 1 imagined the left leg to move. There are two states: start walking and no walking. The result of EEG feature classification of Subject 1 is shown

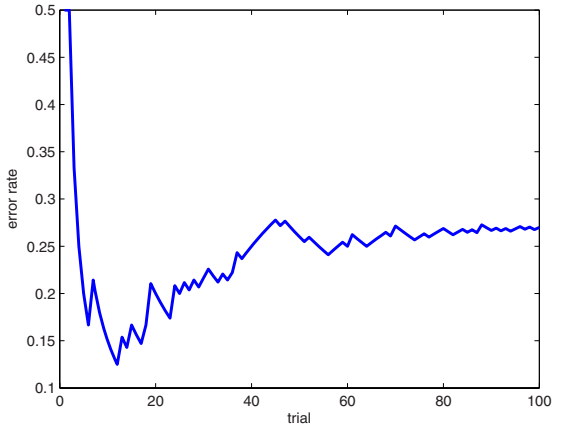


Fig. 6. Error rate of classification (LDA method) on EEG feature of left leg movement vs. no task

in Fig. 5. The error rate of classification is shown in Fig. 6. We found the rate of accuracy is about 75%. However, the rate of accuracy is only about 65% for subject 2.

We also performed the study on classification for the two imaginary states of initiating left leg or right leg to walk. The results are not satisfactory. The possible reason is that the corresponding areas of motor cortex in charging of left leg and right leg movements are too close, since they are both located in the central area of brain cortex. Further exploration will be conducted on this issue.

4 Conclusion and Future Work

In this paper, we proposed the idea to develop a hybrid FES rehabilitation system for locomotion. Both CPG and BCI technologies are incorporated and assembled harmoniously. A neuroprosthetic bridge in such configuration is expected to develop in future.

The current work is focused on the processing of EEG-based BCI. The preliminary results are presented. However, more effortful work are needed to improve the algorithms in order to get more accuracy EEG command. Frequency-weighted method (FWM) is considered to apply into CSP feature extraction. An unsupervised adaptation approach based on variational Gaussian mixture model (VGMM) will be proposed for LDA classification. Also, more experimental work will be conducted to collect EEG data from more subjects to test the performance of algorithms.

CPG design is not discussed in this paper. Actually, we have done much work on CPG including the theoretical work [16], [17] and the application [19]. If the work of BCI part is satisfactory, we will connect CPG with BCI, and then the kernel of FES system can be developed. After that, evaluation of such a hybrid FES system will be performed on a musculoskeletal model.

In the end, it will be tested on healthy subjects or paralyzed patients. Obviously, this is a long journey, but we are quite optimistic about such a promising and interesting research direction.

Acknowledgement

This work is supported by SJTU Interdisciplinary Fund in Medicine and Engineering (Grant No. YG2009MS45), and Science & Technology Commission of Shanghai Municipality (Grant No. 09JC1408400).

References

1. Bentley, A.S.J., Andrew, C.M., John, L.R.: An offline auditory P300 brain-computer interface using principal and independent component analysis techniques for functional electrical stimulation application. In: Proceeding of 30th EMBC, pp. 4660–4663 (2008)

2. Birbaumer, N.: Breaking the silence: Brain-computer interfaces (BCI) for communication and motor control. Psychophysiology 43, 517–531 (2006)
3. Boord, P., Barriskill, A., Craig, A., Nguyen, H.: Brain-computer interface-FES integration: Towards a hands-free neuroprosthesis command system. Neuromodulation 7(4), 267–276 (2004)
4. Grillner, S.: Neurobiological bases of rhythmic motor acts in vertebrates. Science 228, 143–149 (1985)
5. Ijspeert, A.J.: Central pattern generators for locomotion control in animals and robots: a review. Neural Networks 21, 642–653 (2008)
6. Izhikevich, E.M.: Which model to use for cortical spiking neurons? IEEE Trans. Neural Networks 15, 1063–1070 (2004)
7. Lotte, F., Congedo, M., Lecuyer, A., Lamarche, F., Arnaldi, B.: A review of classification algorithms for EEG-based brain-computer interfaces. J. Neural Engineering 4, R1–R13 (2003)
8. Meng, F., Tong, K.Y., Chan, S.T., Wong, W.W., Liu, K.H., Tang, K.W., Gao, X.R., Gao, S.K.: BCI-FES training system design and implementation for rehabilitation of stroke patients. In: Proceed. IEEE IJCNN conference, pp. 4103–4106 (2008)
9. Nicolelis, M.A., Chapin, J.K.: Controlling robots with the mind. Scientific American 287(4), 46 (2002)
10. Peckham, P.H., Knutson, J.S.: Functional electrical stimulation for neuromuscular applications. Annu. Rev. Biomed. Eng. 7, 327–360 (2005)
11. Pfurtscheller, G., Mueller, G.R., Pfurtscheller, J., Gerner, H.J., Rupp, R.: Thought' - control of functional electrical stimulation to restore hand grasp in a patient with tetraplegia. Neuroscience Letter 351, 33–36 (2003)
12. Pfurtscheller, G., Mueller-Putz, G.R., Pfurtscheller, J., Rupp, R.: EEG-based asynchronous BCI controls functional electrical stimulation in a tetraplegic patient. EURASIP Journal on Applied Signal Processing 19, 3152–3155 (2005)
13. Tan, H.G., Zhang, H.H., Wang, C.C., Zhang, D.G., Shee, C.Y., Ang, W.T., Guan, C.T.: A step towards discretized motion control of the upper limb using brain-computer interface and electrical stimulation. In: Proc. 13th IFESS conference (2008)
14. Velliste, M., Perel, S., Spalding, M.C., Whitford, A.S., Schwartz, A.B.: Cortical control of a prosthetic arm for self-feeding. Nature (2008) doi:10.1038/nature06996
15. Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., Vaughan, T.M.: Brain-computer interfaces for communication and control. Clinical Neurophysiology 113, 767–791 (2002)
16. Zhang, D.G., Zhu, K.Y.: Computer simulation study on central pattern generator: From biology to engineering. Intern. J. Neural Systems 16, 405–422 (2006)
17. Zhang, D.G., Zhu, K.Y.: Theoretical analysis on neural oscillator toward bio-mimic robot control. Intern. J. Humanoid Robotics 4, 697–715 (2007)
18. Zhang, D.G., Tan, H.G., Ferdinan, W., Ang, W.T.: Functional electrical stimulation in rehabilitation engineering: A survey. In: Proceed. Intern. Conf. Rehab. Eng. & Assist. Tech., Singapore, pp. 221–226 (2007)
19. Zhang, D.G., Zhu, K.Y.: Modeling biological motor control for human locomotion with functional electrical stimulation. Biol. Cybern. 96, 79–97 (2007)

Introduction to the Development of a Robotic Manipulator for Nursing Robot

Guang Lu, Binbin Tao, Jun Liu, Fang Chen, Jinfei Shi, and Zhisheng Zhang

School of Mechanical Engineering, Southeast University, Nanjing, 211189, China
luguangseu@seu.edu.cn, chris11@seu.edu.cn, oldbc@seu.edu.cn

Abstract. The structure and function system of a practical robotic manipulator applied to nursing robot has been developed. According to the development requirements of practicality and low cost, a 4-DOF lightweight manipulator was firstly designed. Then its kinematics model was established to obtain the forward and inverse kinematics solutions. As to the manipulator control system, CAN-bus architecture was utilized to implement communication between the main control system of robot and the joint control unit of manipulator. Taking the terminal gripper controller as an example, the modular circuit design and the concrete communication rules were emphatically introduced. Furthermore, based on the service task of grasping water cup, a series of core technical schemes on camera calibration and object location were put forward under the measurement condition of monocular vision to achieve visual servo control. The experimental results show that the manipulator can preliminarily realize the predetermined service functions.

Keywords: Service robot, robotic manipulator, CAN-bus, monocular vision, object grasping.

1 Introduction

The rise and popularization of robot technology has already caused a great transformation in many fields of science and technology. Especially from the 90s of last century, robot has not just played a significant role in industrial fields, its application situation has also expended to non-industrial fields. Various robots applicable for service and entertainment have entered families, which gradually become an important part in people's daily lives.

The development of service robot is a research focus of current robot domain. Many valuable international research results have been generated in recent years. A 5-DOF lightweight manipulator applied to home or work situation of the elderly and the disabled was developed by University Carlos of Madrid in Spain. Its self-bring control system can adapt to many situations [1]. A multi-function health wheelchair was developed by National Institute of Standards and Technology in the United States. It can help the aging populations, stroke victims and wheelchair dependents to move independently, to lift themselves and even to do rehabilitation training [2]. During the cooperative research between University

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 1085–1096, 2009.

of Southern California in the United States and Electronics and Telecommunications Research Institute in the Republic of Korea, the vision perception process of personal service robot was divided into three stages: self-location, long range people detection and tracking, and short range human interaction [3]. Also targeting at improving the perception ability of service robot, researchers in University of Essex in the United Kingdom put forward that face recognition can be combined with other surface characteristics to distinguish unknown people effectively [4]. A robot being able to recognize and collect waste in office-like environment was developed by researchers in "POLITEHNIC" University of Bucharest in Romania. The robot can establish three-dimensional virtual map to realize interaction with other robots and people through stereo vision [5].

Presently, with the accelerated aging process of humans, extensive attention has been put on the development of service robot for the elderly and the disabled [6,7]. In China, the National High Technology Research and Development Program (863) has encouraged the cooperation of R&D institutions and enterprises to develop some conception prototypes, which can designate the great prospects of home service robot. The Service Robot Laboratory in School of Mechanical Engineering, Southeast University, participated in the research projects and had obtained preliminary results [8,9,10]. Based on these forward researches, we will present a novel development of a robotic manipulator system which is applied to nursing robot to grasp water cup for humans.

2 Mechanical Design and Kinematics

2.1 Requirements and Schemes

The project requires that the single manipulator should be of simple mechanism and low cost, and combined with control system and vision system, it can realize fetching cup, pouring water and some other services based on transformed household appliances. The final schemes are as follows:

1. The 4-DOF configuration of the manipulator includes: the shoulder parallel movement, the shoulder rotation, the elbow rotation and the wrist rotation round the arm axis. The terminal gripper is designed as humanoid style.
2. The major material of the manipulator is aluminum profile, which not only guarantees structure lightness and attractive appearance, but also has enough stiffness. The length of the manipulator is around 60 cm.
3. The majority joints of the manipulator are driven by DC servo motors. The action of the gripper is controlled by a servo rudder, which has small volume and can be easily mounted and controlled, and thus system cost is reduced.
4. The design weight of the manipulator is around 1.5 kg (not including devices mounted on the robot mobile platform). The weight of water cup for grasping is around 0.5 kg.

2.2 Mechanical Design

The manipulator structure is shown in Fig. 1, and the parts in circles are joints. The shoulder joint shown in Fig. 2 has characteristics as follows.

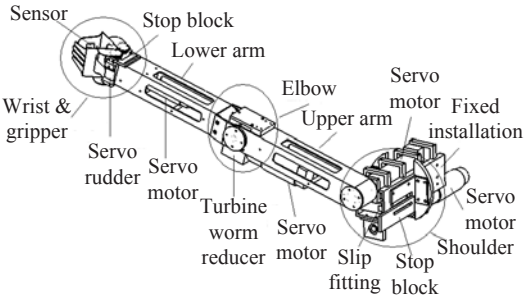


Fig. 1. The structure of the whole manipulator

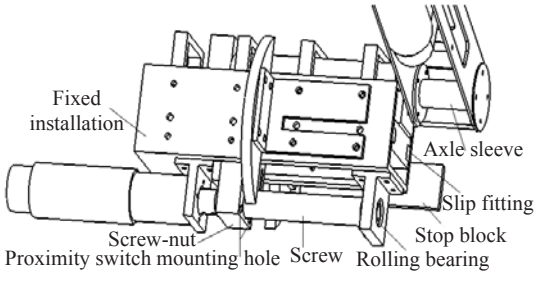


Fig. 2. The structure of the shoulder joint

1. The aluminum profile with sectional area of 60 $mm \times 60$ mm is used for fixed installation, and the guide is mounted on inner side surfaces. In order to prevent deformation of the open profile, C shaped parts are designed to connect two side surfaces of profile into integration.

2. The length of the guide support frame is 200 mm, and the moving distance of shoulder should not exceed 40% of it, which means around $70 - 80$ mm. In order to prevent excessive movement, proximity switch mounting holes are designed on the screw-nut, and proximity sensors are used to set limit position.

3. The servo motor controlling the parallel movement of shoulder is fixed under the support frame. The height of the screw-nut in the vertical direction can be adjusted to make the installation of screw and motor easily. In order to prevent jitter during motor rotation, the other end of the screw uses a rolling bearing as fulcrum.

4. The shoulder rotation is directly driven by a servo motor to eliminate possible errors. The barbell shaped motor axle sleeve is designed to be embedded in the internal aluminum profile of arm, which can transmit torque and at the same time improve stiffness of the shoulder joint.

The structures of the elbow joint and the wrist joint are shown in Fig. 3. The turbine worm reducer realizes right angle turning of the elbow motor motion, and it is fixed on both end face and side face. In addition, another four aspects are concerned in the design. Firstly, the connection part of the upper and lower

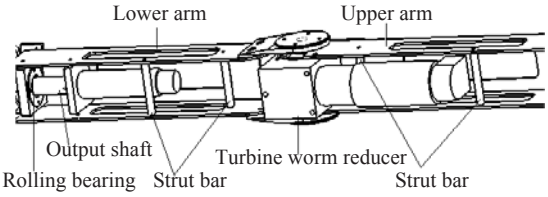


Fig. 3. The structure of the elbow joint and the wrist joint

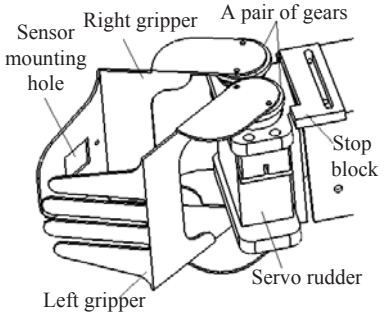


Fig. 4. The structure of the terminal gripper

arms does not generate interference during motion process. Secondly, the offset distance between the upper and lower arms is zero, which is convenient for kinematics calculation. Thirdly, the lengths of the upper arm and the low arm are designed as equal to be human-like. Fourthly, strut bars are embedded in the open profile to improve stiffness.

The structure of the terminal gripper is shown in Fig. 4. The simultaneous motion of the two gripper pieces is realized through a pair of gears with the same modulus, and under the condition of determined center distance, the tooth number of gears is increased to reduce transmission error brought by tooth backlash. The material of the gripper is stainless steel with a thickness of 1 mm, which is light and firm. A rectangular hole is designed on the gripping surface for mounting contact sensor to realize force control during object grasping. The maximum dimension of the terminal gripper is 64 mm.

2.3 Kinematics

The kinematics coordinates are established in Fig. 5. The origin of manipulator coordinate system is located at the intersection of the shoulder rotation axis and the upper arm axis, when the shoulder is at the initial place.

Here, we can establish the transformation matrix between adjacent coordinate systems as follows. Among them, θ_1, θ_2 and θ_3 are rotation angles of the shoulder, the elbow and the gripper.

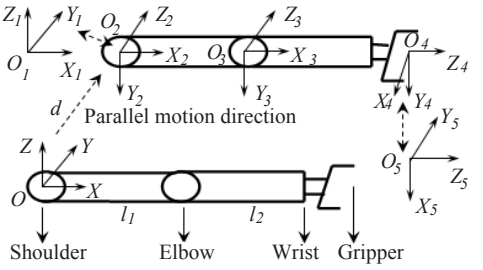


Fig. 5. The kinematics coordinates establishment of the manipulator

$$A_1^0 = \begin{bmatrix} cos\theta_1 & -sin\theta_1 & 0 & 0 \\ 0 & 0 & 1 & d \\ -sin\theta_1 & -cos\theta_1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{1}$$

$$A_2^1 = \begin{bmatrix} cos\theta_2 & -sin\theta_2 & 0 & l_1 \\ sin\theta_2 & cos\theta_2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{2}$$

$$A_3^2 = \begin{bmatrix} 0 & 0 & 1 & l_2 \\ sin\theta_3 & cos\theta_3 & 0 & 0 \\ -cos\theta_3 & sin\theta_3 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{3}$$

Then, the whole kinematics transformation matrix from the origin of the manipulator coordinate system to the final gripper coordinate system is obtained. Pay attention to that s donates sin, c donates cos, c_{12} donates $cos(\theta_1+\theta_2)$, and s_{12} donates $sin(\theta_1+\theta_2)$.

$$A_3^0 = \begin{bmatrix} -s_{12}s_3 & -s_{12}c_3 & c_{12} & c_{12}l_2 + c_1l_1 \\ -c_3 & s_3 & 0 & d \\ -c_{12}s_3 & -c_{12}c_3 & -s_{12} & -s_{12}l_2 - s_1l_1 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{4}$$

Furthermore, we suppose the predetermined posture matrix of the terminal gripper has the following form.

$$P = \begin{bmatrix} n_x & o_x & a_x & p_x \\ n_y & o_y & a_y & p_y \\ n_z & o_z & a_z & p_z \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{5}$$

Therefore, the forward and inverse kinematics solutions of the manipulator can be derived.

$$A_3^0 = P. \tag{6}$$

$$\begin{cases} \theta_1 = arcsin(\frac{p_x}{\sqrt{(sin\theta_2 l_2)^2+(cos\theta_2 l_2+l_1)^2}}) + arctan(\frac{cos\theta_2 l_2+l_1}{l_2 sin\theta_2}) \\ \qquad\qquad p_z > 0, p_z > p_x, p_z > -p_x \\ \theta_1 = 180^\circ - arcsin(\frac{p_x}{\sqrt{(sin\theta_2 l_2)^2+(cos\theta_2 l_2+l_1)^2}}) + arctan(\frac{cos\theta_2 l_2+l_1}{l_2 sin\theta_2}) \\ \qquad\qquad p_z \leqslant 0, p_z < p_x, p_z < -p_x \\ \theta_1 = -arcsin(\frac{p_z}{\sqrt{(sin\theta_2 l_2)^2+(cos\theta_2 l_2+l_1)^2}}) - arctan(\frac{sin\theta_2 l_2}{cos\theta_2 l_2+l_1}) \\ \qquad\qquad p_x > 0, -p_x \leqslant p_z, p_z \leqslant p_x \\ \theta_1 = 180^\circ + arcsin(\frac{p_z}{\sqrt{(sin\theta_2 l_2)^2+(cos\theta_2 l_2+l_1)^2}}) - arctan(\frac{sin\theta_2 l_2}{cos\theta_2 l_2+l_1}) \\ \qquad\qquad p_x \leqslant 0, p_x \leqslant p_z, p_z \leqslant -p_x \\ \theta_2 = \pm arccos(\frac{p_x^2+p_z^2-l_1^2-l_2^2}{2l_1 l_2}) \qquad p_z > 0, \theta_2 < 0; p_z < 0, \theta_2 > 0 \\ \theta_3 = arccos(\frac{o_x}{-sin(\theta 1+\theta_2)}) \\ d = p_y \end{cases} \quad (7)$$

Finally, the initial pose of the manipulator should be determined before executing the grasping function. Because what the motors utilize are incremental encoders, the current angle can not be obtained directly. Therefore, stop blocks mounted on the joints are designed. Once the robot starting, each motor rotates with the appointed direction, and the motor controller can determine whether the joint has already moved to the limit position or not by detecting current change. If the answer is yes, motors will be controlled to rotate an appointed angle in the opposite direction, thus the pose initialization is completed. Preliminary experiments show that the terminal gripper can arrive at the predetermined points in the debugging process, which lays a solid foundation for further realization of visual servo control.

3 Manipulator Control System

The manipulator control system is based on ARM and DSP. The task of ARM is to control robot motion, while DSP is applied to vision image processing, and the communication of them is implemented through HPI interface. We utilize the servo motor controller SDA11 provided by MAXON company, and it saves a lot of time in developing manipulator servo motor controller in the initial experiments, for these kinds of controllers involve a lot of technological details in dynamics of mechanical system. Therefore, we would only emphatically introduce the design of the terminal gripper controller, which is used to control the servo rudder. The interaction between main control board of robot and the gripper control unit is achieved through CAN-bus.

The function diagram of the gripper control system is shown in Fig. 6. LPC2119 produced by Philips is chosen as the core chip. It is a 16/32 bit ARM7TDMI-S CPU which supports the functions of real-time simulation and embedded trace, and it has two independent CAN interfaces.

The control circuit of the gripper control system includes power module, CAN-bus module, debugging module, servo rudder control module and sensor module,

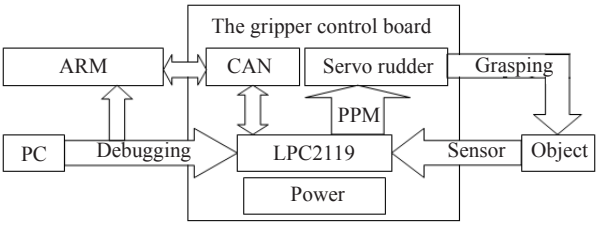


Fig. 6. The function diagram of the gripper control system

etc. Among them, CAN-bus module can be divided into three parts: LPC2119 (integrated CAN communication controller), 6N137 (high speed photoelectric coupler) and 82C250 (CAN-bus transceiver). Sensor module is designed as an external interrupt model, which means that the grippers do not stop closing during grasping until the contact sensor generates signal informing LPC2119 that the object has been grasped.

The program structure of the gripper control system includes CAN communication program, interrupt service routine, servo rudder control program and main program. The main program calls the CAN communication program to complete instruction reception and transmission, uses the interrupt service routine to response to external sensor signal, and controls the gripper action through the servo rudder control program. The instruction format is shown in Table. 1, and attention should be paid to that Addrc is module address and Addrz is host address.

Table 1. CAN-bus communication instruction format

Instruction format	Host transmission	Return value
Address setting	0x00 Addrc Addrz	Addrc Addrz
Motor operation	0x08 Dir Angle	Addrc 0xaa
Position acquisition	0x01 0x00 0x00	Addrc Angle
Instant stop	0x09 0x00 0x00	Addrc 0xbb
Zero resetting	0x06 0x00 0x00	Addrc 0xcc

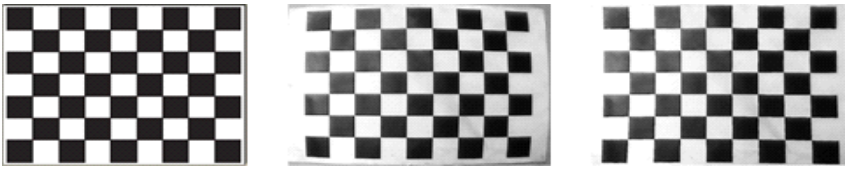
According to above CAN communication rules, the terminal gripper can rotate a certain angle instructed by the main control board of the robot. The motion of the servo rudder is smoothed by a cubic polynomial interpolation, and the velocity is also regulated by decreasing the change rate of PPM signal. After experimental test, the practical gripper satisfies the design requirements on precision and torque, and it can complete the task of grasping water cup for humans.

4 Vision System Scheme

In this paper, one camera is utilized to configure the robot to obtain the space location of object, which plays an important role in guiding path planning. Therefore, the vision system model should be established.

4.1 Camera Calibration

The monocular vision system uses a wide-angle camera to extend the visual view. It is well known that the imaging distortion (see Fig. 7 (a) (b)) should firstly be corrected before camera calibration. When the external parameters are still unknown, the basic method of calibration is to use a standard labeled graph, which here is referred to a printed 7×9 checkerboard.



(a) Theoretical image. (b) Actual image. (c) Distortion correction image.

Fig. 7. Imaging distortion and its correction

According to the distortion model, we only consider the radial distortion while the tangential distortion and the high order terms are ignored. The ideal and actual locations of image points are fitted according to the camera distortion model, and the distortion factor can be calculated with Ordinary Least Square (OSL) method. Substitute the actual coordinates of each pixel to the equations above, and then the theoretical coordinates can be obtained. Furthermore, move the gray value of actual coordinates to theoretical coordinates, and the completed corrected image is shown in Fig. 7(c). As the image is stretched, blank pixels are generated, and their values can be determined by linear interpolation of adjacent pixels. Finally, a linear calibration is implemented with the modified camera model, and the transformation form from the imaging coordinate system to the space coordinate system is shown as follows.

$$x_d = \begin{bmatrix} x_{d1} \\ x_{d2} \end{bmatrix} = (1 + kc_1 r^2 + kc_2 r^4 + kc_5 r^6) x_n + \begin{bmatrix} 2kc_3 xy + kc_4(r^2 + 2x^2) \\ kc_3(r^2 + 2y^2) + 2kc_4 xy \end{bmatrix}. \quad (8)$$

$$z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} ff_u & -ff_u cos\theta & u_0 \\ 0 & ff_v / sin\theta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R\, t \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}. \quad (9)$$

4.2 Object Location

The study of object location is carried on through the procedures such as threshold segmentation, connected domain extraction, object recognition, and object location. The chosen algorithm is balanced between efficiency and precision.

The threshold segmentation is primarily based on an operation of grayscale image. But generally, as a result of bad illumination and complex background, the object can not be fully extracted from the image background through grayscale image threshold segmentation, which affects the further image analysis. Therefore, we adopt a color segmentation method, and YUV coding of a PAL camera is used to directly set the threshold values of chromatism components Cr and Cb. The preliminary experiments demonstrate the robustness of color segmentation to different illumination, and the object can still be segmented from even more complex background.

As to connected domain extraction, the interval list algorithm is utilized and its core idea is as follows. The connected domain is regarded to be composed by line segments, and each line segment can be represented by its two ending points. While scanning the image, encode the line segments for interval lists, make the same label to line segments in the same connected domain, and then region extraction is accomplished. The conventional algorithm need to scan the image four times, however, in order to improve efficiency in each scanning time, during programming we combine related procedures when scanning the adjacent rows. Firstly, detection on connected domain is carried on rightly after the starting point and the ending point are attained, and the same label is assigned to the connected line segments in the current and the upper line. If a line segment in the current line is connected to multiple line segments in the upper line, establish equivalence relation to these labels of line segments, and then represent all these line segments using the minimum label. Secondly, rescanned the image for endowing uniform labels to the connected line segments with equivalent labels, and then all the connected domains can be obtained. Finally, the connected

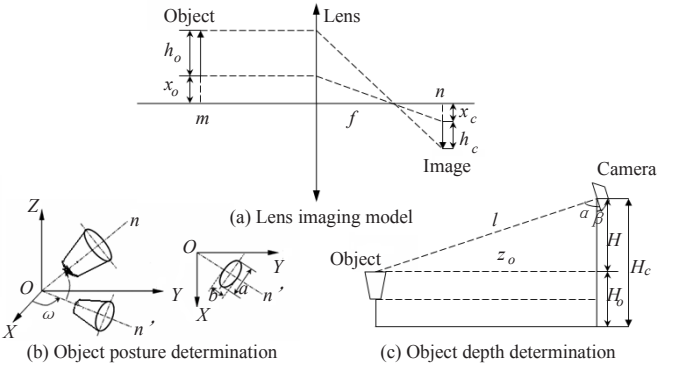


Fig. 8. Object location for manipulator grasping

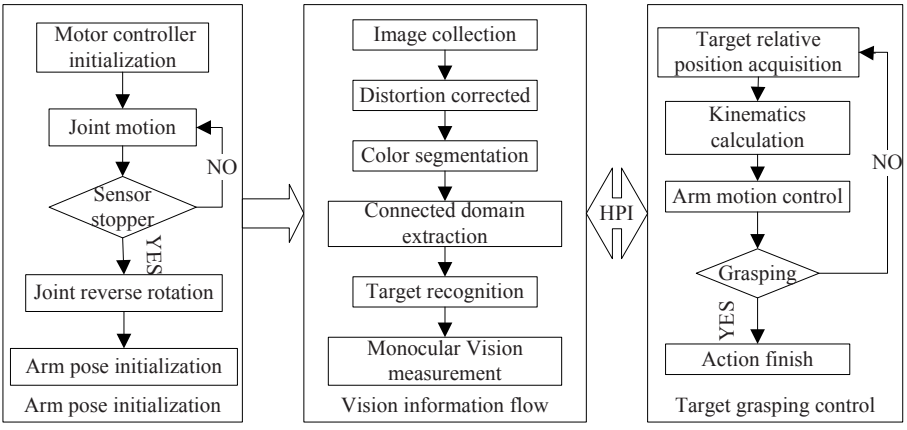


Fig. 9. The whole working procedure architecture of the nursing robot

domain with area smaller than a predetermined threshold value is excluded in advance so as to reduce calculation load of further object recognition.

When the connected domain extraction is finished, the vision system recognizes a cylindrical water cup according to its geometrical property. Width and height of the connected domain are calculated to obtain the ratio of them, and center of the connected domain is calculated to use the symmetry property as a help to object recognition. As to object location, the prior knowledge (the real size of the water cup) is needed to obtain depth information when only one camera is utilized. The concrete procedures are listed.

1. The lens imaging model is shown in Fig. 8 (a), and the relationship among focal length, object distance and image distance can be donated as $1/f = 1/m + 1/n$. When $m \gg n$, n can be approximately regarded equal to f, and then the pinhole imaging model is obtained. The camera intrinsic parameters have been calibrated, thus the focal length f has already been determined. Then the distance l from the upper surface of the water cup to the origin of camera coordinate system can be determined using proportion relation between the upper surface geometry size of the water cup and its imaging size. The position information of object in XY plane of the camera coordinate system can be provided by coordinate difference between the projection center and the camera principle point in the imaging plane.

2. As Fig. 8 (b) shows, XY plane is the imaging plane, Z axis is the optical axis of camera, and n' is the projection of symmetry axis n of the water cup in the imaging plane. Because the projection of a space circle is generally an ellipse, $\alpha = sin^{-1}(b/a)$ is the angle between the symmetry axis n and the imaging plane, where a is the projected long axis and b is the projected short axis. Then in Fig. 8 (c), under the condition of knowing camera pitching angle β, the object depth can be determined by $z_0 = lsin(\alpha - \beta)$.

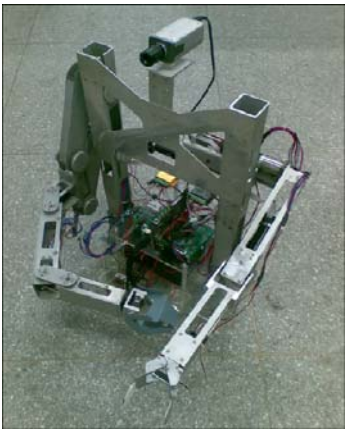


Fig. 10. The nursing robot prototype

3. Finally, the manipulator coordinate system and the camera coordinate system should be established transformation relation through a fixed point on the robot platform, and each joint angle can be obtained through solving the inverse kinematics equations. Thus, the preliminary visual servo control of manipulator has been achieved, and the whole working procedure architecture of the nursing robot is shown in Fig. 9.

5 Conclusions and Future Work

Based on prophase research achievements of the service robot prototype for the elderly and the disabled, we have further developed a robotic manipulator system applicable to nursing robot, and this paper mainly introduces its mechanical structure, control system and vision scheme. Currently, the robot prototype is shown in Fig. 10. It is an autonomous mobile platform with double manipulators. The left manipulator is applied to accomplishing the functions of grasping water cup and pouring water, etc; while the right manipulator is designed as a parallel linkage mechanism with the characteristic of spatial decoupling, and more complex functions can be realized through the dual-arm interaction. The mobile platform of the service robot utilizes differential drive system, and it includes two independent universal wheels and two wheels driven separately by two servo motors. The arrangement of wheels is in diamond shape, which means that both lines connecting two drive wheels and two driven wheels are diagonals of the diamond with intersection at the center point of the round mobile platform, thus can enable the robot to rotate around its own axis so as to enlarge workspace of the manipulator. The future work will mainly lie in improving the robot intelligent, so that it can accomplish service tasks under more complex indoor environment.

Acknowledgements. This work was supported by the National High Technology Research and Development Program of China (No.2006AA040202), the Hexa-type Elites Peak Program of Jiangsu Province (No.2008144) and the Natural Science Foundation of Jiangsu Province (No.BK2007514).

References

1. Jardon, A., Gimenez, A., Correal, R., Cabas, R., Martinez, S., Balaguer, C.: A Portable Light-weight Climbing Robot for Personal Assistance Applications. Industrial Robot: An International Journal 33(4), 303–307 (2006)
2. Bostelman, R., Albus, J., Chang, T., Hong, T., Agrawal, S.K., Ryu, J.C.: HLPR Chair: A Novel Indoor Mobility-assist and Lift System. In: Proceedings of International Design Engineering Technical Conference & Computers and Information in Engineering Conference, Las Vegas, Nevada, USA, pp. 663–670 (2007)
3. Medioni, G., Francois, A.R.J., Siddiqui, M., Kim, K., Yoon, H.: Robust Real-time Vision for a Personal Service Robot. Computer Vision and Image Understanding 108, 196–203 (2007)
4. Bellotto, N., Hu, H.: Multimodal Perception and Recognition of Humans with a Mobile Service Robot. In: Proceedings of the 2008 IEEE International Conference on Robotics and Biomimetics, Bangkok, Thailand, pp. 401–406 (2009)
5. Buiu, C.: Design and Development of a Waste Cleanup Service Robot. In: Proceedings of the First International EUROBOT Conference, Heidelberg, Germany, pp. 194–202 (2008)
6. Li, X.S., Kong, L.F., Zhao, F.D.: An Approach for Elderly Daily Behavior Habits Extraction for Service Robot Using Position Information. In: Proceedings of the 7th World Congress on Intelligent Control and Automation, Chongqing, China, pp. 1619–1623 (2008)
7. Kim, M., Kim, S., Park, S., Choi, M.T., Kim, M., Gomaa, H.: Service Robot for the Elderly - Software Development with the COMET/UML Method. IEEE Robotics & Automation Magazine 16(1), 34–45 (2009)
8. Shi, J.F., Zheng, J.Y., Zhang, Z.S., Zhang, L., Lu, G.: Implementation of a Multijoint Service Robot Manipulator with Z-direction Translational Decoupling. China: 200810021607.2 (August 05, 2008) (in Chinese)
9. Zhang, Z.H., Zheng, J.Y., Liu, Y., Lu, G.: Development of a Movable Service Robot with Double Manipulators for the Elderly and the Disabled. In: Proceedings of 2009 International Conference on Electronic Computer Technology, Macau, China, pp. 636–640 (2009)
10. Hu, Z.J., Dai, M., Zhang, L., Zheng, J.Y., Lu, G., Zhang, Z.S.: Overall Design of a Vision-guided Robot for Elderly People. Mechanical & Electrical Engineering Magazine 26(4), 1–4 (2009) (in Chinese)

On the Design of Exoskeleton Rehabilitation Robot with Ergonomic Shoulder Actuation Mechanism

Wenbin Chen[1,*], Caihua Xiong[1], Ronglei Sun[1], and Xiaolin Huang[2]

[1] State Key Lab of Digital Manufacturing Equipment and Technology
Institute of Rehabilitation and Medical Robotics
Huazhong University of Science and Technology, Wuhan 430074, China
[2] Department of Rehabilitation Medicine, Tongji Hospital
Tongji Medical College of Huazhong University of Science and Technology,
Wuhan 430030, China
wbchen@smail.hust.edu.cn

Abstract. Due to the advantages of more intensiveness, long duration, repeatability and task-orientation, robot-assistant training has become a promising technology in stroke rehabilitation. Comparing to the end-effector guided robots, exoskeleton robots provide better guidance on the posture of upper extremity, especially during movements with large ranges. Regarding the upper extremity, the natural coordination called shoulder rhythm is the most challenge to the ergonomic design of shoulder exoskeleton. Based on kinematic analysis of movement of shoulder complex, a nine degree-of-freedom exoskeleton rehabilitation with six degree-of-freedom shoulder actuation mechanism is proposed. In order to verify the manipulability of the proposed robot during assisting patient with performing activities of daily living (ADLs), the performance criteria, i.e., dexterity measure and manipulability ellipsoid, are used to evaluate and compare with human upper extremity. The evaluated result confirms the ergonomic design of shoulder mechanism of the rehabilitation robot on providing approximate dexterity matching that of human upper extremity in ADLs.

Keywords: Exoskeleton rehabilitation robot, upper extremity, ergonomic design, matching dexterity.

1 Introduction

Among the patients admitted to hospital for neurologic diseases, the most of survivors who suffer from serious neurologic disorders are evoked by stroke. Nowadays the stroke is the leading cause of disability[1]. There is a strong belief that early mobilizations, such as passive stretching and moving the joints through a full range of motion, is beneficial to prevent joints contracture and improve the state of the illness

* To whom correspondence should be addressed.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 1097–1110, 2009.

[1]. In order to recover motor function, the therapy for rehabilitation should be long duration, intensive, repetitive and task-oriented[2].

Obviously, the manually assisted movement training has many limitations to meet these criteria. First the therapists are likely to be fatigue during the labor-intensive training. The training duration would be terminated in advanced, and then the training sessions, are often shorter than required to gain an optimal therapeutic outcome. The one-to-one manually training in medical center is also expensive in the sense of long-term session. Finally, manually assisted training can not guarantee the quality of repetitive movement and it is difficult to give the quantitative evaluation of functional performance and progress of patient outcomes. In contrast, the robot assisted movement training is immune from these problems. Rehabilitation therapy with robot can give more accurate and effective stimulation to the neuromuscular system for promoting the progress of neuro-plasticity. So far, many rehabilitation robots have been developed[3] [4] [2, 5, 6]. The end-effector based robots have been widely studied [7-9]. Due to needless of rotation axis alignment between robot joints and human joints, this type of rehabilitation robots are easier to realize. Unlike the end-effector based robots, the exoskeleton robots [2, 3, 6, 10-12], should have to ensure their technical rotation axis in consistent with human rotation axis for safe and comfortable training. This requirement brings difficult on designing adaptable structure to different body size. However, because the exoskeleton therapy robot can entirely determine the posture of upper extremity, it is more suitable for task-oriented training. Moreover, it is possible to separately control the torque applied to each joint of the upper extremity and prevent the possibility of hyperextension by mechanical stops. The possibility to control the interacted reaction force and torque between shoulder joint and elbow joint is essential, especial for patient with glenohumeral subluxation and flaccid paralysis.

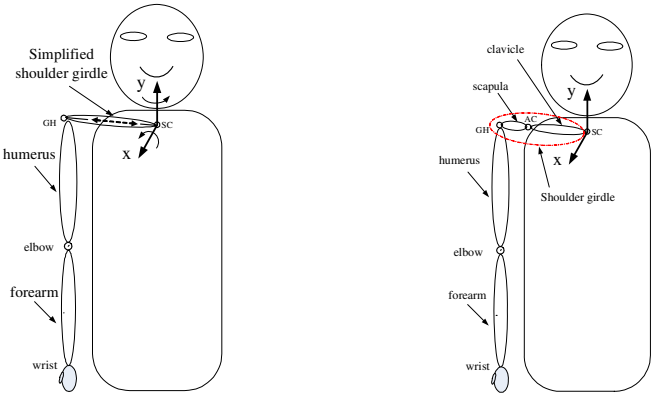
On implementation of exoskeleton rehabilitation robot, the most involved work is to deal with the shoulder mechanical structure to ensure automatically match the center of glenohumeral joint (CGH) which has no fixed position. Several groups presented their simplified design of shoulder mechanism. The work in [6] proposed reduced DoF design of shoulder joint for specified training movement. The principle of three orthogonal axes crossing one point was used to design the shoulder joint for realizing ball-socket movement[10] [3, 6, 13]. These designs assume that CGH is relatively fixed in the shoulder region. The shoulder girdle rhythm was not considered during humerus elevation[14], therefore, the harmful pressure resulted from interference and misalignment between the robot link and human bones will emerge. The ARMin-II in [2] introduced one vertical translation DoF in their shoulder mechanism, however, this improvement is just approximation of the movement of CGH, because the real trajectory CGH in frontal plane is not exactly linear. Carignan et al [15] proposed a shoulder rehabilitation exoskeleton which considered the scapula motion. The shoulder girdle translation relatively to the thorax was approximated by a single rotary joint with front-to-back axis. This design can not give enough flexibility to realize shoulder rhythm. Since shoulder girdle movement is significantly dependent on rotation of sternoclavicular (SC) joint, it is seems useful to provide the shoulder mechanism with some additional DoFs to match the SC joint. To address this issue, two different exoskeleton rehabilitation mechanisms developed in [11, 16] have divided the shoulder

mechanism into two parts: inner part and outer part. The inner structure corresponding to the SC joint has 2 DoFs whose crossing point approximate to the SC joint, and the outer structure corresponding to the GH joint has 3 DoFs whose crossing point approximate to the CGH. However, according to the kinemactic analysis of shoulder complex [14], the length of shoulder girdle is variable during the humerus elevation. The fixed link length between inner structure and outer structure can not match the mobility of human limb. Without actively match the movement of shoulder girdle, it is not possible to prevent interference between the robot link and human bones, nor is there a means to properly regain strength and coordination of the shoulder girdle. The purpose of this study is to analyze the motion of human shoulder girdle and to present an ergonomic design of rehabilitation robot with shoulder mechanism to match GH joint motion as much as possible.

2 Method

2.1 Ergonomic Design of Shoulder Actuation Mechanism

The large range of motion of upper extremity significantly roots from the flexibility of shoulder complex. During humeral elevation, the clavicle rotates at the SC joint, the scapula rotates at the acromioclavicular joint and slides on scapulothoracic joint, tilting the glenoid fossa upwards. It is a complex movement with no fixed centre of rotation [14] [17]. In fact, if the shoulder girdle has no motion ability, the humerus cannot be elevated more than 120 degree.



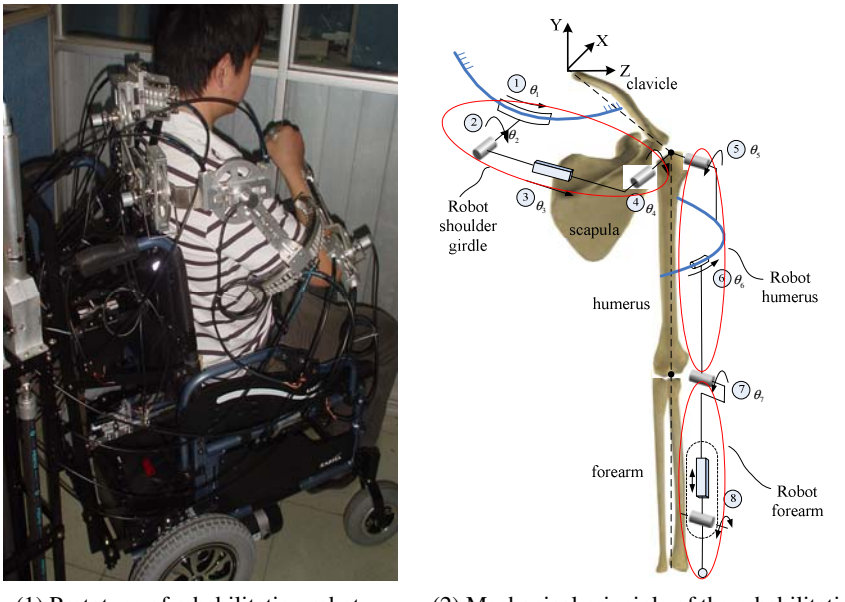
(1) Simplified shoulder girdle (2) Anatomical shoulder girdle of human

Fig. 1. Two definitions of the kinematic chain of human shoulder girdle. The global reference coordinate system of two kinematic chain parallels to the thorax coordinate system[18] but locates in marker SC.

In current literature, there is no shoulder rehabilitation mechanism considering the length change of the shoulder girdle. Most of the current devices were less than 5 DoFs [2, 15, 16] in the shoulder mechanism with the objective that used the lowest number of DoFs possible to allow for full exercise therapy of the shoulder complex[15]. However, this simplification gives birth to the non-ergonomic interface with human.

In this paper, the shoulder girdle of human is simplified as a variable-length link with 2 rotary DoF at the SC joint and 3 rotary DoF at the GH joint. This simplified strategy is similar with the presentation in [19]. The figure 1 gives the sketch of simplified shoulder girdle and un-simplified shoulder girdle of human upper extremity.

According to the simplified shoulder girdle of human, the exoskeleton shoulder rehabilitation mechanism should not only provide the rotation of elevation/depression and protraction/retraction but also have the additional DoF to match the length variation of shoulder girdle. Under this requirement, a 9-DoFs exoskeleton rehabilitation robot with 6-DoFs shoulder mechanism has been developed. The robot prototype and mechanism principle are showed in figure 2.



(1) Prototype of rehabilitation robot

(2) Mechanical principle of the rehabilitation robot

Fig. 2. A 9-DoFs exoskeleton rehabilitation robot with 6-DoFs shoulder mechanism

In the proposed rehabilitation robot, two arc guides (① and ⑥) were used to implement the rotational movement around axis which is internal embedded in the human body, i.e. protraction/retraction of SC joint and internal/external rotation of humerus. The rotational DoFs (and ②) are used to assist the shoulder girdle for elevation/depression and protraction/retraction, the translational DoF corresponds to the length change of simplified shoulder girdle, and the DoFs (, and ⑥) are use to implement the sphere movement of GH joint. In the elbow joint, two passive joints

(one translation plus one rotation) are designed to compensate the misalignment between exoskeleton elbow joint and human elbow joint during elbow flexion [11]. Obviously, when the patient were taken for training, the CGH of human upper extremity can be automatically matched with the rotational center of joints , and ⑥ via the coordination of joints , ②and .The intersection of the axes of joint ① and joint ② is assumed to be coincident with incisura jugularis of patient during rehabilitation training. In this proposed prototype of rehabilitation robot, all active joint are actuated by pneumatic muscles through housing cables. In order to realize bidirectional movement, each active joint is actuated by dual pneumatic muscles. The separately driving unit which includes pneumatic muscles, housing cables, and automatic cable tension mechanism is not provided due to the limited space in this paper. The mechanism is supported by external carriage to eliminate the pressure on human body.

2.2 Movement Relationship between Human Joint and Robot Joint

Due to the inevitable vertical and horizontal misalignments between robot elbow joint and human elbow joint, the rotation angle of robot joint is minor different from that of human elbow joint. The general relative position between robot forearm and human forearm can be illustrated in figure 3.

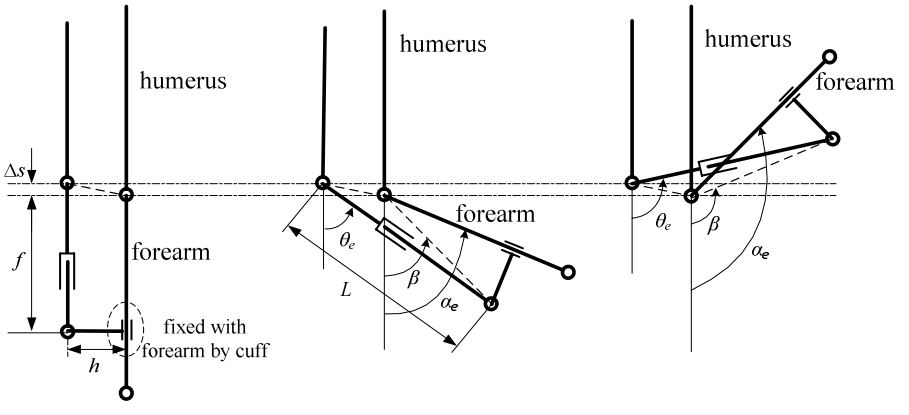


Fig. 3. Three configurations of the robot elbow joint during human elbow flexion

The relationship of flexion/extension angle between robot elbow θ_e and human elbow α_e is given by:

$$\theta_e = \begin{cases} \pi - \lambda - \cos^{-1}\left(\dfrac{\Delta s^2 + L^2 - f^2}{2L\sqrt{\Delta s^2 + h^2}}\right), & \beta < \pi - \lambda \\ \pi - \lambda \qquad\qquad\qquad\qquad , & \beta = \pi - \lambda \\ \pi - \lambda + \cos^{-1}\left(\dfrac{\Delta s^2 + L^2 - f^2}{2L\sqrt{\Delta s^2 + h^2}}\right), & \beta > \pi - \lambda \end{cases} \tag{1}$$

$$L = \sqrt{\Delta s^2 + 2h^2 + f^2 - 2\sqrt{(\Delta s^2 + h^2)\cdot(h^2 + f^2)}\cos(\lambda + \beta)} \tag{2}$$

$$\lambda = \frac{\pi}{2} + \tan^{-1}(\frac{\Delta s}{h}) \tag{3}$$

$$\beta = \alpha_e - \tan^{-1}(\frac{h}{f}) \tag{4}$$

where Δs is the center displacement between human elbow joint and robot elbow joint along the vertical direction, and f is the initial length of robot forearm, and h is the horizontal displacement between robot forearm and human forearm at the initial posture. The passive translation and rotation of joint , if denoted separately by $\Delta\chi$ and $\Delta\theta_8$, can be written by:

$$\Delta\chi = L - f \tag{5}$$

$$\Delta\theta_8 = \begin{cases} \frac{\pi}{2} - \tan^{-1}(\frac{f}{h}) - \cos^{-1}(\frac{L^2 + f^2 - \Delta s^2}{2L\sqrt{f^2 + h^2}}), & \beta < \pi - \lambda \\ \frac{\pi}{2} - \tan^{-1}(\frac{f}{h}) & , \beta = \pi - \lambda \\ \frac{\pi}{2} - \tan^{-1}(\frac{f}{h}) + \cos^{-1}(\frac{L^2 + f^2 - \Delta s^2}{2L\sqrt{f^2 + h^2}}), & \beta > \pi - \lambda \end{cases} \tag{6}$$

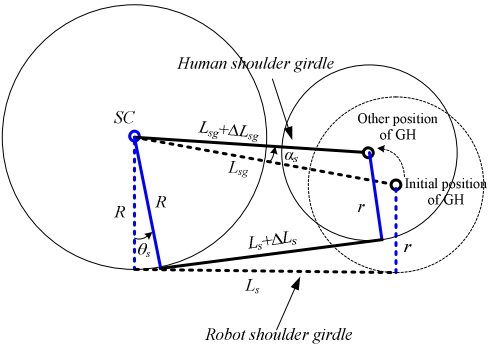


Fig. 4. The movement of human shoulder girdle and robot shoulder girdle in horizontal plane

The protraction/retraction between robot shoulder girdle and human shoulder girdle is minor different due to the inequality between radius of arc guider and distance from CGH to the axis of prismatic joint . Because distance from CGH to the axis of prismatic joint is relatively constant (this distance is denoted to be *r*), in horizontal

plane, it could be considered that the axis of prismatic joint is co-tangent to arc guider (the radius is denoted to be *R*) and the circle whose center is CGH .

As showed in figure 4, L_{sg} and L_s are the initial length of human shoulder girdle and robot shoulder girdle separately, ΔL_s and ΔL_{sg} are the variation of L_s and L_{sg} separately. The relationship of protraction /retraction angle between robot shoulder girdle θ_s and human shoulder girdle α_s is given by:

$$\theta_s = \phi + \alpha_s - \tan^{-1}(\frac{L_s + \Delta L_s}{R - r}) \tag{7}$$

$$\phi = \tan^{-1}(\frac{L_s}{R - r}) \tag{8}$$

$$L_s + \Delta L_s = \sqrt{(L_{sg} + \Delta L_{sg})^2 - (R - r)^2} \tag{9}$$

The elevation/depression angle of robot shoulder girdle is similar to the angles of human shoulder girdle because the two shoulder girdles are coincident in front plane. The three Euler-angles of GH joint of the robot are similar to Euler-angles of human GH joint.

3 Kinematic Analysis of the Proposed Rehabilitation Robot

Although the proposed prototype of rehabilitation robot has the potential to assist human by performing ADLs, the ability of matching dexterity of human upper extremity during assistance is essentially important and needed to verify. Here we utilize two performance criteria to evaluate the robot dexterity, i.e., dexterity measure and manipulability ellipsoid. In order to compare the kinematic performance of the rehabilitation robot with the human upper extremity, the kinematic models of the rehabilitation robot and human upper extremity are introduced in following sections.

3.1 Kinematic Models of the Proposed Rehabilitation Robot and Human Upper Extremity

The formula of exponential product was used to model the robot and human upper extremity. In these two kinematic models, the wrist was considered as the end-effector. The local anatomical coordinate system of human upper extremity follows the proposal of ISB [18]. As showed in figure 1, the configuration of the end-effector in the reference frame is:

$$\mathbf{g} = e^{\hat{\xi}_1\theta_1} \cdots e^{\hat{\xi}_n\theta_n}\mathbf{g}_0 \tag{10}$$

where $\mathbf{g}_0 \in \mathbf{SE}(3)$ is the initial configuration of end-effector in the reference frame. $\xi_i \in \Re^{6\times 1}$ ($i = 1,\cdots,7$ for robot and $i = 1,\cdots,10$ for human upper extremity.) is the

rotation twist corresponding to each joint. For human upper extremity, the twist set (ξ_1 , ξ_2 , ξ_3)corresponds to the SC joint in Y-X-Z order, twist set (ξ_4, ξ_5, ξ_6)corresponds to the AC joint in Y-X-Z order of joints , twist set (ξ_7, ξ_8, ξ_9) corresponds to the GH joint in Y-X-Z order, and twist set (ξ_{10}, ξ_{11}) corresponds to the flexion/extension and pronation/supnation of elbow joint. The table 1 lists the needed Kinematic parameters of robot and upper extremity.

Table 1. Kinematic parameters of robot and upper extremity

Robot		Human upper extremity	
Initial length of robot shoulder girdle	L_{sg}	Clavicle length	L_c
Length of robot humerus	L_{rh}	Scapula length	L_g
Length of robot forearm	L_{rf}	Humerus length	L_h
Initial length of robot shoulder girdle	L_{sg}	Forearm length	L_f
$\mathbf{g}_0$	$\begin{bmatrix} & & 0 \\ & \mathbf{E}^* & -L_{rh}-L_{rf} \\ & & L_{sg} \\ 0\,0\,0 & & 1 \end{bmatrix}$	$\mathbf{g}_0$	$\begin{bmatrix} & & 0 \\ & \mathbf{E}^* & -L_h-L_f \\ & & L_c+L_g \\ 0\,0\,0 & & 1 \end{bmatrix}$
ξ_1 ,	$(0\ 0\ 0\ 0\ 1\ 0)^T$	ξ_1	$(0\ 0\ 0\ 0\ 1\ 0)^T$
ξ_2	$(0\ 0\ 0\ -1\ 0\ 0)^T$	ξ_2	$(0\ 0\ 0\ 1\ 0\ 0)^T$
ξ_3	$(0\ 0\ 1\ 0\ 0\ 0)^T$	ξ_3	$(0\ 0\ 0\ 0\ 0\ 1)^T$
ξ_4	$(0\ 0\ -L_{sg}\ -1\ 0\ 0)^T$	ξ_4	$(-L_c\ 0\ 0\ 0\ 1\ 0)^T$
ξ_5	$(0\ 0\ 0\ 0\ 0\ 1)^T$	ξ_5	$(0\ L_c\ 0\ 1\ 0\ 0)^T$
ξ_6	$(-L_{sg}\ 0\ 0\ 0\ 1\ 0)^T$	ξ_6	$(0\ 0\ 0\ 0\ 0\ 1)^T$
ξ_7	$(-L_{rh}\ 0\ 0\ 0\ 0\ 1)^T$	ξ_7	$(-L_g-L_c\ 0\ 0\ 0\ 1\ 0)^T$
		ξ_8	$(0\ L_g+L_c\ 0\ 1\ 0\ 0)^T$
		ξ_9	$(0\ 0\ 0\ 0\ 0\ 1)^T$
* $\mathbf{E} \in \Re^{3\times3}$ is identity matrix.		ξ_{10}	$(-L_h\ 0\ 0\ 0\ 0\ 1)^T$
		ξ_{11}	$(-L_g-L_c\ 0\ 0\ 0\ 1\ 0)^T$

The Jacobian matrix $\mathbf{J}(\boldsymbol{\theta}) \in \Re^{6\times7}$ of the model (10) is represented by:

$$\mathbf{J}(\boldsymbol{\theta}) = \left[\xi_1 \quad \xi_2' \quad \cdots \quad \xi_n'\right] \tag{11}$$

$$\xi_i' = \mathbf{Ad}_{(e^{\hat{\xi}_1\theta_1}\cdots e^{\hat{\xi}_{i-1}\theta_{i-1}})}\xi_i \ , \ i = 2\cdots n \tag{12}$$

where $n=7$ for robot and $n=11$ for upper extremity. The matrix $\mathbf{Ad}_{(\mathbf{g})} \in \Re^{6\times 6}$ which transforms twist from one coordinate frame to another is known as adjoint transformation associated with $\mathbf{g} \in \mathbf{SE}(3)$ [20]. The linear space velocity and angular space velocity of end-effector, according to the definition in [20] , corresponds to the first three rows and the last three rows of $\mathbf{J}(\boldsymbol{\theta})$. Here we denote the linear velocity by $\mathbf{J}_v(\boldsymbol{\theta})$. The dexterity measure [21] corresponding to the linear velocity of end-effector is:

$$\omega = \sqrt{\det[\mathbf{J}_v(\boldsymbol{\theta})\mathbf{J}_v^{\ T}(\boldsymbol{\theta})]} \tag{13}$$

where ω is the dexterity value at the configuration $\boldsymbol{\theta}$. The value of the measure ω describes the dexterity of the robot at a certain configuration. The bigger the measure ω is, the better the motion ability for the robot at a certain configuration is. In order to compare the dexterity measure between upper extremity and robot, the dexterity measure is normalized by:

$$\bar{\omega} = \frac{\omega - \min(\omega)}{\max(\omega) - \min(\omega)} \tag{14}$$

However, the dexterity measure is only a scalar index corresponding to certain configuration of the robot; the motion ability of the end-effector along different directions at current configuration can not be revealed directly. Fortunately, we can make use of the equation $\dot{\boldsymbol{\theta}}^T\dot{\boldsymbol{\theta}} = 1$ to define a supersphere in the joint space. The supersphere will be mapped into an ellipsoid in the manipulation space. The ellipsoid is referred to as manipulability ellipsoid which is similar to the force superellipsoid defined by Xiong [22]. The equation of the manipulability ellipsoid corresponding to the linear velocity is defined by

$$\mathbf{v}^T(\mathbf{J}_v\mathbf{J}_v^{\ T})^{-1}\mathbf{v} = 1 \tag{15}$$

where the vector $\mathbf{v} \in \Re^{3\times 1}$ is end-effector's linear space velocity following the definition in [20]. The volume of the manipulability ellipsoid has been proved proportional to the dexterity measure in [22]. The direction along longest principle of the manipulability ellipsoid is the most dexterous motion direction for the robot at a certain configuration. If the manipulability ellipsoid approximates to a sphere, then the motion ability along all direction is almost same and the corresponding configuration of the robot is referred to as the optimal kinematic configuration.

3.2 Kinematic Performance Evaluation

A healthy subject with height $h = 175cm$ was employed as the reference to determinate the size of the robot structure. Moreover, other necessary assumptions which simplify the movement of different subjects were needed and listed as:

1). The movement of shoulder girdle is independent of the internal/external rotation of humerus
2). The glenohumeral joint movement is similar between different individuals who have the same body size.
3). Assume that the movement of shoulder girdle is spontaneous and without conscious control during humeus elevation.

The initial length of robot shoulder girdle L_{sg} is $195mm$, the length of robot humerus L_{rh} is $334mm$, the length of robot forearm L_{rf} is $231mm$ and the other two dimensions relatively to the robot shoulder girdle are $R = 150mm$ and $r = 90mm$. The dimensions of misalignment between robot elbow and human elbow are given by: $\Delta s = 14mm$, $h = 6mm$. In this section, two representative activities highly frequently involved in daily living, i.e., combing hair and touching contra lateral shoulder, are used to evaluate the motion ability of the robot. The natural movements of each anatomical joint of upper extremity were recorded by the motion capture system (VICON, Oxford Metrics Ltd, UK). ALL the bony local coordinate systems follow the ISB recommendation. The angle mapping from human upper extremity to robot is given by equations(1), (6) and(7). The normalized dexterity measure (13) and manipulability ellipsoids (15) of the proposed rehabilitation robot and human upper extremity are given by figure 5 and figure 6. In these two figures, the robot is represented as simplified mechanism for representation simplification when visualizing the human postures.

As showed in figure 5 and figure 6, four representative stages were marked out by red color point and the corresponding manipulability ellipsoid of upper extremity and robot are given. The dexterity measure of the proposed rehabilitation robot is well matched to that of upper extremity before upper extremity reaching the optimal dexterity posture, i.e. stage 3. In this duration, the motion ability of the robot in space is approximately consistent with upper extremity. The configuration of the robot at upper extremity's optimal configuration, i.e., stage 3, provides the robot with approximate dexterity with that of upper extremity. Relatively large deviation between robot and upper extremity, especially at the end of tasks, i.e., stage 4, is found after the upper extremity leaving the optimal configuration. However, the motion ability of robot still have relatively better motion ability along other direction which are indicated by the less oblateness of manipulability ellipsoids. During the period from stage 3 to stage 4, the dexterity of the proposed robot is smaller than that of human upper extremity. This dexterity deviation may result from the simplified structure of shoulder girdle of the robot and one DoF less than the human elbow joint. At the optimal configuration of upper extremity, the most dextrous direction of robot is along direction from left upper side to right lower side with respect to the tested subject, and this most dexterous direction is consistent with the movement of human upper extremity in the duration of the tasks.

4 Discussion and Conclusion

To our knowledge, matching the dexterity of upper extremity as much as possible in ADLs is essentially significant and basic requirement on designing ergonomic rehabilitation robot with the characteristics of safety and comfort. The proposed evaluation method in this paper gives insight into the dexterity distribution of upper extremity in ADLs. By comparing the dexterity of rehabilitation robot with that of human upper extremity, this paper provides an alternatively effective method for evaluating the design of ergonomic rehabilitation robot. The proposed rehabilitation robot for upper extremity therapy is verified to be approximately ergonomic during assisting upper extremity with performing ADLs.

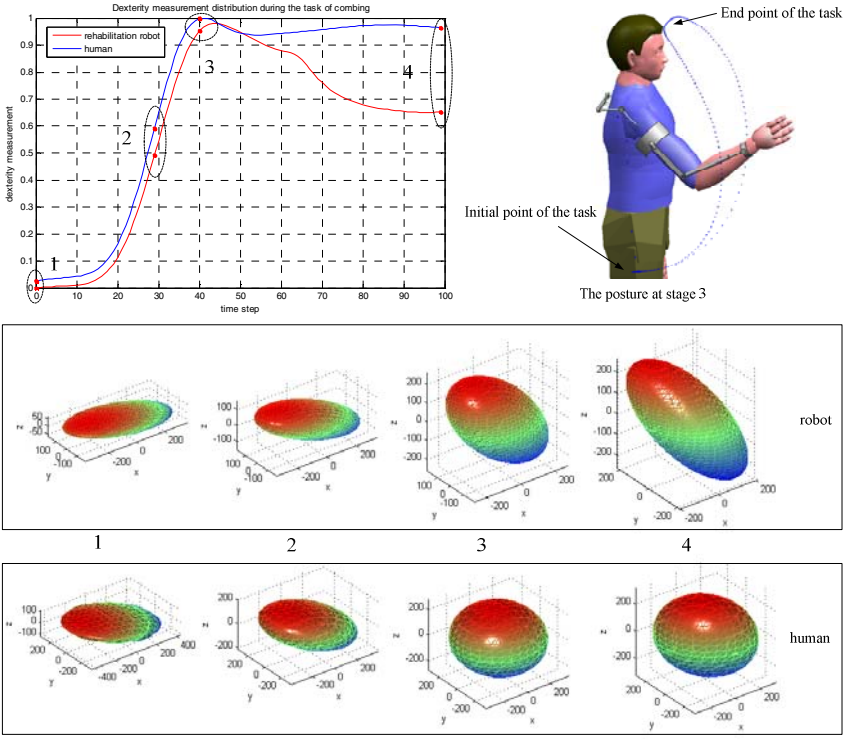


Fig. 5. The normalized dexterity measure and manipulability ellipsoid of robot assisting subject with combing hair

In conclusion, the main results obtained in this paper are as follows:

1) An exoskeleton rehabilitation robot based on the analysis of shoulder complex is proposed which consists of seven active joints and two passive joints. The shoulder girdle rhythm of human is considered and ergonomically implemented by two rotary joint and one prismatic joint allowing a self adaptation to the shoulder girdle movement.

2) A kinematical model of human upper extremity is represented. The kinematic performance of the robot is evaluated and compared with the human upper extremity. The mapping relationship of joint angles from upper extremity to the robot is established which enable employ the kinematic data from human limb to drive the robot for assisted therapy.

3) The suitability of the proposed robot to the acute phase rehabilitation is confirmed by investigating the ability of robot on dexterity matching with human upper extremity in ADLs.

We are now conducting to improve the mechanical design for more suitable therapy in clinic environment. The inverse kinematics and motion planning of the robot also will be carried out to reduce the dependency on the motion capture system for realizing human-like trajectory.

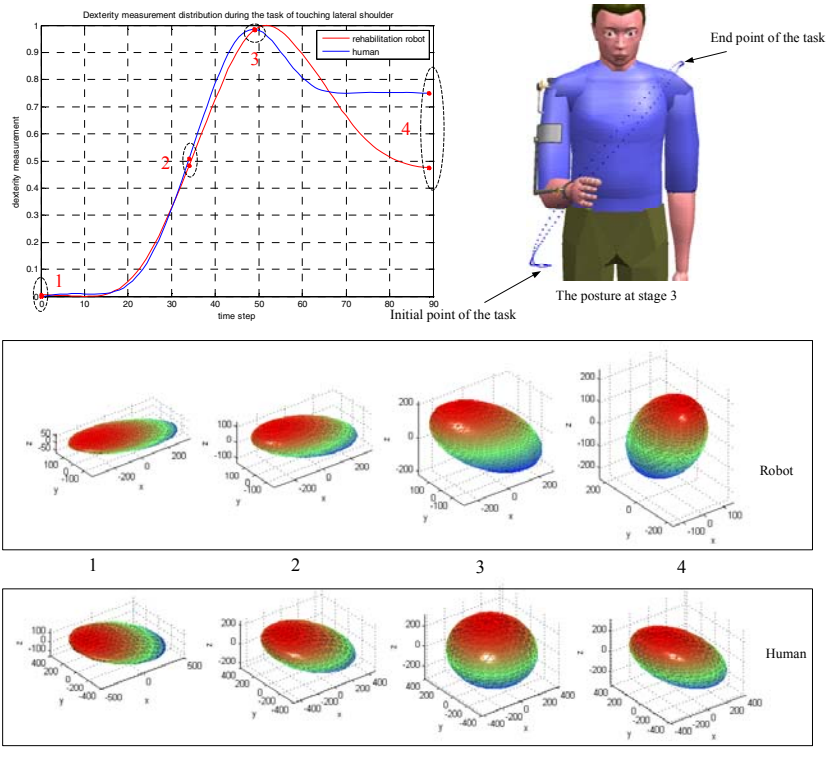


Fig. 6. The normalized dexterity measure and manipulability ellipsoid during robot assisting subject with touching lateral shoulder

Acknowledgments. This work was supported by the National Natural Science Foundation of China under Grant 50775080 and 50875100, the "863" Hi-Tech Research and Development Program of China under Grant 2007AA04Z204, and the National Key Technology R&D Program of China under Grant 2008BAI50B04.

References

1. DeLisa, J.A., Gans, B.M., Walsh, N.E.: Physical medicine and rehabilitation: principles and practice.Lippincott Williams & Wilkins (2004)
2. Nef, T., Riener, R.: Shoulder actuation mechanisms for arm rehabilitation exoskeletons. In: 2nd IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics, Scottsdale, AZ, USA, pp. 862–868 (2008)
3. Perry, J.C., Rosen, J., Burns, S.: Upper-Limb Powered Exoskeleton Design. IEEE/ASME Transactions on Mechatronics 12(4), 408–417 (2007)
4. Prange, G.B., et al.: Systematic review of the effect of robot-aided therapy on recovery of the hemiparetic arm after stroke. Journal of Rehabilitation Research and Development 43(2), 171 (2006)
5. Carignan, C., Liszka, M., Roderick, S.: Design of an arm exoskeleton with scapula motion for shoulder rehabilitation. In: 12th International Conference on Advanced Robotics, ICAR 2005, Proceedings, pp. 524–531 (2005)
6. Sugar, T.G., et al.: Design and Control of RUPERT: A Device for Robotic Upper Extremity Repetitive Therapy. IEEE Transactions on Neural Systems and Rehabilitation Engineering 15(3), 336–346 (2007)
7. Rosati, G., Gallina, P., Masiero, S.: Design, Implementation and Clinical Tests of a Wire-Based Robot for Neurorehabilitation. IEEE Transactions on Neural Systems and Rehabilitation Engineering 15(4), 560–569 (2007)
8. Lum, P.S., Burgar, C.G., Shor, P.C.: Evidence for improved muscle activation patterns after retraining of reaching movements with the MIME robotic system in subjects with post-stroke hemiparesis. IEEE Transactions on Neural Systems and Rehabilitation Engineering 12(2), 186–194 (2004)
9. Jackson, A., et al.: Initial patient testing of iPAM-a robotic system for Stroke rehabilitation. In: IEEE 10th International Conference on Rehabilitation Robotics, pp. 250–256 (2007)
10. Stienen, A.H.A., et al.: Dampace: dynamic force-coordination trainer for the upper extremities. In: IEEE 10th International Conference on Rehabilitation Robotics, ICORR 2007, pp. 820–826 (2007)
11. Schiele, A., Helm, F.C.T.v.d.: Design to Improve Ergonomics in Human Machine Interaction. IEEE Transactions on Neural Systems and Rehabilitation Engineering 14(4), 456–469 (2006)
12. Riener, R., et al.: Human-centered rehabilitation robotics. In: 9th International Conference on Rehabilitation Robotics, Chicago, USA, pp. 319–322 (2005)
13. Sanchez, R.J., et al.: Automating Arm Movement Training Following Severe Stroke: Functional Exercises With Quantitative Feedback in a Gravity-Reduced Environment. IEEE Transactions on Neural Systems and Rehabilitation Engineering 14(3), 378–389 (2006)
14. Klopcar, N., Lenarcic, J.: Bilateral and unilateral shoulder girdle kinematics during humeral elevation. Clinical Biomechanics 21(Supplement 1), S20–S26 (2006)
15. Carignan, C., Liszka, M., Roderick, S.: Design of an arm exoskeleton with scapula motion for shoulder rehabilitation. In: Proceedings of 12th International Conference on Advanced Robotics, Washington, pp. 524–531 (2005)
16. Ball, S.J., Brown, I.E., Scott, S.H.: MEDARM: a rehabilitation robot with 5DOF at the shoulder complex. In: IEEE/ASME international conference on Advanced intelligent mechatronics, Switzerland, pp. 1–6 (2007)
17. Doorenbosch, C.A., et al.: Determination of functional rotation axes during elevation of the shoulder complex. The Journal of orthopaedic and sports physical therapy 31(3), 133–137 (2001)

18. Wu, G., et al.: ISB recommendation on definitions of joint coordinate systems of various joints for the reporting of human joint motion–Part II: shoulder, elbow, wrist and hand. Journal of Biomechanics 38(5), 981–992 (2005)
19. Klopcar, N., Tomsic, M., Lenarcic, J.: A kinematic model of the shoulder complex to evaluate the arm-reachable workspace. Journal of Biomechanics 40(1), 86–91 (2007)
20. Murray, R.M., Li, Z., Sastry, S.S.: A Mathematical Introduction to Robotic Manipulation. CRC Press, Inc., Florida (1994)
21. Yoshikawa, T.: Manipulability of Robotic Mechanisms. The International Journal of Robotics Research 4(2), 3–9 (1985)
22. Xiong, C.H., Ding, H., Xiong, Y.L.: Fundamentals of Robotic Grasping and Fxturing. World Scientific Publishing Co. Pte. Ltd, Singapore (2007)

A Novel Five Wheeled Rover for All Terrain Navigation

Arun Kumar Singh, Arun H. Patil, and Anup Kumar Saha

Mechanical Engineering Department National Institute of Technology, Durgapur, India
aks1812@gmail.com, arun.h.patil@gmail.com,
saha_ak2001@yahoo.co.in

Abstract. This paper presents a new concept for rough terrain navigation of rovers. The proposed design has reduced number of joints and links from existing suspension concepts. The suspension mechanism is derived from planar four bar mechanism and hence we present the singularity and trajectory analysis of the proposed mechanism. We derive the quasi-static equations of motion of the rover and a linear programming based approach is proposed for the optimum wheel motor torque control. Efficacy of the proposed mechanism is proved by simulations on undulating terrains as well as on terrains having discontinuity.

Keywords: Four bar mechanism, singularity, linear programming and dynamics.

1 Introduction

To design an effective suspension mechanism with minimum design and control complexity is the focus of the research here. Past research on wheeled all terrain vehicles has led to the development of two types of suspension mechanisms: active and passive. Passive suspension rovers adapt passively to the underlying by the virtue of contact forces and hence do not require any actuators for controlling the internal configuration of the vehicle thus significantly reducing the control architecture. Rocky7 [1] is one such vehicle which utilizes one of the simplest suspension mechanisms called rocker bogie. But the climbing ability and specially the lateral stability is limited as compared to shrimp[2,9] which utilizes a more sophisticated design derived from the four bar mechanism to enhance climbing ability. But as sophistication increases the number of joints and links also increases significantly increasing the overall complexity and weight of the system. In general joints are heavy parts and can easily lead to trouble in space environments [3]. Passive suspension rovers are usually multi wheel drive system [MWD] e.g. some rovers such as Lunokhod [4] and Marshakhod [5] have 6 or more wheels. Some other passive suspension mechanisms can be found in [10],[11] which are basically modifications of rocker bogie suspensions. Though the system has higher degree of mobility the system is intended to be heavier and hence not ideally applicable to medium to small scale rovers. Moreover the closed kinematic structures of passive suspension rovers pose additional constraint on the kinematic analysis and motion planning of the robot. Hence the aim of this paper is to significantly reduce the kinematic complexity by reducing the number of joints

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 1111–1122, 2009.

and closed loops but without compromising with the climbing ability of the rover. On the other hand active suspension rovers are kinematically simple but employ complex control architectures to command the actuators controlling the internal configuration of the rover. These control architectures critically depend on number of sensor inputs which can be easily corrupted by noise in uncertain environments. Owing to this reason the focus of the research here is concentrated on passive suspension with an aim to reduce the overall complexity of the design while maintaining the same obstacle climbing ability. We present the trajectory and singularity analysis of the mechanism proposed and a wheel motor torque control technique under quasi-static conditions. Efficacy of the proposed suspension is proven by extensive simulation on rough terrains.

2 Kinematic Model of the Rover

The kinematic mode of the rover is shown in figure1. It consists of two planar mechanisms on two sides of the chassis. The front wheel on both sides is connected to the end of the fork which is derived from the four bar mechanism. The common back leg is attached to the chassis through some compliance in the form of rotational joint passively controlled by torsion spring of high spring constant. This is one of the novelties of the design because unlike other designs such as [2] where back leg is directly connected by rigid joint, we introduce some compliance between the back leg and the wheel. This modification allows the rover to have some level of adaptability on uneven terrains even while moving backward. The other important novelty of the proposed design is the reduction in number of joints and links. The Planar view of the robot with its joints is shown in figure2. All designated joints are revolute with 5, 6, 7 being the actuated joint for driving the wheels and joint 4 being controlled passively by a torsion spring. The number of joints is 16 in total and number of bodies (excluding ground) is 15. A similar functioning design found in [2] employs 18 bodies and 22 joints. Joints are generally heavier and critical part because they have maximum tendency of failing and hence corresponding decrease of the number of joints in the

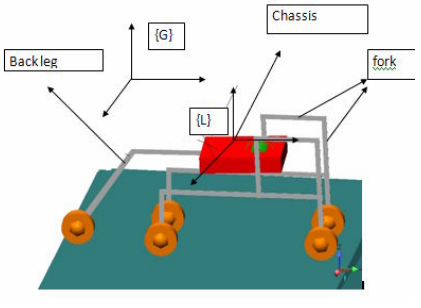


Fig. 1. Kinematic model of the rover

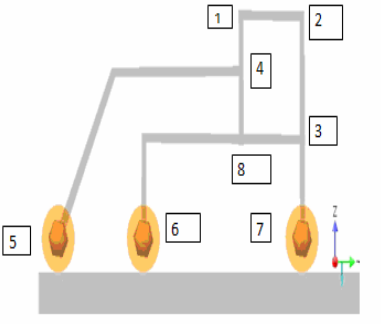


Fig. 2. Front view of the rover

proposed design not only reduces the overall weight of the system but also increases the reliability of the system.

3 Kinematics of the Fork

The fork with its geometrical parameters is shown in figure 4

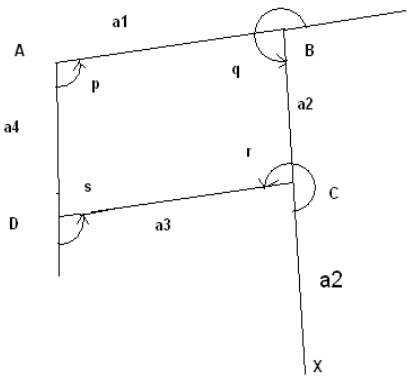


Fig. 3. Geometry of the Fork

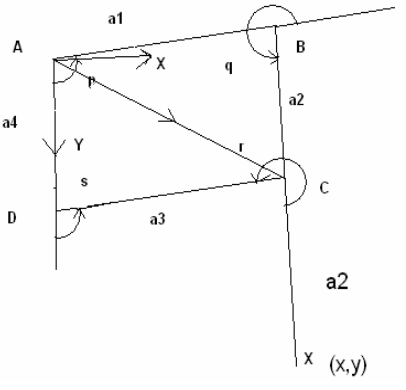


Fig. 4. Breaking the mechanism at point C

We are interested in finding out the trajectory of point X as the link AB is rotated to a particular angle. This simulates the situation encountered while the front fork is climbing the obstacle. We analyze the optimisable parameters from the trajectory plot of X. To generate the trajectory of point X considering joint A to be the input angle we derive the two loop closure equations for the four bar mechanism ABCD following the work of [6] and is briefly reviewed here. Consider the four bar mechanism to be broken at point C which will result in two planar 2R and 1R serial kinematic chain. This is shown in figure 4. Since the mechanism has been broken at point C passive variable C will not appear in the loop closure constraint equation. We have from figure 4

$$y = a_1\cos(p) + 2a_2\cos(p+q) \tag{1}$$

$$x = a_1\sin(p) + 2a_2\sin(p+q) \tag{2}$$

$$y = a_3\cos(s) + a_{4+}a_2\cos(p+q) \tag{3}$$

$$x = a_3\sin(s) + a_2\sin(p+q) \tag{4}$$

$$y = a_1\cos(p) + 2a_2\cos(p+q) = a_4 + a_3\cos(s) + a_2\cos(p+q) \tag{5}$$

$$x = a_1\sin(p) + 2a_2\sin(p+q) = a_3\sin(s) + a_2\sin(p+q) \tag{6}$$

from equation (5) and (6) we can eliminate the passive variables q, s and get x and y as a function of active joint variable p. We solve equation (5) and (6) using non linear

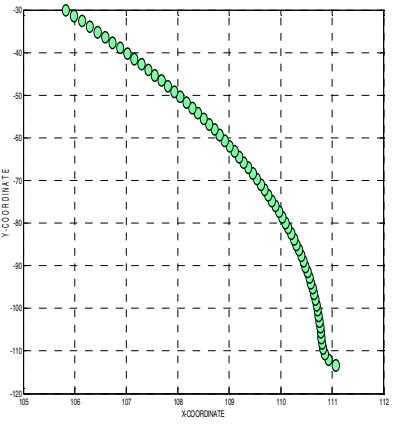


Fig. 5. Trajectory of the point X of the Fork

methods such as Newton's method. The deduced trajectory of point X is shown in figure 5.

From the trajectory of the fork we can see that for y co-ordinate varying from -110 to -90 there is very minimal variation in the x co-ordinate i.e. the slope during that part of the trajectory is almost 90^0. This signifies that for step like obstacles to the height of 20cm the fork will show a smooth climbing sequence with minimal loss of contact with the vertical surface of the step. For obstacles of height greater than 20cm the fork will show smoother climbing for lesser slope angles which will decrease with corresponding increase with obstacle diameter. Moreover the actual climbing ability of the fork will depend greatly on the overall dimension of the fork as well as of the whole rover.

4 Singularities of the Fork

A singularity analysis of the fork will enable us to make provision for its prevention by incorporating some mechanical stoppers at those positions to prevent the mechanism from reaching that state. To derive the point of singularity we follow the work of [6]. We have from closed loop ABCD of figure the following two loop closure constraint equations.

$$\eta_1 = a_1\cos(p) + a_2\cos(p+q) - a_4 - a_3\cos(s) = 0 \tag{7}$$

$$\eta_2 = a_1\sin(p) + a_2\sin(p+q) - a_3\sin(s) = 0 \tag{8}$$

In equation (7) and (8) we have q and s as passive variables and p as active variables. Differentiating both the equations with respect to time and writing them in matrix form yields

$$\begin{bmatrix} -a_1\sin(p) - a_2\sin(p+q) \\ a_1\cos(p) + a_2\cos(p+q) \end{bmatrix} (d(p)/dt) + \begin{bmatrix} -a_2\sin(p+q) + a_3\sin(s) \\ a_2\cos(p+q) \;\; -a_3\cos(s) \end{bmatrix} \begin{bmatrix} (dq/dt) \\ (ds/dt) \end{bmatrix} = 0 \tag{9}$$

Let $K = \begin{bmatrix} -a_1\sin(p) - a_2\sin(p+q) \\ a_1\cos(p) + a_2\cos(p+q) \end{bmatrix}$, $\overset{*}{K} = \begin{bmatrix} -a_2\sin(p+q) + a_3\sin(s) \\ a_2\cos(p+q) \;\; -a_3\cos(s) \end{bmatrix}$

The singular positions can be found by equation

$$\det[(\overset{*}{K})] = 0 \tag{10}$$

which reduces to

$$a_2\, a_3 \cos(s) \sin(p + q) - a_2\, a_3 \cos(p + q) \sin(s) = 0 \tag{11}$$

$$a_2 a_3 \sin(p + q - s) = 0 \tag{12}$$

$$\text{Or} \;\; p + q - s = n\pi, n = 1 \tag{13}$$

Also from figure 3 we have

$$p + q + r + (\pi - s) = 4\pi \tag{14}$$

$$\text{So} \;\; r = 2\pi \tag{15}$$

Which means that the link 2 and 3 are parallel or B,C,D are in straight line. This can be prevented by adding a stopper a joint C which will allow for the restricted rotation of joint.

5 Quasi-Static Analysis of the Rover

In this section we prepare a quasi-static model of the rover and solve the inverse problem where the orientation of the chassis and wheel ground contact force vectors are given and we are required to compute the feasible set of motor torques that can be applied to this system under equilibrium and no-slip constraints. In real life situations orientation of the chassis can be easily determined from sensors such as accelerometers and gyroscopes. At the wheel ground contact point two sets of forces are acting, normal and traction forces as shown in figure 6.

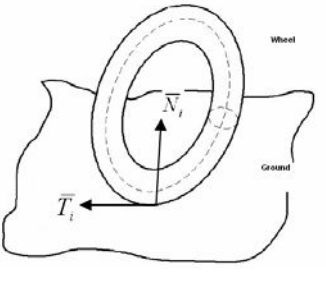


Fig. 6. Forces at the wheel ground contact point

We assume that the normal force unit vectors are known. This is a valid assumption if the rover is equipped with tactile wheels as proposed in [7]. The traction force can be calculated by the methodology proposed in [8]. Let Ψ, α and β be the pitch, roll and yaw angles respectively of the chassis about the global $\{G\}$ axes. The local and global frames ({L},{G}) are shown in figure1 respectively. The resultant rotation matrix relating position vectors and the force components at various points to the reference frame of the chassis is:

$$R = R_z(\beta)R_y(\alpha)R_x(\psi) \tag{16}$$

Where $R_x(\psi)$, $R_y(\alpha)$ and $R_z(\beta)$ are the rotation matrices corresponding to the Euler angles about the X, Y and Z axes respectively

From Fig.6 the forces acting at the i^{th} wheel of the vehicle from the chassis (Local frame {L}) reference frame are

i) The normal force $\overrightarrow{N_{iL}} = \begin{bmatrix} N_{xi} & N_{yi} & N_{zi} \end{bmatrix}^T$

ii) The traction force $\overrightarrow{T_{iL}} = \begin{bmatrix} T_{xi} & T_{yi} & T_{zi} \end{bmatrix}^T$

We assume that the motion of the vehicle to be non-holonomic in nature i.e the velocity of the vehicle along the direction perpendicular to the wheels is negligible. This is shown to be satisfied in the results and simulations section. And hence from the chassis reference frame we have

$$T_{xi} \approx 0 \text{ and } T_{yi} > 0 .$$

Under no slip conditions we have $\left|\vec{T}_{iL}\right| \le \mu_i \left|\vec{N}_{iL}\right| \quad \forall \quad i \in \{1,2,3,4\}$ (17)

Where μ_i is the coefficient of friction between the point of contact of i^{th} wheel and the terrain.

Since $\vec{T}_{iL}$ is always perpendicular to $\vec{N}_{iL}$ we have $dot(\vec{T}_{iL}, \vec{N}_{iL}) = 0$

$$\Rightarrow N_{xi}T_{xi} + N_{yi}T_{yi} + N_{zi}T_{zi} = 0 \tag{18}$$

From (17)

$$\left|\vec{T}_{iL}\right| = \eta\mu_i \left|\vec{N}_{iL}\right| \Rightarrow T_{xi}^2 + T_{yi}^2 + T_{zi}^2 = (\eta\mu_i)^2 N_{iL}^2 \tag{19}$$

Where $0 < \eta \le 1$

If $\left|\vec{N}_{iL}\right| \ne 0$ then any one of the components $N_{xi}, N_{yi}, N_{zi} \ne 0$, In general $N_{zi} \ne 0$,as long as the wheels remain in contact, then from (18) and (19) we get

$$(\eta\mu_i)^2 \left|\vec{N}_{iL}^2\right| - T_{xi}^2 = T_{yi}^2 + \left[\frac{N_{xi}T_{xi} + N_{yi}T_{yi}}{N_{zi}}\right]^2 \tag{20}$$

Simplifying the above we get in the form:

$$aT_{yi}^2 + bT_{yi} + c = 0 \tag{21}$$

Solving the above quadratic equation we get

$$T_{yi} = \frac{\eta\mu_i \left|\vec{N}_{iL}\right| \left|N_{zi}\right|}{\sqrt{N_{zi}^2 + N_{yi}^2}} \quad \text{and} \quad T_{zi} = \frac{-\eta\mu_i N_{yi} \left|\vec{N}_{iL}\right| \left|N_{zi}\right|}{N_{zi}\sqrt{N_{zi}^2 + N_{yi}^2}}$$

Now the unit vectors of $\vec{T}_{iL}$ and $\vec{N}_{iL}$ in the global reference frame are

$$\hat{t}_i = R\frac{\vec{T}_{iL}}{\left|\vec{T}_{iL}\right|} = [t_{Xi}\ t_{Yi}\ t_{Zi}]^T \tag{22}$$

and

$$\hat{n}_i = R\frac{\vec{N}_{iL}}{\left|\vec{N}_{iL}\right|} = [n_{Xi}\ n_{Yi}\ n_{Zi}]^T \tag{23}$$

Where $\vec{T}_i = R\vec{T}_{iL}$ and $\vec{N}_i = R\vec{N}_{iL}$ are the traction and normal forces at the point contact in the global reference frame.

To get the moment arm vectors we use the following procedure. We get the vector connecting the wheel ground contact point to the centre of the chassis by first considering a planar view of the mechanism and then multiplying the arm vector hence derived with the rotation matrix combining the roll and yaw angles. This is exemplified by figure8 which shows the planar view of the suspension which is nothing but the projection of the suspension in the pitch plane.

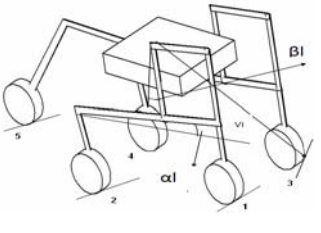


Fig. 7. Orientation on an arbitrary terrain

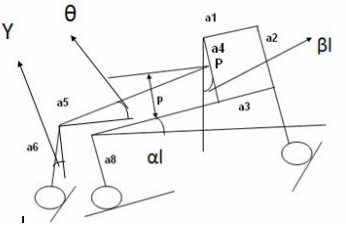


Fig. 8. Planar projection of the suspension mechanism

As shown in figure7 P represents the point of attachment of the chassis. Vi represents the radius vector form the centre of the mass to the wheel ground contact point of the ith wheel. The planar projection showing the link and geometrical parameters are shown in figure8. It is to be noted that the first four wheels are symmetrical with each other. So the link parameters for the front part of the suspensions on both the font and back side are same. αl and βl are the angles made by the corresponding links on the front and αr and βr on the back side of the suspension mechanism and. The radius vector of the wheels in the pitch plane are given by the following expression

$$V_{1}x = a_1 \sin(\alpha_l) - a_2 \sin(\beta_l) \tag{24}$$

$$V_1 y = a_2 \cos(\beta_l) - p\cos(\beta_l) - a_1 \sin(\alpha_l). \tag{25}$$

$$V_2 x = a_4 \sin(\beta_l) + a_7 \cos(\alpha_l) - a_8 \sin(\alpha_l) \tag{26}$$

$$V_2 y = a_8 \cos(\alpha_l) + a_7 \sin(\alpha_l) - a_8 \sin(\alpha_l) \tag{27}$$

$$V_3 x = a_1 \sin(\alpha_r) - a_2 \sin(\beta_r) \tag{28}$$

$$V_3 y = a_2 \cos(\beta_r) - p\cos(\beta_r) - a_1 \sin(\alpha_r) \tag{29}$$

$$V_4 x = a_4 \sin(\beta_r) + a_7 \cos(\alpha_r) - a_8 \sin(\alpha_r) \tag{30}$$

$$V_4 y = a_8 \cos(\alpha_r) + a_7 \sin(\alpha_r) - a_8 \sin(\alpha_r) \tag{31}$$

The radius vectors in the global reference frame are given by

$$rfai = R_z(\beta) R_y(\alpha) * \begin{bmatrix} Vix \\ Viy \end{bmatrix} + R\left[\frac{(2.5-i)}{|2.5-i|} H \ (-1)^{i+1} W \ 0 \right]^T + r. \begin{bmatrix} nix \\ niy \\ niz \end{bmatrix} \forall i \in \{1,2,3,4\}$$

where *H* and *W* are the width and depth of the chassis in global x and z direction respectively, r represents the wheel radius and *nix, niy*, *niz* represents the unit vector f normal forces in the global frame

Let $[m_{tXi}\ m_{tYi}\ m_{tZi}\ ]^T$, $[m_{nXi}\ m_{nYi}\ m_{nZi}\ ]^T$ be the unit moment vectors due to $\vec{T}_i$ and $\vec{N}_i$ respectively.

Hence the quasi-static equations that relate the normal and traction forces to the forces on the chassis of a three dimensional (3D) suspension vehicle can be put in the form:

$$\bar{A}.\bar{C} = \bar{D} \tag{32}$$

where

$$\bar{C} = [|\vec{T}_1| \ |\vec{N}_1| \ |\vec{T}_2| \ |\vec{N}_2| \ |\vec{T}_3| \ |\vec{N}_3| \ |\vec{T}_4| \ |\vec{N}_4|]^T$$

$$\bar{D} = [F_x \ F_y \ F_z \ M_x \ M_y \ M_z]^T$$

$$\bar{A} = \begin{bmatrix} t_{netx1} & n_{netx1} & t_{netx2} & n_{netx2} & t_{netx3} & n_{netx3} & t_{netx4} & n_{netx4} \\ t_{nety1} & n_{nety1} & t_{nety2} & n_{nety2} & t_{nety3} & n_{nety4} & t_{nety4} & n_{nety4} \\ t_{netz1} & n_{netz1} & t_{netz2} & n_{netz2} & t_{netz3} & n_{netz3} & t_{netz4} & n_{netz4} \\ m_{tx1} & m_{nx1} & m_{tx2} & m_{nx2} & m_{tx3} & m_{nx3} & m_{tx4} & m_{nx4} \\ m_{ty1} & m_{ny1} & m_{ty2} & m_{ny2} & m_{ty3} & m_{ny3} & m_{ty4} & m_{ny4} \\ m_{tz1} & m_{nz1} & m_{tz2} & m_{nz2} & m_{tz3} & m_{nz3} & m_{tz4} & m_{nz4} \end{bmatrix}$$

The above set of equations are under-constrained in the sense that they have more number of variables than equations. To reach a unique solution we solve the following optimization problem.

$$\min(S),\ S = \sum_{i=1}^{4} \left|\vec{T}_i\right| \tag{33}$$

Subjected to the following constraints

$$\left|\vec{N}_i\right| \geq 0 \quad \forall\, i = \{1,2,3,4\} \tag{34}$$

$$\left|\vec{T}_i\right| \leq \mu_i \left|\vec{N}_i\right| \forall\, i = \{1,2,3,4\} \tag{35}$$

$$\Gamma_i^{\min} \leq (\left|\vec{T}_i\right|.r) \leq \Gamma_i^{\max} \quad \forall\, i = \{1,2,3,4\} \tag{36}$$

Where (34) corresponds to the constraint that the wheel maintains contact with the ground always, (35) corresponds to the no-slip constraint and (36) corresponds to the constraint that the torque required to generate the required traction is between $\Gamma_i^{\min}$ and $\Gamma_i^{\max}$.

6 Simulations and Results

Simulations were performed using Matlab and Msc Visual Nastran. Extensive simulations for 5-wheeled rover on uneven terrain were performed. Maximum coefficient of friction and maximum torque requirement were selected as the major parameters for evaluating the performance of the mechanism. We set the maximum permissible torque of the motor at 4N-m.The coefficient of friction between the wheel and terrain was varied and minimum co-efficient of friction required to overcome the given terrain was noted for the proposed suspension mechanism. Figure9 shows 5-wheeled rover climbing a terrain with 70^0 degree discontinuity about two times the wheel diameter. The plot of the wheel motor torques is shown in figure 10.

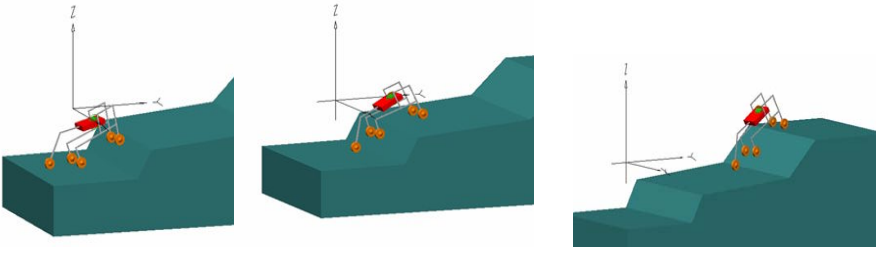
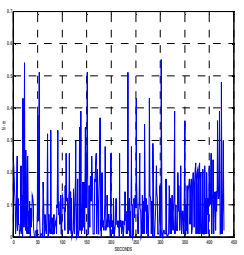
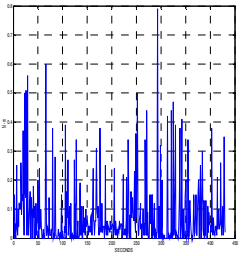


Fig. 9. 5-wheeled rover climbing a sample terrain

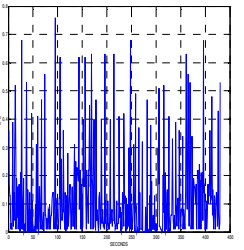
As can be seen from figure10 the maximum torque reading comes out to be around 0.8 N-m when the fourth wheel is climbing the obstacle. The coefficient of friction between the wheel and the terrain was kept at 0.5 which was to be found to be the minimum for the rover to climb over the given terrain. Figure11 shows the satisfaction of the non-holonomic constraint i.e the rover's velocity along the direction



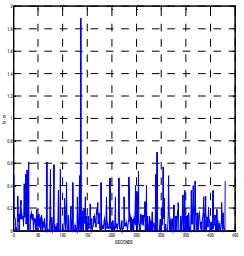
Torque Wheel1



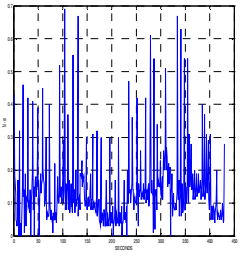
Torque Wheel2



Torque Wheel3



Torque Wheel4



Torque Wheel5

Fig. 10. Wheel torques in N-m

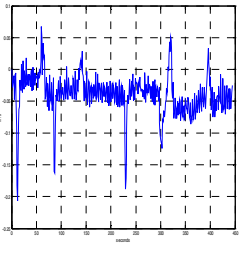


Fig. 11. Velocity Vx of the chassis

perpendicular to the wheel is shown to be negligible in the plot which justifies our assumption that $T_{xi} \approx 0$

Figure 12 shows the 5-wheeled rover traversing on a fully 3D terrain which has slopes in all three orthogonal directions. The minimum coefficient of friction requirement for this terrain went up to 0.7. The increment in the requirement of the coefficient of friction was to ensure minimum lateral slippage of the wheels since the terrain also includes lateral slopes. The plot of wheel motor torques for the terrain is shown in figure13.

As can be seen from figure13 that the maximum torque requirement of the rover has increased to 2.6 N-m, this follows from the fact that due to the complexity of the terrain there is corresponding increase in the amount of traction force required to maintain equilibrium of the system.

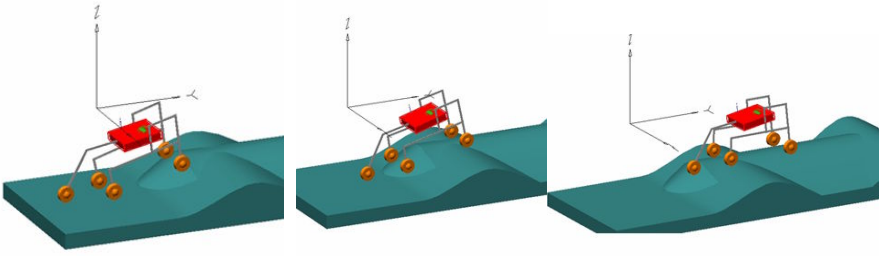
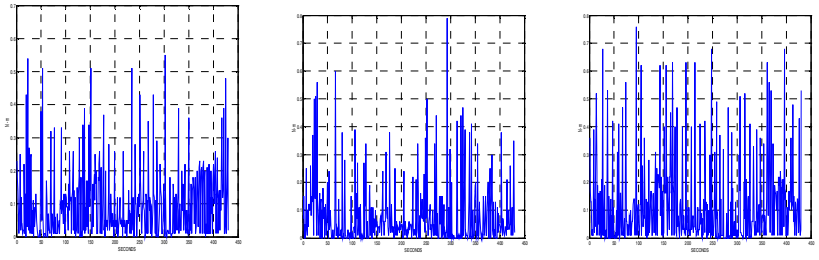
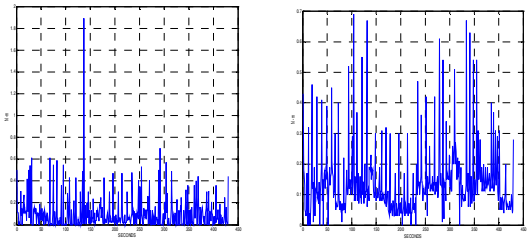


Fig. 12. 5-Wheeled rover on a 3D Terrain



Torque Wheel1 Torque Wheel2 Torque Wheel3



Torque Wheel 4 Torque Wheel5

Fig. 13. Wheel torques for 3D terrain in N-m

7 Conclusions and Future Work

In this paper we have proposed a new five wheeled passive suspension rover which is deduced from the current existing suspension mechanism Shrimp by significantly reducing the number of joints and the number of wheels but yet maintaining a good terrain adaptability. The compliance added to the back wheel is an added novelty in the proposed mechanism. A quasi-static model for wheel motor torque control was presented. Uneven terrain simulations confirm the efficacy of the proposed mechanism. The future work is related to the development of the rover and analyzing path tracking capability of the rover in a fully 3D rough terrain.

References

[1] Volpe, R., Balaram, J., Ohm, T., Ivlev, R.: Rocky 7: A next generation mars rover prototype. J. Advanced Robotics 11(4), 341–358 (1997)
[2] Estier, T., Crausaz, Y., Merminod, B., Lauria, M., Piguet, R., Siegwart, R.: An innovative space rover with extended climbing abilities. In: Proc. Int. Conf. Robotics in Challenging Environments, Albuquerque, USA (2000)
[3] Low Power Mobility System for Micro Planetary Rover "Micro5"in Proc i-SAIRAS-99,ESTEC Noorwidijik, The NetherLands
[4] Lunokhod-1, FTD-MT-24-1022-71, pp.66-77
[5] Eremenko, A., et al.: Rover in The Mars 96 MISSION. Missions, Technologies and Design of Planetary Mobile Vehicles, 277–300 (1992)
[6] Ghosal, A.: Robotics –Fundamental Concepts and Analysis, 1st edn. Oxford University Press, Oxford
[7] Lauria, M., Piguet, Y., Siegwart, R.: Octopus: An Autonomous Wheeled Climbing-Robot. In: CLAWAR (2002)
[8] Eathakota, V.P., Singh, A.K., Kolachalama, S., Madhava Krishna, K.: Linear Force Actuator based Active Suspension for Rough Terrain Navigation. In: Proc. 18th IEEE-International Symposium on measurements and Control in Robotics, Banglore India,
[9] Lamon, P., Krebs, A., Lauria, M., Siegwart, R.: Wheel Torque Control for rough terrain tover. In: Proc. International Conference on Robotics and Automation ICRA-(2005)
[10] Barlas, F.: Design of Mars Rover Supension Mechanism.,Master of Science Thesis, Izmir Institute of Technology, Izmir Turkey
[11] Chugo, D., Kawabata, K., Kaetsu, H., Asama, H., Mishima, T.: Mechanical Design of Step Climbing Vehicle with Passive Linkages. In: Bioinspiration and Robotics: Walking and Climbing Robot, pp. 544, I-Tech, Vienna, Austria, EU (September 2007) ISBN 978-3-902613-15-8

A Study on Wafer-Handling Robot with Coaxial Twin-Shaft Magnetic Fluid Seals

Ming Cong*, Penglei Dai, and Huili Shi

School of Mechanical Engineering, Dalian University of Technology,
Box 116023, Dalian, China
Tel.: +0411-84707147; Fax: +0411-84707147
congm@dlut.edu.cn

Abstract. Wafer-handling robot is an important IC equipment in wafer manufacturing system. As such robot works in vacuum environment, there is a high requirement for the sealing device of it. Magnetic fluid rotary seal shows the effectiveness in machinery operating in a vacuum chamber, and the advantages of simple design, low friction and zero leakage at almost any rotation speed. This paper deduces the seal differential pressure formulas for magnetic fluid rotary seals. A coaxial twin-shaft magnetic fluid seals for wafer-handling robot is designed, of which the important structure parameters are optimized, the magnetic flux density distribution in the sealing gaps is analyzed and the pressure resistance of inner and outer shaft is calculated. Finally, a wafer-handling robot equipped with coaxial twin-shaft magnetic fluid rotary seals and bellows seal is devised. It is shown that this kind of robot can enhance the grade of vacuum and cleanliness in wafer processing system.

Keywords: Wafer-handling robot, magnetic fluid, seals, coaxial twin-shaft, magnetic field.

1 Introduction

In general, a wafer's production is required a dust free vacuum environment to prevent the surface of the wafer from pollution and achieve higher IC integration levels. So the wafer-handling robot working in semiconductor cluster tools need to be free of dust during operation [1], [2]. To isolate pollution source from the vacuum chamber, a reliable hermetical technique is crucial for wafer-handling robot.

Magnetic fluid seals serve to keep magnetic fluid inside a gap when subjected to a magnetic force. They ensure total sealing with no need to come into contact with the sliding surface [3]. Compared with traditional mechanical seals, magnetic fluid seals have unique characters of zero leakage, lowest contamination level, minimal friction torque requirements and no maintenance required [4], [5]. Considering above advantages, magnetic fluid seal is appropriate for the sealing of torque transmission in ultrahigh vacuum, highly clean, hazardous or corrosive environments, especially in semiconductor processing equipment.

* Corresponding author.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 1123–1137, 2009.

As an important kind of wafer-handling robot, the R-θ robot shown in Fig.1 has three DOF (R, θ and Z) to fulfill the mission of transporting wafer among processing units in vacuum environment [6]. To achieve the DOF of R and θ, two rotational motions need to be transmitted from atmosphere to vacuum chamber by the aid of reliable rotary sealing device. Then, pulleys and belts are used to drive the robot arms to achieve the desired motion. It's obvious that magnetic fluid seal is a good choice for rotary hermetic seal of wafer-handling robot. In this paper, an R-θ robot is devised, and the design of coaxial twin-shaft magnetic fluid rotary seals will be introduced emphatically.

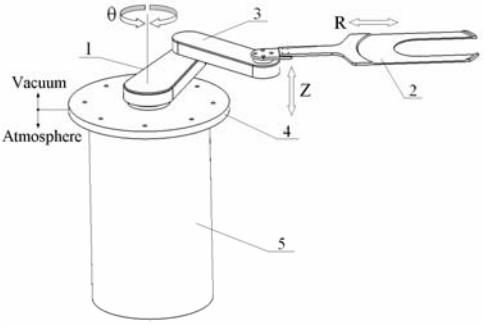


Fig. 1. The model of wafer-handling robot: 1- Back arm; 2- Wafer holder; 3- Forearm; 4- Flange; 5- Enclosure

2 Magnetic Fluid Seals

The common structure of monaxial magnetic fluid rotary seals is shown in Fig.2. The device normally consists of bearings, magnetic pole pieces, permanent magnetic ring, magnetically permeable rotary shaft and magnetic fluid [7]. Its working principle is: the

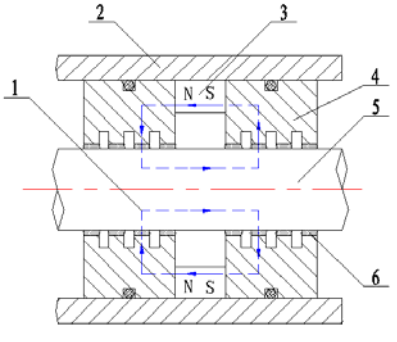


Fig. 2. The typical structure of monaxial magnetic fluid rotary seals: 1- Magnetic circuit; 2- Housing; 3- Permanent magnetic ring; 4- Pole piece; 5- Rotary shaft; 6- Magnetic fluid

magnetic pole pieces, the permanent magnetic ring, and the magnetically permeable rotary shaft compose a closed magnetic circuit. Because of the magnetic energy in the permanent magnetic ring, a strong magnetic field is generated in the gaps between rotary shaft and magnetic pole pieces, and holds the magnetic fluid in gaps tightly. The magnetic fluid kept in gaps forms "O"-rings, which lock gaps and achieve sealing [8], [9]. When the magnetic fluid seal makes isolation for the two chambers whose pressures are different, it's the magnetic fluid that withstands the superfluous pressure. With the increasing of superfluous pressure, the sealing becomes difficult.

In general, magnetic fluid is treated as continuous and incompressible, according to the ferrohydrodynamic analysis, the ferrohydrodynamic Bernoulli equation is represented as [10]

$$p^* + \frac{1}{2}\rho v^2 + \rho g h - \int_0^B M(B)dB = const \tag{1}$$

Where p^* is the pressure in the magnetic fluid, ρ denotes the flow velocity, M is the magnetization of magnetic fluid, B is the magnetic flux density, g is the acceleration of gravity, $\rho g h$ is the gravitational potential with a height h measured relative to the reference level for the gravitational potential.

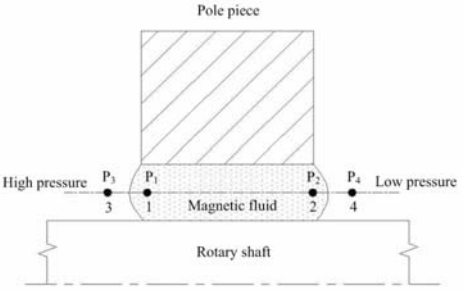


Fig. 3. The microcosmic schematic diagram of magnetic fluid rotary seals

The seal pressure differential between high pressure side and low pressure side, with reference to Fig.3, can be maintained under a rotating shaft condition. The seal pressure differential $\Delta p = p_3 - p_4$ can be estimated by the static equilibrium equation in (1), assuming that the gravitational force is ignored and the magnetic fluid in the sealing gap is in a static state with a condition of the magnetic field to be tangential at the interface, With the magnetic condition at the interface, there would not be a magnetic pressure jump, so that $p_1^* = p_3$ and $p_2^* = p_4$. Thus, the seal pressure differential can be estimated by

$$\Delta p = p_3 - p_4 = p_1^* - p_2^* = \int_{B_1}^{B_2} M(B)dB \tag{2}$$

Where B_1 and B_2 are the magnetic flux density on the left and right surfaces respectively. ΔP is a pressure differential held by one stage of the magnetic fluid seals. When B_1 reaches the maximum value $B_{\max}$ and B_2 reaches the minimum value $B_{\min}$ and the total magnetic fluid is in the saturation, a differential pressure sustained by the seal reaches the maximum

$$\Delta p_{\max} = \int_{B_{\min}}^{B_{\max}} M_s dB = M_s (B_{\max} - B_{\min}) \tag{3}$$

On the basis of formula (3), we can obtain the seal differential pressure formulas as follow

$$\Delta p_{total} = \sum_{i=1}^{N} M_s (B_{\max} - B_{\min}) \tag{4}$$

Where Δp_{total} is the maximum burst pressure sustained by the magnetic fluid rotary seals, M_s is the saturation magnetization of magnetic fluid, N is the number of seal stages, $B_{\max}$ and $B_{\min}$ are the maximal and minimal value of the magnetic flux density in the magnetic fluid respectively.

Equation (4) shows that the maximum seal pressure differential can be increased by optimizing following parameters:

First, the number of seal stages N should be as large as possible. But a large number of seal stages will result in complicated structures. In addition, for the same seal length, a high N value will result in a relative high value for $B_{\min}$ which will reduce the gain in seal pressure per seal step [11]. Therefore, we choose eight stage seals in the design of this paper.

Second, the material constant M_s should be as high as possible. But the magnetic fluid with high M_s will bring a high cost in the design of sealing device. In this research we choose a lower vapor pressure magnetic fluid with saturation magnetization M_s =30 KA/m.

The last design parameter that should be maximized is $B_{\max} - B_{\min}$. Using finite element simulations of the magnetic flux density within the sealing gap, we can optimize this parameter by varying structure parameters of magnetic rotary seals, and this is what we investigate in this paper.

3 Coaxial Twin-Shaft Magnetic Fluid Rotary Seals

The generic structure of coaxial twin-shaft magnetic fluid rotary seals is shown in Fig.4. For monaxial magnetic fluid rotary seal, when the stage number and the permanent magnet volume are fixed, the magnetic field distribution in the sealing gap is determined by the tooth parameters such as the sealing gap g , the tooth width t , the slot width s and the tooth height h . In the paper "A Study of Magnetic Fluid Rotary Seals for Wafer Handling Robot" [12], we analyzed the relationship of the seal pressure differential and the tooth parameters. According to that paper, the tooth parameters can

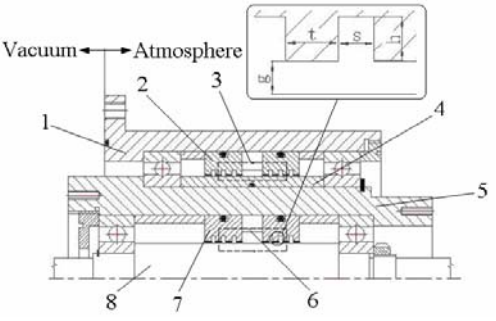


Fig. 4. The generic structure of coaxial twin-shaft magnetic fluid rotary seals: 1- Magnetism isolating housing; 2- Outer magnetic poles; 3- Outer permanent magnetic ring; 4- Magnetic conductive sleeve of outer shaft; 5- Magnetism isolating outer shaft; 6- Inner permanent magnetic ring; 7- Inner magnetic poles; 8- Magnetic conductive inner shaft.

Table 1. Tooth parameters of the coaxial twin-shaft magnetic fluid seals

Tooth parameter	numerical value
Sealing gap (g)	0.1 mm
Tooth width (t)	0.6 mm
Slot width (s)	3 mm
Tooth height (h)	2 mm

be fixed in the design of coaxial twin-shaft magnetic fluid rotary seals. Table.1 shows the fixed tooth parameters.

Besides tooth parameters, the axial relative position of two permanent magnetic rings, magnetic conductive sleeve of outer shaft and magnetism isolating outer shaft are all main factors that influence sealing performance. Optimize the structure parameters of above factors is significant for getting a high $B_{\max} - B_{\min}$ which determines the sealing property of coaxial twin-shaft magnetic fluid rotary seals.

In this paper, we use magnetic finite element method program (ANSYS10.0) to analyze and optimize the structure parameters of sealing device. The seal model constructed by us is two-dimensional and axis-symmetrical, it includes NdFeB magnet (the permanent magnet), stainless steel (pole piece and inner shaft), aluminum (sleeve and outer shaft) and atmosphere. The material parameters for NdFeB are: the remanent magnetic induction Br =1.20 T, the coercivity Hc=935 KA/m, the differential permeability MURX=1.02. The grade of stainless steel is NO.10, it's B-H curve is shown in Fig.5. As magnetism isolating material, aluminum's differential permeability is MURX=1.00, which is as same as that of atmosphere.

As magnetic fields generated by inner and outer permanent magnetic rings affect with each other, the axial relative position of two rings is an important factor that influences sealing performance. In this paper, four different models shown in Fig.6 are

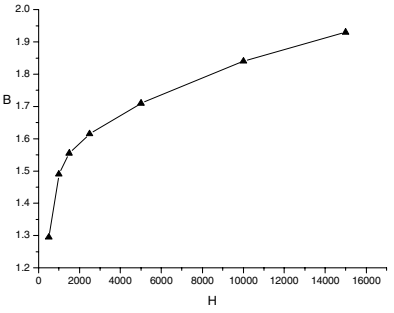


Fig. 5. B-H curve of 10# steel

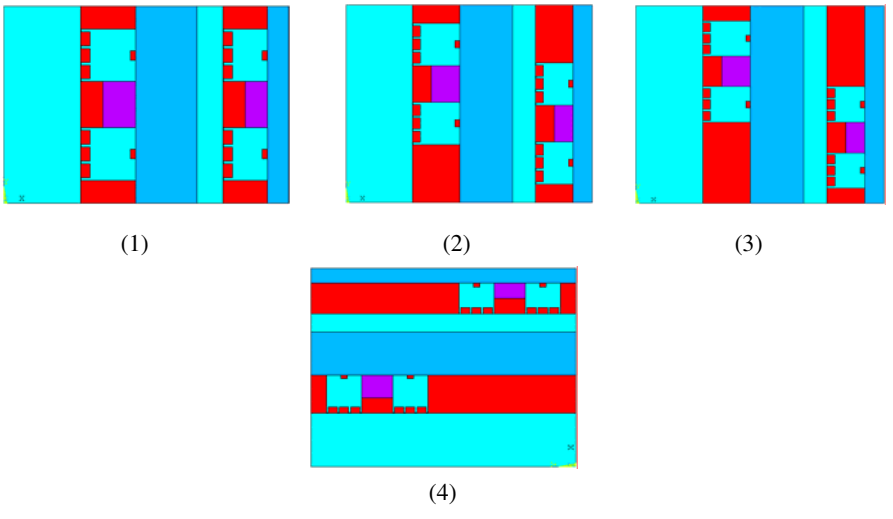


Fig. 6. Four different models of twin-shaft magnetic fluid rotary seals

analyzed by the software of ANSYS10.0. If the radial thickness of outer shaft and shaft sleeve are designed as 14 mm and 6 mm, the analyzing results are presented in Table.2.

From Table.2, we can summarize that model 1 has relative poor hermetic performance compared with other three models. The differentia of P_{total} between model 2, 3 and 4 is very small. In view of retrenching axial size of sealing device, we choose model 2 as final design.

The function of magnetism isolating outer shaft is to insulate the mutual influence between magnetic fields generated by inner and outer magnetic rings. If the radial thickness of outer shaft was too small, the influence between magnetic fields will become severe and the inner P_{total} will be small. On the other hand, an oversize outer shaft will add the radial size of the whole sealing device. So it is very important to choose a right size for magnetism isolating outer shaft. Fig.7 illustrates the relationship

Table 2. The analyzing results of four different models

Num	Inner $\sum_{i=1}^{n} B_{\max}$ (T)	Inner $\sum_{i=1}^{n} B_{\min}$ (T)	Inner ΔP_{total} (Mpa)	Outer $\sum_{i=1}^{n} B_{\max}$ (T)	Outer $\sum_{i=1}^{n} B_{\min}$ (T)	Outer ΔP_{total} (Mpa)
1	12.589	3.884	2.61	9.691	2.364	2.20
2	12.849	3.846	2.70	10.341	2.620	2.32
3	12.883	3.801	2.72	10.364	2.858	2.33
4	12.911	3.769	2.74	10.402	2.562	2.35

of the inner seal pressure differential and the radial thickness of outer shaft when the radial thickness of shaft sleeve is 6 mm. It's obvious that 10 mm is the right size for the radial thickness of magnetism isolating outer shaft.

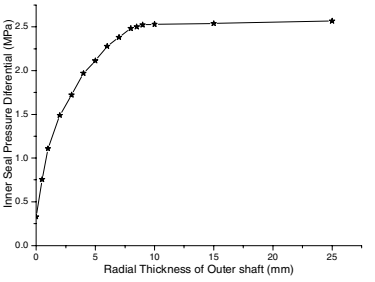


Fig. 7. Inner seal pressure differential versus radial thickness of outer shaft

Magnetic conductive outer shaft sleeve's function is to permit transit of magnetism and ensure an adequate magnetic field in the sealing gaps. A right radial thickness size of outer shaft sleeve can insure a high outer $\triangle P_{total}$ and pinch the radial size of sealing device. Fig.8 illustrates the relationship of the outer seal pressure differential and the radial thickness of outer shaft sleeve when the radial thickness of outer shaft is 10 mm. We can confirm that the right value for the radial thickness of outer shaft sleeve is 9 mm.

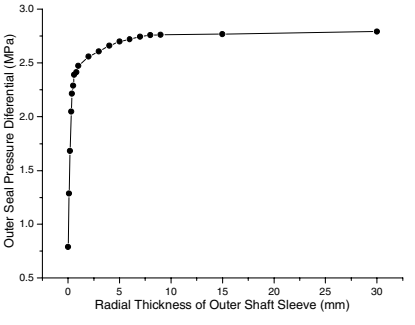


Fig. 8. Outer seal pressure differential versus radial thickness of outer shaft sleeve

Table 3. Parameters of the coaxial twin-shaft magnetic fluid seals

Parameters	Inner shaft	Outer shaft
Shaft diameter	35 mm	80 mm
Stage number	8	8
Sealing gap	0.1 mm	0.1 mm
Tooth width	0.6 mm	0.6 mm
Slot width	3 mm	3 mm
Tooth height	2 mm	2 mm
Inside diameter of permanent magnet	45mm	110mm
Outside diameter of permanent magnet	60mm	120mm
Axial length of permanent magnet	10mm	10mm
Radial thickness of outer shaft	10 mm	
Radial thickness of outer shaft sleeve	9 mm	

As the axial relative position of two rings and the radial thicknesses of outer shaft and outer shaft sleeve are confirmed, all important parameters of coaxial twin-shaft magnetic fluid seals can be designed, which are shown in Table.3. On the basis of these parameters, we construct the seal model shown in Fig.9 in ANSYS10.0.

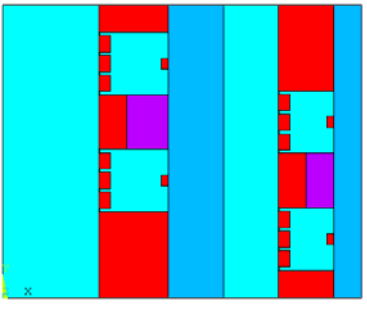


Fig. 9. The seal model constructed by ANSYS10.0

By analyzing the constructed seal model, we can gain the magnetic flux-distribution plot in the cross sections of the gap, shaft, soft magnetic pole piece, permanent magnet and surrounding air region, which is shown in Fig.10. The effects of fringing, leakage and nonlinearity can easily be seen.

The magnetic induction vector of seal structure is shown in Fig.11, from which we can see the flow of magnetic induction vector. The color of vector reflects the intensity of magnetic induction. It's obvious that the intensity under pole pieces is relatively high.

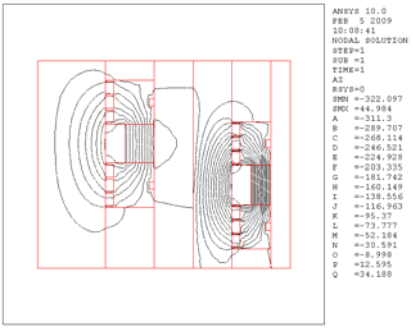


Fig. 10. Magnetic flux-distribution plot for magnetic fluid rotary seals

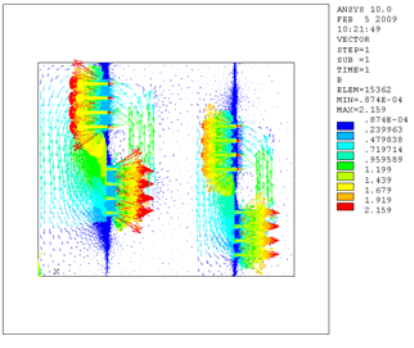


Fig. 11. The magnetic induction vector of seal structure

Fig.12 shows the distribution of the magnetic flux density in the sealing gaps under the pole piece. It is easy to make out that the maximum flux density is under tooth and the minimum flux density is under slot.

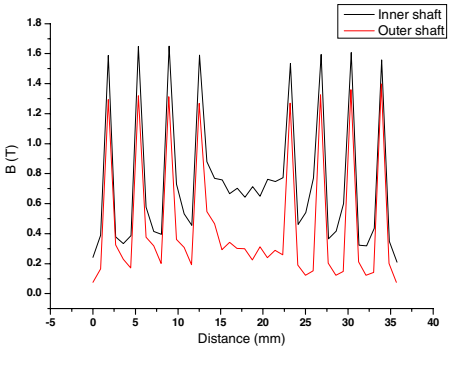


Fig. 12. The distribution of the magnetic flux density in the sealing gaps

According to eq.4, the seal capacity can be numerated on the basis of magnetic field distribution. If M_s was evaluated as 30 K A/m, the $\triangle P_{total}$ of inner and outer shaft can be calculated as follow:

Inner shaft:

$$\Delta p_{total} = \sum_{i=1}^{8} M_s (B_{\max} - B_{\min}) = M_s (\sum_{i=1}^{8} B_{\max} - \sum_{i=1}^{8} B_{\min})$$
$$= 3 \times 10^4 \times (12.77 - 3.74) = 2.71\,(Mpa)$$

Outer shaft:

$$\Delta p_{total} = \sum_{i=1}^{8} M_s (B_{\max} - B_{\min}) = M_s (\sum_{i=1}^{8} B_{\max} - \sum_{i=1}^{8} B_{\min})$$
$$= 3 \times 10^4 \times (10.55 - 1.47) = 2.72\,(Mpa)$$

From calculating, the $\triangle P_{total}$ of inner and outer shaft are all about 2.7 Mpa. The coaxial twin-shaft magnetic fluid rotary seals designed in this paper can meet the sealing requirements of vacuum in wafer processing system.

The final design of coaxial twin-shaft magnetic fluid rotary seals for wafer-handling robot is illustrated in Fig.13.

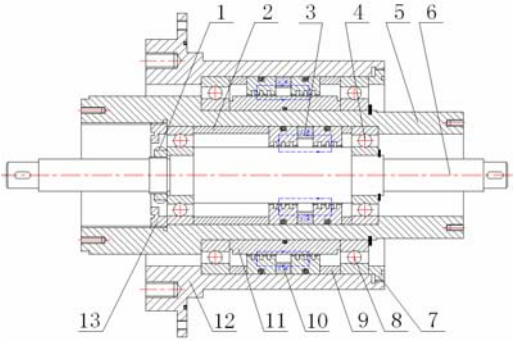


Fig. 13. The final structure of coaxial twin-shaft magnetic fluid rotary seals: 1- Nut; 2- Inner sleeve; 3- Magnetic fluid rotary seals of inner shaft; 4- Bearing of inner shaft; 5- Magnetism isolating outer shaft; 6- Inner shaft; 7- Outer end cover; 8- Bearing of outer shaft; 9- Outer sleeve; 10- Magnetic fluid rotary seals of outer shaft; 11- Magnetic conductive sleeve of outer shaft; 12- Magnetism isolating housing; 13- Inner end cover.

4 A Wafer-Handling Robot Equipped with Coaxial Twin-Shaft Magnetic Fluid Rotary Seals

The wafer-handling robot designed in this paper belongs to the polar coordinate robot, and can finish R (radial linear) movement, θ (rotary) movement and Z (vertical) movement. R, θ and Z movements are driven by AC servomotors. The robot can hold wafer by vacuum absorbing system.

The robot's R movement is illustrated in Fig.14. Based on the principle of epicyclic gear system, the rotation angles of three rotary shafts shown in Fig.14 are designed to be 1:-2:1, so the R movement of end-effector can be achieved by the rotation of back arm. This movement is in charge of radial linear transmission of wafer. Fig.15 shows the robot's θ movement. When the back arm and the big belt pulley of back arm are driven at the same speed and rotary direction, the small belt pulley of back arm keeps still relative to O_2-axes because of movement coupling, thus the θ movement around O_1-axes can be achieved in the precondition of no R movement. This motion enables the wafer to be transferred among wafer processing units in the circumferential direction.

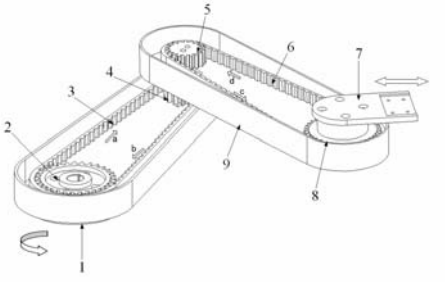


Fig. 14. Schematic diagram of wafer-handling robot's R movement: 1- Back arm; 2- Big belt pulley of back arm; 3- Isochronous cincture of back arm; 4- Small belt pulley of back arm; 5- Small belt pulley of forearm; 6- Isochronous cincture of forearm; 7- Wafer hand; 8- Big belt pulley of forearm; 9- Forearm

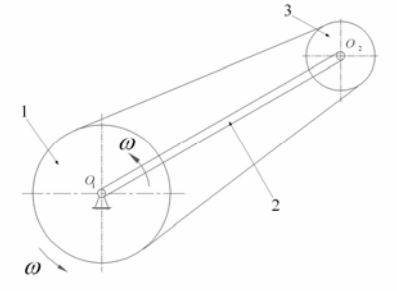


Fig. 15. Schematic diagram of wafer-handling robot's θ movement: 1- Big belt pulley of back arm; 2- Back arm; 3- Small belt pulley of back arm

To realize above R and θ movements of robot, we let the torques of two AC servomotors be transferred into vacuum by the aid of coaxial twin-shaft magnetic fluid rotary seals. Fig.16 is the descriptive diagram of R movement, the AC servomotor is joined to the outer shaft of seal device by harmonic reducer and isochronous cincture,

and the upper side of outer shaft is connected to the back arm body directly. So the R movement can be achieved by the rotation of AC servomotor of outer shaft. The descriptive diagram of θ movement for wafer-handling robot is shown in Fig.17. When the two driving AC servomotors rotate at the same direction, their torques are transferred into the inner and outer shafts of seal device respectively. The control system of robot can let big belt pulley and back arm body that attached on seal shafts rotate at the same speed to achieve the θ movement of robot.

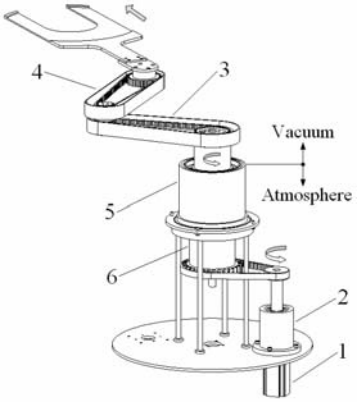


Fig. 16. Descriptive diagram for R movement of wafer-handling robot: 1- AC servomotor of outer shaft; 2- Harmonic reducer of outer shaft; 3- Back arm; 4- Forearm; 5- Coaxial twin-shaft magnetic fluid rotary seals; 6- Outer shaft

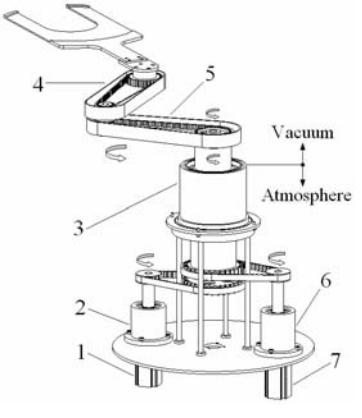


Fig. 17. Descriptive diagram for θ movement of wafer-handling robot: 1- AC servomotor of inner shaft; 2- Harmonic reducer of inner shaft; 3- Coaxial twin-shaft magnetic fluid rotary seals; 4- Forearm; 5- Back arm; 6- Harmonic reducer of outer shaft; 7- AC servomotor of outer shaft

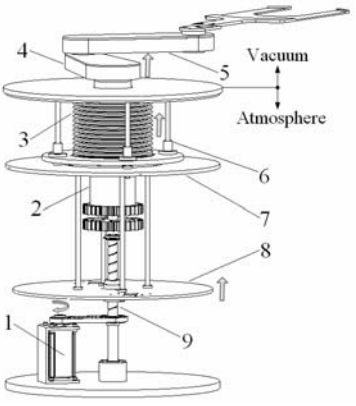


Fig. 18. Descriptive diagram for Z movement of wafer-handling robot: 1- AC servomotor of screw; 2- Outer shaft; 3- Bellows seal; 4- Back arm; 5- Forearm; 6- Linear bearing; 7- Immovable platform; 8- Movable platform; 9- Screw

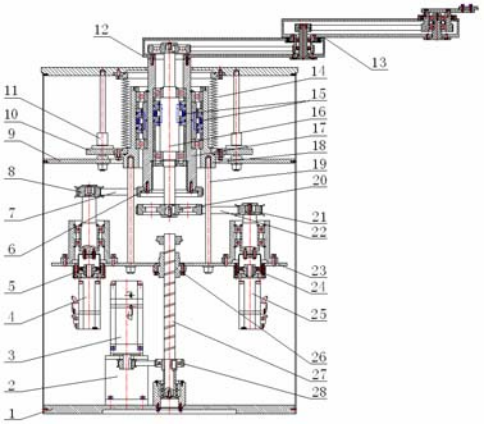


Fig. 19. The overall structure of wafer-handling robot: 1- Base; 2- Holder of servomotor; 3- Lift AC servomotor; 4- AC servomotor of outer shaft; 5- Harmonic reducer of outer shaft; 6- Big belt pulley of outer shaft; 7- Isochronous cincture of outer shaft; 8- Small belt pulley of outer shaft; 9- Immovable platform; 10- Lift plate; 11- Linear bearing; 12- Driving end of back arm; 13- Subassembly of robot arm; 14- Bellows seal; 15- Magnetic fluid rotary seals; 16- Inner shaft; 17- Outer shaft; 18- Sleeve; 19- Draw-bar; 20- Big belt pulley of inner shaft; 21- Small belt pulley of inner shaft; 22- Isochronous cincture of inner shaft; 23- Movable platform; 24- Harmonic reducer of inner shaft; 25- AC servomotor of inner shaft; 26- Screw-nut; 27- Screw; 28- Belt pulley of screw

The function of Z movement is to adjust the vertical position of wafer hand when the robot take or set wafer among processing units. When the vertical movement is transferred from atmosphere to vacuum, a reliable linear seal device is also needed to isolate pollution source from the vacuum chamber. As a simple and effective linear seal

structure, bellows seal can still perform a good linear seal property in the condition of enduring several atmospheric intensity of pressure. In the design of this paper, we use bellows seal to fulfill the sealing requirement of Z movement. Fig.18 shows the descriptive diagram for Z movement of wafer-handling robot. The bellows seal enclose the coaxial twin-shaft magnetic fluid rotary seals to form a three shafts sealing structure which is attached to the movable platform by draw-bars. By the aid of screw-nut pairs and isochronous cincture, the rotary movement of AC servomotor is transform into vertical movement of movable platform. As the three shafts sealing structure move vertically with movable platform, the Z movement of robot is realized.

Fig.19 shows overall structure of wafer-handling robot designed in this paper. This robot can achieve radial linear movement, rotary movement and vertical movement to fulfill the mission of transferring wafer in the vacuum environment.

5 Conclusion

Magnetic fluid seal is an effective method for designing the rotary feedthrough of wafer-handling robot. This paper designs a coaxial twin-shaft magnetic fluid rotary seals, which has high sealing capability and can commendably meet the sealing requirements of wafer processing system. A wafer -handling robot equipped with above seals is devised, it can enhance the grade of vacuum and cleanliness of wafer manufacturing environment. By using the most advanced magnetic fluid technology and seal design technology, magnetic fluid rotary seals can provide satisfactory performance in the challenging new application of wafer-handling robot.

Acknowledgments

This research is supported by the National Natural Science Foundation of China Grant#50675027 entitled “Research on Key Technologies of wafer-handling robot oriented to IC manufacturing” Ming Cong (PI).

References

1. Korczynski, E.: Design challenges in vacuum robotics. Solid State Technology 39, 70–76 (1996)
2. Takashi, K.: Vacuum manipulator for semiconductor manufacturing equipment. Industrial Robot 29, 324–328 (2002)
3. Mitsuya, Y., Tomita, K., Takeuchi, K.: Meniscus shape and splashing of magnetic fluid in a magnetic fluid seal. Journal of Tribology 1999, 377–383 (1999)
4. Ochonski, W.: New designs of magnetic fluid exclusion seals for rolling bearings. Industrial Lubrication and Tribology 57, 107–115 (2005)
5. Bhimani, Z., Wilson, B.: New low cost ferrofluidic sealing challenges the mechanical seal. Industrial Lubrication and Tribology 49, 288–290 (1997)

6. Ming, C., Yumin, Z., Ying, J., Renke, K., Dongming, G.: An automated wafer-handling system based on the integrated circuit equipments. In: IEEE International Conference on Robotics and Biomimetics, ROBIO, July 2005, vol. 2005, pp. 240–245 (2005)
7. Szydlo, Z., Ochonski, W., Zachara, B.: Experiments on magnetic fluid rotary seals operating under vacuum conditions. TriboTest 11, 345–354 (2005)
8. Mitamura, Y., Arioka, S., Sakota, D., Sekine, K., Azegami, M.: Application of a magnetic fluid seal to rotary blood pumps. Journal of Physics Condensed Matter 20, 204–208 (2008)
9. Raj, K., Moskowitz, B., Casciari, R.: Advances in ferrofluid technology. Journal of Magnetism and Magnetic Materials 149, 174–180 (1995)
10. Rosensweig, R.E.: Ferrohydrodynamics. Cambridge University Press, Cambridge (1985)
11. Sarma, M.S., Stahl, P., Ward, A.: Magnetic-field analysis of ferrofluidic seals for optimum design. Journal of Applied Physics 55, 2595–2597 (1983)
12. Cong, M., Shi, H.: A study of magnetic fluid rotary seals for wafer handling robot. In: The 15th international conference on mechatronics and machine vision in practice, December 2008, vol. 2008, pp. 269–273 (2008)

A Study on the Key Technology of Autonomous Underwater Vehicle–AUV's Self Rescue Beaconing System

Guohua Xu*, Kun Yu, and Xiaoliang Chen

Underwater Technology Laboratory, School of Naval Architecture and Engineering, Huazhong University of Science and Technology, Wuhan, 430074, China

Abstract. It is dangerous for Autonomous UnderwaterVehicle (AUV) to work in the deep sea. so varies of risk may be faced by an AUV when working in such environment. Thus, it costs a lot of humanbeing resource to develop an AUV while the data or the collection that attained by the working of an AUV in the deep sea is valuable. Especially for an AUV which carries person, the safety of itself is much more important. This Paper introduces an AUV underwater robot system based on MSP430 microcontroller. The technology of miniaturization, low power consumption, airproof and reliability are given in detailed. It will be disastrous if the AUV is lost in the sea and, at the mean time, an expense of oceanic research and development. So it is extremely significant to explore the self-rescue system for AUV.

Keywords: AUV, Self-Rescue System, Prediction of reliability, Low Power Consumption Design, MSP430.

1 Introduction

Autonomous Underwater Vehicle (AUVs) is one of the most development and application underwater vehicle currently, which have excellent prospects with its own energy equipment and the active regions that aren't limited to the length of the cable. As we all know, the working environment of Autonomous Underwater Vehicle (AUV) is very dangerous, so kinds of risk may be faced by an AUV when working in the deep sea. Then, it costs a lot to develop an AUV and also the data or the collection that attained by an AUV each time is valuable especially for a ROV which carries people, the safety of itself is much more important. It will be disastrous if the AUV is lost in the sea and, at the mean time, an expense of oceanic research and development. So it is extremely significant to explore the self-rescue system for AUV [1] [2] [3].

This article is mainly discussed the key technology of the beacon system include Anti-jamming technology, small bulk technology, low cost technology, good concealment and so on.

* The nation eleventh "Five-year Plan" – "developing the self-rescue equipment" support.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 1138–1145, 2009.

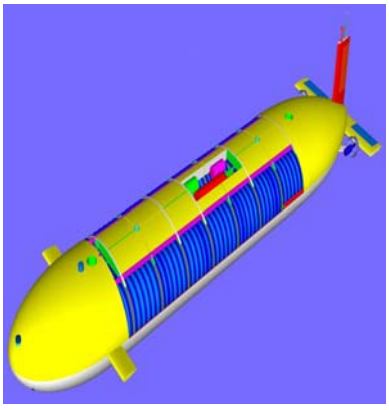


Fig. 1. The ThreeD model of the AUV

Fig. 2. The 1000 meter depth AUV in the pool

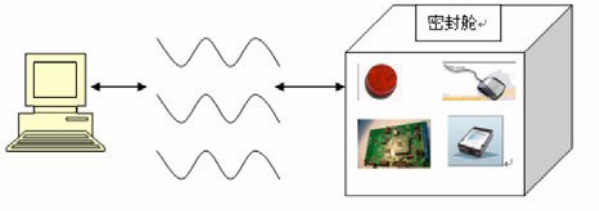


Fig. 3. the self-rescue beacon system plan. This shows the self-rescue system has three parts-surface computer, communication system and underwater information controlling device.

2 Component and Function of the Beaconing System

2.1 Hardware Architecture

As a self-rescue beaconing system, its aim is obtain the crash position Real-time information and send them to the mother ship. The underwater device gain the GPS positioning information of the crash AUV and do the integration and data compression. Meanwhile this position information can be transmitted to the surface computer at the same time by radio device. For the long working target, msp430 also deal the power management and time monitoring features. [2]

2.2 Software Architecture

The software system includes two parts, the main programmer and interruption service subprogramme. The main programme takes charge of initialization, processing senor's data, filter, processing GPS data, and checking the code. The interruption service subprogramme's function is receiving and sending GPS data.[4]

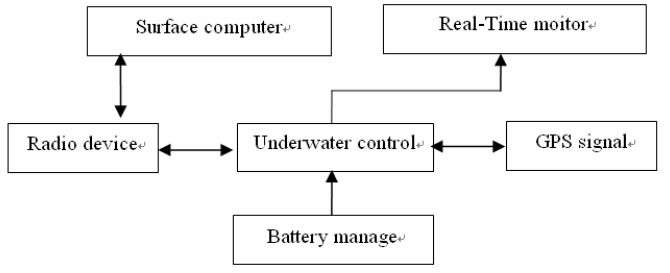


Fig. 4. The Beaconing system hardware architecture diagram

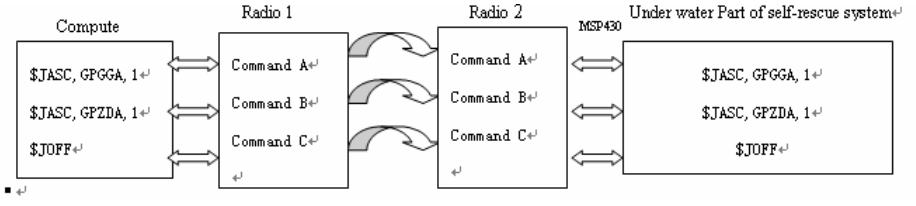


Fig. 5. Multi-computers communication software schematic

Since GPS signal relatives to frequency, analysis of algorithm's time complexity and space complexity is necessary for low power consumption system.

Controlling by computer above water, it can enter into the subprogramme for data interception soon.

3 Key Technology of the Self-rescue Beaconing System

3.1 Low Cost

In order to ensure long practical working in the deep-sea, the long-term effective is a very important for the beacon system. This target requires the low-power beaconing system.The

(1) Micro-controller

The battery consume of the Msp430f149 in RAM data registing only is 0.1 μA while 250 $\mu A / MIPS$ in the active mode .The highest Current drain only is 50 nA with the interrupt Vector working mode. As it is said, the msp430 is the lowest cost Single-chip. For its cpu interrupt request time can be limited in wake up 6μs. Through the rational programming, it is not only can reduce system power consumption, but also respond quickly for external events. Thus, this serial Single-chip is chosen in this project.

(2) Lithium batteries

In our designed circuit, we choose the lithium batteries as the power afford device.There are six groups lithium batteries, were divided into two parts.One is

supplied GPS and the other is wireless radio power supply. For its rated capacity is 11AH, the nominal voltage is 3.7v, battery 0.2CA average discharge voltage. As a result of performance requirements, in low-power state, it can be maintained 41.56 hours throughout the whole-day working. [5]

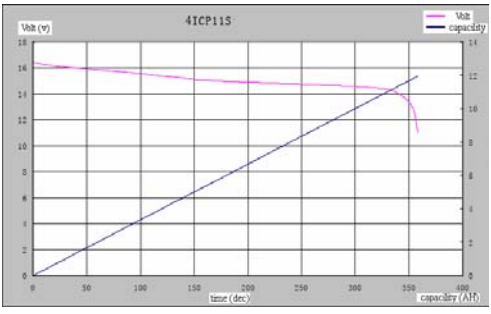


Fig. 6. Charge time –volt relation of the power supply – 4IPIIS Lithium battery. Red line refers to capability, the blue line refers to volt, and transverse refers to time. These figure describe in the caption can be receive a long working Cycle. [7]

(3) Battery Protection Board
To prevent battery Excessive discharge, Battery Protection Board offers a Strategy to allocation of the use of batteries rationally.

3.2 Reliability

Reliability is a main significant System target for a longtime-harsh working environment. Above all, the beacon system require a high positioning accuracy and low error rate of the GPS Measurement. Secondly, it's important to preclude the interfere such as temperature and humidity, while the low-power MSP430 deal with GPS data. The method can offer a well experiment environment for CPU to accomplish accurate data acquisition and compression tasks. Finally, in order to essure the stability working in the deep-sea, the serial of GPS should be selected for all-day test to see whether or not to meet the requirements of positioning accuracy. At the same time, for a long period of testing of the radio transceiver to see whether it has a lower error rate or not. Through the key experiment of each single components can rich the reliability of the beaconing system.[8]

3.3 Airproof and Antijam

Tow parts of the airproof and antijam technology:

(1) Package the beacon GPS, wireless radio and strobe, in a sealed container ring hollow, while the wall is metal. The method can shield the electromagnetic interference the between GPS and wireless radio signal communication.

(2) GPS antenna was packaged in long or short cover glass. Sealed the bottom, Record the receiving data. Compare the affection data receiving in the open-air environment and sealed glass .Curve data relative in matlab; we can come to interfere data range.

3.4 Miniaturization

Since the beaconing system is on AUV, the miniaturization should be achieved from three aspects.

(1)Make it smaller in struct theory
Under the premise of its normal working, the control circuit should be simpler and more practical.

(2)Make it smaller in function
Choose SMD and highly integrated chips because the beaconing system don't work incessantly.

The heat sink is unnecessary because the system is underwater.

(3)Make it smaller in design
Make electronic components on circuit board closer and use Multilayer PCB.

4 Experiments about Self-rescue Beaconing System

4.1 Low Power Consumption Experiments

(1) The performance of Lithium batteries testing
The Li-batteries ICP11S are used to do the charging and discharging experiment. In the experiment, the batteries are charged by a Power Supply. Firstly, the power supply is set to 4.2V/cell and 0.2 CA. When the current timed down to 0.02cA, the batteries' stabile voltage can reach 4.16V. Then, discharge the batteries and record the time of charging and discharging. Finally, repeat these steps in order to test whether the batteries can meet the requirements of long hours supply.

Experimental results show that the battery for the self-rescue beaconing system is qualified.

(2) test the power chip
The power chip is chosen for supplying to GPS, MSP430 and the A/D module of the radio. According to the requirements for input voltage and low power consumption, the power chip SPX1129M is a good choice.

When the system is working, the battery will supply for GPS, MSP430 and other chips such as MAX3232 incessantly. So the performance of power chip SPX1129M should be tested to ensure that the current will not overload. In the experiments, we monitored the voltage on SPX1129M's output port for long hours to make sure GPS and MSP430 can work normally.

4.2 Software Experiment

For the requirements of low power consumption, MSP430's watch dog is used for timer's delay. Turning system on and external reset signal would give the POR signal which would bring the PUC signal. Besides, timer's overflow and security key error

can give the POR signal too. The POR signal starts up the watch dog, and after POR, MSP430 would enter into AM state. In the state, it can switch to any low-power mode and exit the state when it receives the interrupt signal from peripheral module.

4.3 GPS Experiment

(1) Experiment for GPS performance
This experiment using radio, GPS, controller and computer could test the stability of GPS signal. After the series of experiments, the experimental data from each moment on sunny day and rainy day can be compared with the practical latitude and longitude.

Conclusion: the weather can hardly affect GPS signal but the different moments affect a lot. By the experimental data, it can be educed that the signal between 11 am and 2 pm is not only weak but also sometime totally wrong.

(2)Real-time test
Measure a circuit and a straight road by the GPS in self-rescue beaconing system and a handheld GPS at one time. Then, draw it by Matlab, calculate the error and get a method for data compensation.

Data of measuring a 400m long circuit:

Table 1. The data of measuring a 400m long circuit

Method	Calculated GPS data km)
Handheld GPS	0.373
GPS in self-rescue system	0.343
Actual length	0.400

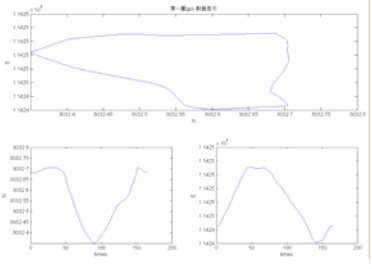


Fig. 7. The data after drawn by Matlab

4.4 Cooperation Experiment in Boat Pool

Fig. 8. shows the initial state of our security experiment stage in the boat pool of Harbin Engineering University.

Fig. 9. shows the surface controller of this beaconing system, the computer receive and integrate the position information from GPS.

Fig. 10. draws the final state of the beaconing system especially the float body.

Fig. 8. Beaconing system in the water

Fig. 9. Surface system

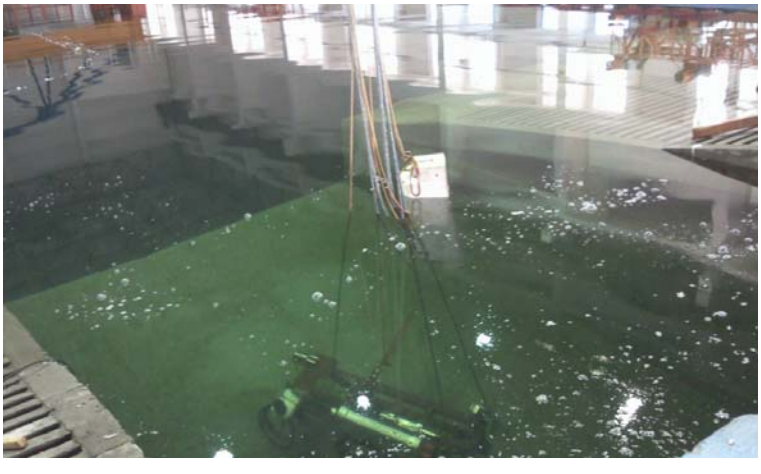


Fig. 10. The final state of the beaconing system

5 Conclusion

In accordance with the research on the technique of underwater unmanned system integration and function-underwater self-safety and sampling work technology of key pre-research project of Department of Military Equipments. Several aspects are detailed to dissertate and study. This article is based on the key technology of AUV beaconing system. Through half-year experiments, we can conclude that the system is practical. Each target is accord with the demand of the longtime-harsh ideal environment. However, there are also two aspects to improve and further study: A better method to boost the feature of the real time response and A better method to remove interference which can enhance the monitoring accuracy.

References

1. Jiang, X.-s., Feng, X.-s., Wang, D.-t.: AUV. Shen Yang. Science Press, Liao Lin (2000)
2. Shen, J.-h., Yan-qin, Y., Qu, X.-s.: MSP430 family of ultra-low power 16-bit single-chip Principle and Application. PekingTsinghua University Press (2004)
3. Hu, D.-k.: FLASH-based MSP430 series ultra-low power 16-bit single-chip, pp. 15–17. Beijing University of Aeronautics and Astronautics Press (2004)
4. Xiang, X.-b., Cai, T.: Intelligent self-rescue system of AUV. Journal of Huazhong University of Science and Technology (Natural Science Edition) 34(7), 111–114 (2006)
5. Xu, G.-h., Duan, G.-q., Tong, J.: A study on fuzzy self-rescue expert system of AUV. Shipbuilding of China 12, 38–41 (2004)
6. Wanninger: Real-Time Differential GPS Error Modeling in Regional Reference Station Networks. In: Proc.IAG Scientific Assembly, IAG Symp., vol. 118, pp. 86–92. Springer, Heidelberg (2006)
7. Saridis, D.R.: Toward the Realization of Intelligent. Proceedings of the IEEE Controls 67(8), 1115–1120 (1979)
8. Garche, J.: Influence of phosphoric acid on both the electroche mistry and the operating behavior of the lead/acid system. Power Sources 33, 213–220 (1991)

The Research on Mechanism, Kinematics and Experiment of 220kv Double-Circuit Transmission Line Inspection Robot

Cheng Li, Gongping Wu, and Heng Cao

College of Power and Mechanical Engineering, Wuhan University, Wuhan, Hubei, China
lichenganny@163.com, gpwu@whu.edu.cn

Abstract. Multi-circuit (extra) high-voltage transmission line places important role in China Power Grid. To ensure its running safety, it is better using inspection robots than helicopters or manual inspections. This paper first presents the mechanism of a kind of inspection robot for 220kv double-circuit transmission line, which has 7 degrees of freedom, designed into double anti-symmetrical retractable arms, interactive sliding and suspending structure. Then build up its kinematics equation using D-H method. In order to solve its reverse solution, apply Back Propagation (BP) neural network in MATLAB. The results have shown the method is effective. To verify the mechanism whether is feasible, build up the model in CATIA, translate it into ADAMS to be simulated. The tests of its prototype in lab shows the simulation is correct. Pictures of the robot overcoming the obstacle on the horizontal line and on the jumpers are shown in this paper.

Keywords: Transmission line inspection robot, Robot mechanism, Kinematics equation, Kinematics Simulation.

1 Introduction

220kv and above multi-circuit high-voltage transmission line is the main line of China Power Grid, which has long distance with complex geographical environment around and strong magnetic field. So it is a heavy task to inspect and maintain the line. Using inspection robots is cheaper than helicopters, and can avoid inspection blind zone which manual inspection can't. Now many countries have been researching on transmission line robot.

The "Expliner" remote control inspection robot [1] developed by Kansai Electric Power (KEPCO, Japan) and Japan Power System Corp (JPS) can overcome spacers and suspension clamps on horizontal cables. It can't move along the cables with tension towers right now. Canada Hydro-Quebec research institute(IREQ) has developed a remote control inspection robot called "Linescout"[2], which can move along the cables with insulation strings, spacers and vibration dampers, but also can't overcome the tensile tower. In [3], Zhou F, etc, introduced an inspection robot with three arms for 110kv transmission line. It can overcome vibration dampers, suspension clamps and the tensile clamps with no other obstacles on the jumpers. In [4], Sun, C, etc,

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 1146–1155, 2009.

introduced an inspection robot with two arms for less than 500kv ground wire or high voltage wire. It can't cross the tensile clamps. The inspection robot presented in [5] is for 220kv single bundle transmission line, which can cross the tensile clamps.

The inspection robot presented in this paper is designed for 220kv double-circuit transmission line, which has spacers on the jumpers to overcome, so this research lays the root for developing other multi-circuit transmission lines. The robot has 7 degrees of freedom (DOF), which can overcome the tensile tower and other hardware.

2 The Mechanism of Inspection Robot

220kv double-circuit transmission line concludes double-circuit cables, suspension towers, tensile towers and different kinds of hardware (such like vibration dampers, suspension clamps, insulation strings). In it, there are spacers between the bundles and also on the jumpers at intervals. The line is shown in Fig.1 in details.

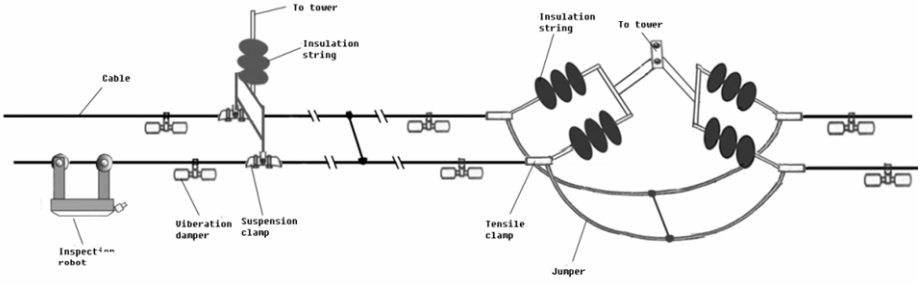


Fig. 1. Double-circuit transmission line

The inspection robot applies wheel mobile mechanism and double anti-symmetrical arms. It has eight joints, in which, the joint 8 is a imaginary joint; joint 1 and 7 are retractable joints on the two arms; joint 2 and 6 are rotation joints of vertical plane; joint 3 and 5 are rotation joints of horizontal plane; joint 4 is a translation joint, which makes the two arms move across. The axis of joint 1 and 7 are parallel with the axis of joint 3 and 5. The axis of joint 2 and 6 are horizontal, intersecting perpendicularly with that of joint 3 and 5. The axis of joint 4 is horizontal, intersecting perpendicularly with that of joint 3 and 5.

In order to keep the robot on the line in security and avoid it falling down, the gripper mechanism is designed under this circumstance. It is composed of the clamping bar (1-1), rotary pair (1-2), constraint rotary pair (1-3) and gripper jaw (1-4). The gripper mechanism is arranged besides the arm-wheels and symmetrically in the transverse direction of the cable.

Its general mechanism schematic and the gripper mechanism schematic are shown in Fig.2.

The odd (even) movable arm can slide along its relative immovable arm, which is driven by lead screw pair fixed in the arm. The odd (even) immovable arm can swing in vertical plane and rotate in horizontal plane, which are driven by worm gear pairs

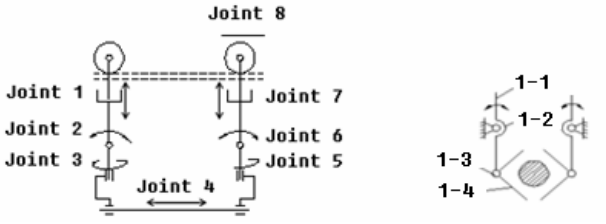


Fig. 2. Inspection robot mechanism schematic and its gripper mechanism schematic

fixed on and besides the slipway. The whole odd (even) arm system can slide along the slipway path, which is implemented by ropes with lead screw pair stimulated by electromotor. The gripper mechanism stalled on the top of the odd (even) movable arm is implemented by lead screw pair to make the gripper open and close. The physical prototype model of the inspection robot is shown in Fig.3.

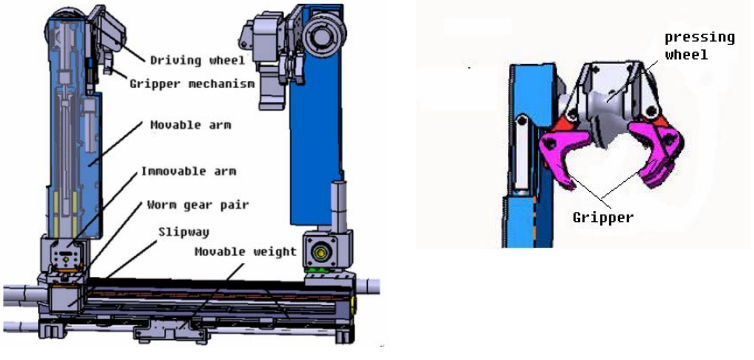


Fig. 3. The inspection robot model and its gripper mechanism

The robot has 7 degrees of freedom, without the gripper mechanism and the driving wheel. Because their freedom is partial freedom, little influence to the whole robot kinematic analysis.

3 7-DOF Inspection Robot Kinematic Analysis

3.1 Kinematics Equation

Since the robot is anti-symmetrical in structure, so we just discuss the kinematics under the situation that when the even arm is stable and the odd arm crosses the obstacle. According to D-H method, the robot is composed of 7 linkages and 7 joints. The 7 linkages are even immovable arm, worm gear box, the linkage composed of slipway and frame, odd immovable arm, odd movable arm, the slipway and worm gear box of the odd arm. The 7 joints are two translation joints between odd (even)

immovable arm and its movable arm, two rotation joints between the immovable arms and the worm gear boxes, two rotation joints between worm gear boxes and slipways, one translation joint between the slipway of the odd arm and the frame. Follow the settings above, the linkages coordinates are shown in Fig.4. The linkages parameters are shown in Table.1.

Coordinate {0} and {1} are congruent. The origin is fixed at the point of intersection that the translation pair of the even movable arm and the even immovable arm with the rotation pair of the worm gear box.

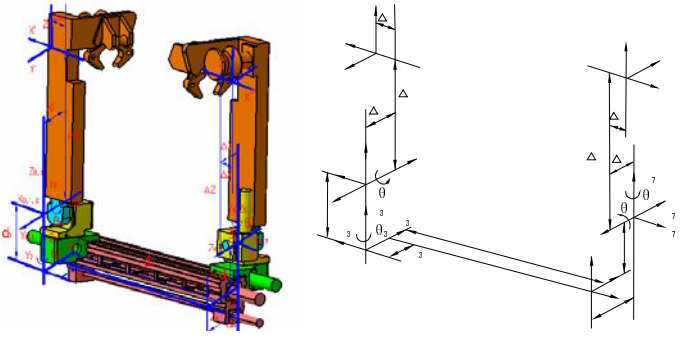


Fig. 4. Linkage coordinates

Table 1. Linkage parameters

Linkage i	a_{i-1}	α_{i-1}	θ_i	d_i	Joint variables	Value range
1	0	0	0	d1	d1	–200 ~ 0
2	0	90°	θ2	0	θ2	–60° ~ 60°
3	0	–90°	θ3	–74	θ3	–270°~ –90°
4	105.5	90°	0	d4	d4	–584 ~ 516
5	105.5	–90°	θ5	74	θ5	–270°~ –90°
6	0	90°	θ6	0	θ6	–60° ~ 60°
7	0	–90°	0	d7	d7	0 ~ 200

Coordinate { } is transformed from coordinate {0}. Coordinate { } is transformed from coordinated {7}. The origins of them are at the center of the cable which under the wheel centroid. The two translation matrixes are as follows:

$$ {}_{0}^{'}T = \begin{bmatrix} 1 & 0 & 0 & -\Delta x \\ 0 & 1 & 0 & \Delta y \\ 0 & 0 & 1 & -\Delta z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad {}_{n}^{7}T = \begin{bmatrix} 1 & 0 & 0 & \Delta x \\ 0 & 1 & 0 & -\Delta y \\ 0 & 0 & 1 & \Delta z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{1} $$

$\Delta x, \Delta y, \Delta z$ are constant. $\Delta x = 9.5, \Delta y = 105.5, \Delta z = 425$

According to the linkage coordinates and linkage parameters, the translation matrixes relative to the ones before have been gotten. Then calculate the translation through (1), finally get the matrix (2).

$$ {}^{'}_{n}T = {}^{'}_{0}T\,{}^{0}_{1}T\,{}^{1}_{2}T\,{}^{2}_{3}T\,{}^{3}_{4}T\,{}^{4}_{5}T\,{}^{5}_{6}T\,{}^{7}_{7}T\,{}^{7}_{n}T = \begin{bmatrix} n_x & o_x & a_x & p_x \\ n_y & o_y & a_y & p_y \\ n_z & o_z & a_z & p_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2} $$

3.2 Kinematics Reverse Solution

Using BP neural network to solve the robot kinematics reverse solution is to train plenty of the samples, in order to get the non-linear mapping from the work variable space to joint variable space. This method can avoid formulas deduction and programming process.

The approach of using BP neural network is as follows: select 300 groups of joints variables: d_i , θ_i by "rand" function in MATLAB. Then the unknown in (2): $n_x, n_y \ldots p_z$ can be calculated; normalize d_i θ_i $n_x, n_y \ldots p_z$ by "premnmx" function, as the input variable and output variable; use 3 layers BP neural network structure, set 20,50,7 as the numbers of each layer; use "tansig", "tansig" and "purlin" as each layer's translation function, the train the network by "trainlm" function. In the network, the target error is 0.001, rate of learning 0.025, the maximum step 3000. After 56 steps, the network can get the error require. Then through the pn value using the net, anti-normalizing function "postmnmx", the joint variable $d_i^{'}$ $\theta_i^{'}$ can be calculated out.

The comparison of the results from using BP neural network and the expected value is shown in Table.2.

Table 2. The results from BP neural network and its error

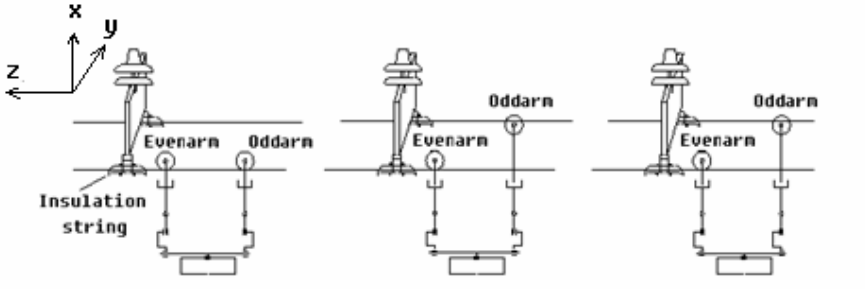
	d_1	θ_2	θ_3	d_4	θ_5	θ_6	d_7
Expected value	–124.33	0.75401	–2.0306	68.919	–3.1524	0.83727	164.33
BP NN Results	–124.84	0.77346	–1.9801	65.563	–3.1486	0.84872	163.17
Error	0.51	0.01945	0.0505	3.356	0.0038	0.01145	1.16
Relative error	0.41%	2.5%	2.49%	4.9%	0.12%	1.37%	0.705%

Through the table above, the error is acceptable, and the speed is fast. So the BP neural network method is workable.

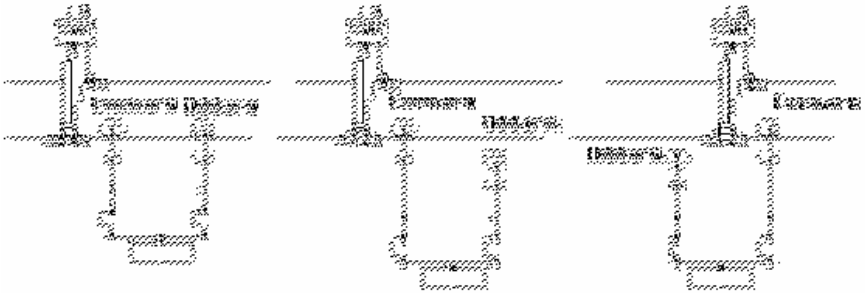
3.3 Kinematics Stimulation Analysis

The approach to solve the kinematics simulation analysis based on the inspection robot prototype model is as follows: first build up the robot 3-D model in CATIA, also build up the cable, vibration dampers, insulation strings, etc; import the models into ADAMS through SimDesigner module interface to add kinematical pairs; impose motion drivers, do the kinematics analysis in ADAMS/View. The inspection robot working environment is seen as rigid. The condition to analyze is at the time when the robot is crossing the suspension insulation strings.

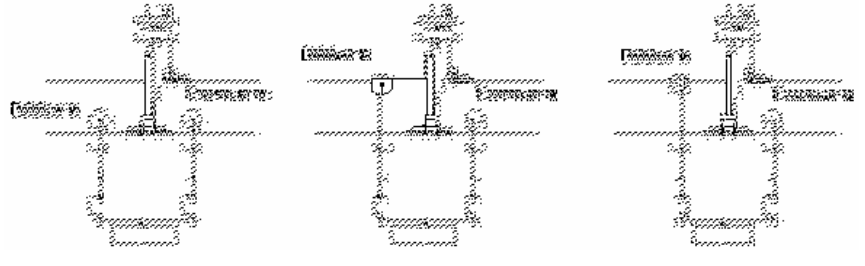
The sequence to cross the suspension insulation strings is as below:



(a) Initial position (b) Odd arm lifts (c) Odd arm rotates 180°clockwise(CW)

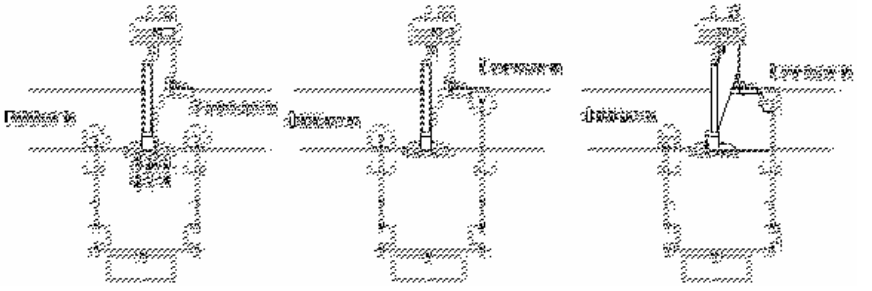


(d) Odd arm descends (e) Odd arm stays, others descend (f) Arms cross first time

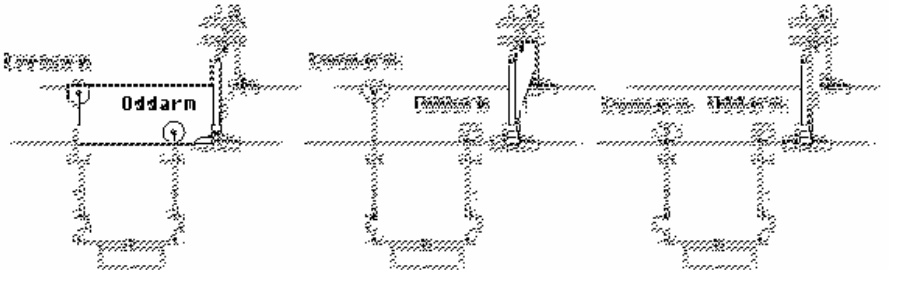


(g) Even arm stays, others lift (h) Odd arm lifts (i) Odd arm rotates180° CCW

Fig. 5. The sequence to cross insulation string



(j) Odd arm descends (k) Even arm lifts (l) Even arm rotates180°CW



(m) Arms cross second time (n) Even arm rotates 180°CCW (o) Even arm descends

Fig. 5. (*continued*)

In order to improve the reality of the simulation, collision detection is needed. Using restriction tools in ADAMS/View can observe the collision directly. So between the odd arm and the insulation string, the even arm and the insulation string, we set up solid to solid contact. Considering the inspection robot is using two arms crossing each other to overcome the insulation string, so the arms will be in different restriction conditions. From (a) to (j), the even arm is fixed with the cable. From (k) to (o), the odd arm is stimulated to be fixed with the cable, and the even arm is released from the full restriction situation. The change described above is written in ADAMS/Solver Commands.

All the joints' actions are settled as velocity-time functions. For example, the "general_motion_1" is presented as the odd movable arm's moving relative to its immovable arm. Its function is "–10*(STEP5(time, 0 , 0 , 1 , 1)–STEP5(time,9 , 0 , 10 , 1) –STEP5(time, 46 , 0 , 47 , 1)+STEP5(time, 55 , 0 , 56 , 1)+STEP5(time, 131 , 0 , 132 , 1)–STEP5(time, 140 , 0 , 141 , 1)–STEP5(time, 177 , 0 , 178 , 1)+STEP5 (time, 186 , 0 , 187 , 1))". It can be shown in Fig.6.

The position of the two arms is important to estimate the collision. So the points on the centerline of the cable under the wheels' centroid are marked. In ADAMS, the

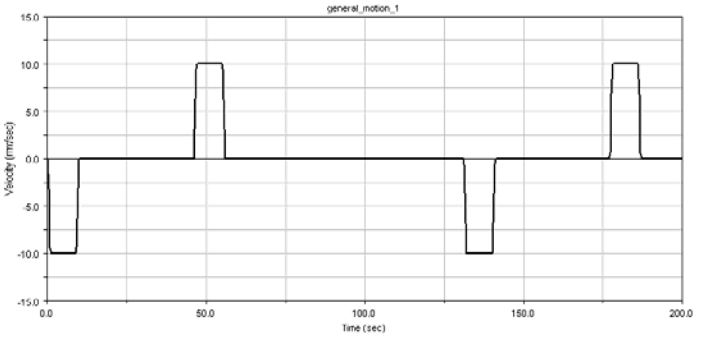


Fig. 6. General_motion_1 function

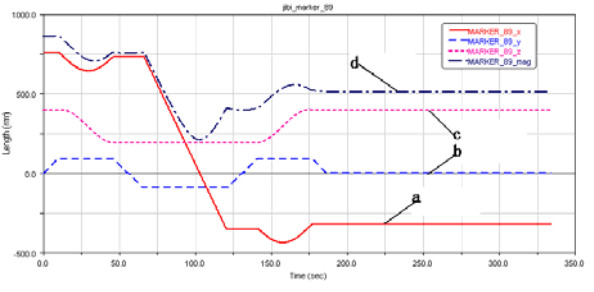


Fig. 7. Displacement of the odd wheel point (a-x direction displacement part, b-y direction displacement part, c-z direction displacement part, d-displacement)

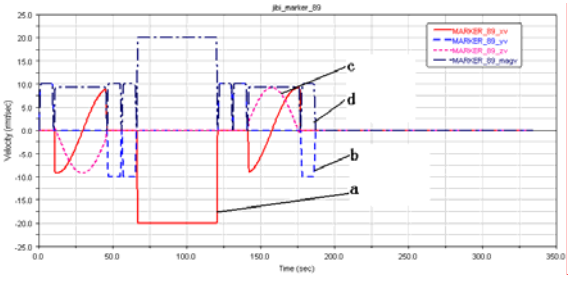


Fig. 8. Velocity of the odd wheel point (a-x direction velocity part, b-y direction velocity part, c-z direction velocity part, d-velocity)

two points' displacement and velocity can be plotted. The displacement and velocity of the point under the odd arm wheel are shown in Fig.7 and Fig.8.

From the two figures we can see that the odd arm moves during 0s to 187s. According to the insulation string size in three directions, it is possible for the robot to overcome this obstacle. The collision detection simulation also improves that the robot can overcome the insulation string. All of above show that, the robot design is feasible.

4 The Prototype and Its Experiment

Use the prototype to verify the kinematical calculation above. A 220kv double-circuit transmission line for testing is built up of 1:1 proportion in the lab. The pictures of the robot crossing the insulation string on horizontal line, translating itself from the horizontal line to the jumper, crossing the spacer on the jumper, shown in Fig.9 proves that the kinematic analysis is correct.

a) The robot crosses the insulation string on horizontal line

b) The robot moves from the horizontal line to the jumper

c) The robot crosses the spacer on the jumper

Fig. 9. The prototype and its test

5 Conclusions

In this paper an inspection robot with 7 DOF for 220kv double-circuit transmission line is introduced. Through the kinematics equation, its reverse solution, the kinematical simulation and the test in lab, which the paper all presented, the mechanism has

been proved to be feasible and effective. The work in next step is to develop the integrated system, including the machine vision, the multi-electromagnetic sensors, GPS-GIS and so on, in order to make the robot be autonomous.

Acknowledgement

This paper is supported by the National High Technology Research and Development Program of China under Grant NO: 2002AA420110, 2005AA420010, 2006AA04Z202, National Natural Science Foundation of China under Grant NO: 50575165. Thanks for the cooperation and support from Wuhan electric power Supply Company, Hubei province, Nanning electric power Supply Company, Guangxi power Grid Company.

References

1. Debenest, P., Guarnieri, M., Takita, K., Fukushima, E.F., Hirose, S., Tamura, K., Kimura, A., Kubokawa, H., Iwama, N., Shiga, F.: Expliner - Robot for inspection of transmission lines. In: IEEE International Conference on Robotics and Automation, 2008. ICRA 2008, pp. 3978–3984 (2008)
2. Montambault, S., Pouliot, N.: Design and validation of a mobile robot for power line inspection and maintenance. In: 6th International Conference on Field and Service Robotics - FSR 2007. Chamonix, France (2007)
3. Zhou, F., Li, Y., Wang, J., Xiao, H.: Design of the mechanical configuration of inspection robot for high voltage power transmission lines and analysis on its inverse kinematics with constraints, vol. 17, pp. 4–8 (2006)
4. Sun, C., Wang, H., Zhao, M.: Vibration analysis of obstacle-avoidance for EHV power transmission lines inspection robot. In: 33rd Annual Conference of the IEEE Industrial Electronics Society, IECON, United States, Taipei, Taiwan, November 5-8, pp. 2824–2828. Institute of Electrical and Electronics Engineers Computer Society, Piscataway (2007)
5. Wu, G., Xu, X., Xiao, H., Dai, J., Xiao, X., Huang, Z., Ruan, L.: A novel self-navigated inspection robot along high-voltage power transmission line and its application. In: Xiong, C.-H., Liu, H., Huang, Y., Xiong, Y.L. (eds.) ICIRA 2008. LNCS (LNAI), vol. 5315, pp. 1145–1154. Springer, Heidelberg (2008)
6. Zhang, X., Mao, E.: Study and Development of Mechanical System Virtual Prototyping. Journal of China Agriculture University (1999)
7. Xiong, Y.: Robotics. China Machine Press, Beijing (1993)

The Design of the GPS-Based Surveying Robot Automatic Monitoring System for Underground Mining Safety

Chenguang Jiang[1], Jianguo Peng[2], Chunqiao Yuan[3], Guohui Wang[3], Yong He[4], Shaohong Li[4], and Bo Liu[4]

[1] College of Environmental and Civil Engineering, Jiangnan University, Jiangsu Wuxi 214122, P.R. China
jiangcg999@yahoo.com.cn
[2] Communication Plan and Investigation and Design Institute of Hunan Province, Changsha, Hunan 410008, P.R. China
[3] Planning Management Bureau of Yantai City, Shandong Yantai 264003, P.R. China
[4] Qingdao Agricultural University, Shandong Qingdao 266109, China

Abstract. Earth subsidence in underground mining is an unavoidable problem in mining production, and timely and scientific observation and early warning is one of the important factors in the security of mining production. Though the surveying robot (i.e. automatic electronic total station) can automatically (or semi-automatically) monitor ground deformation for underground mining, the stability of the station location (monitor base station) has great impact on the monitor precision and when the measurement vision is covered, the surveying robot fails to monitor the corresponding deformation point. In order to tackle the above problem, the author and the research team have integrated the technology of GPS (Global Positioning System) with surveying robot and developed the GPS-based surveying robot automatic monitoring system for underground mining safety, which completely solves the foresaid problem, simplifies the monitor program and reduces the fixed investment cost of monitor. The article introduces the structure and working principle of the GPS-based surveying robot automatic monitoring system for underground mining safety, presents examples of monitor.

Keywords: Automatic monitoring technique, mining engineering, underground mining, ground deformation, integrated monitor system, research and manufacture of instrument, surveying robot, GPS technology.

1 Introduction

The subsidence of ground, as a kind of global artificial geological disaster, is resulted from complicated and diversified causes, including the underground mining. Earth subsidence in underground mining is an unavoidable problem in mining production, and timely and scientific observation and early warning is one of the important factors in the security of mining production[1]-[3]. The process of subsidence of ground

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 1156–1162, 2009.

caused by underground mining is a complex systematic problem. So it is of great realistic significance to accurately forecast the development trend of the subsidence of ground led by mining so that people can avoid its harm. The author and the research team have integrated the technology of GPS (Global Positioning System) with surveying robot and developed the GPS-based surveying robot automatic monitoring system for underground mining safety. Here we will introduce it particularly.

2 The GPS-Based Surveying Robot Automatic Monitoring System for Underground Mining Safety

The GPS-SR-AMS-MS(GPS-based surveying robot automatic monitoring system for underground mining safety) consists of two parts, namely the host and auxiliary equipment. For the host of the GPS-SR-AMS-MS, please see Fig. 1.The host of the GPS-SR-AMS-MS is composed of three components, namely surveying robot, 360°omni-directional reflection prism and RTK-GPS receiving antenna. Besides, it includes the supporting wireless data transmission system, distribution system, GPS communication system and so on.

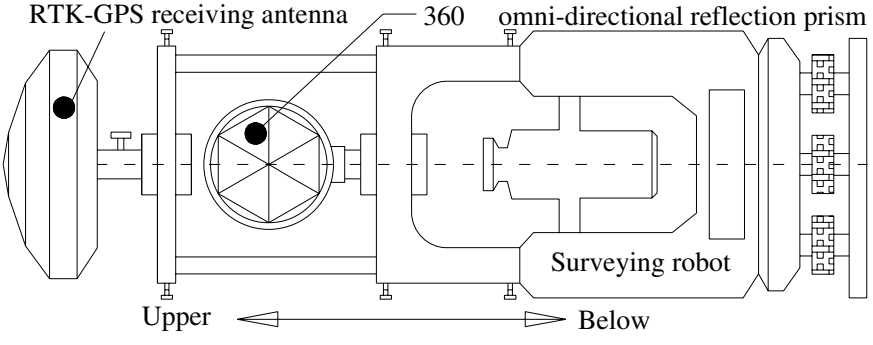


Fig. 1. Sketch of the host of GPS-SR-AMS-MS

The independent three-dimensional coordinate system must be adopted in application of the GPS-SR-AMS-MS. The plane coordinate system adopted should be the Gauss plane right angle coordinate system based on WGS-84 elliptical ball, central meridian be the meridian crossing the center of mining area, distance projecting plane be the water level over the mining area top and height system be the GPS geodetic height system.

2.1 Basic Characteristics of Surveying Robot

The author terms the intelligent electronic total station with the function of automatic tracking as surveying robot, which prides in functions of motor drive, laser point search, CCD technology, automatic target recognition, automatic collimation, automatic angle measurement, automatic distance measurement, automatic target tracking

and automatic record, has the standard data interface (such as RS232 serial interface) or Modem, and can achieve synchronic data communication between field operation and distant office. The surveying robot has the function of self-study. Currently its measurement speed can reach several points per second and its precision of three-dimensional positioning can be better than ±1mm. It high precision results from the standard length comparison technology (or termed as distance difference technology) between base stations, which can get rid of or reduce the air refraction error and the inborn systematic error of the instrument and equipment. The core technology of surveying robot is automatic target recognition and tracking system (ATR--Automatic Target Recognition).

2.2 Basic Technical Features of RTK-GPS

GPS (Global Positioning System) technology is the second generation satellite navigation system invented by the U.S.A. RTK-GPS technology is one of the important progresses of GPS technology. It positioning speed has reached within 2 seconds and its positioning precision has achieved sub-millimeter level. If the difference technology is adopted, its three-dimensional positioning precision can reach millimeter level.

The operation method of RTK-GPS technology is as follows. First, a fixed GPS base station must be set up or installed. Second, one or more mobile GPS receivers (referred to as mobile station) are adopted in the control area of base station (usually a round area 10km-20km away from the base station) for instant positioning. The base station achieves synchronic contact with the mobile station through data transmission radio to get the three-dimensional coordinate information of each instant of the mobile station. When GPS—RTK is monitoring, the base station should be high in location and no GPS signal reflective body or objects possible to cause multi-channel effect (such as imposing buildings, large area of water, high voltage wire, communication equipment, microwave station and so on) can show up within 200 meters around. One outstanding advantage of GPS—RTK is no requirement of visibility between GPS base station and GPS mobile station.

At present, RTK-GPS technology can conduct data transmission through radio, GSM mobile phone, fax and Modem and available communication formats include CMR, CMR+ and RTCM.

3 Monitor Process of GPS-SR-AMS-MS

Fig. 2 is the arrangement of lazitun coal mine ground deformation monitor points. There are many ground deformation monitor points located on the mining area. See a1, …, a5, b1, …, b6, c1, c2, d1, …, d3 in Fig. 2. Q and P are the deformation monitor mutual visibility double base station. B is the monitor three-dimensional base origin. A and C are the checkpoints of three-dimensional base origin.

3.1 Monitor Field Arrangement of GPS-SR-AMS-MS

Lazitun coal mine GPS-SR-AMS-MS adopts Leica TCA2003 surveying robot and Leica SR530 GPS receiver. The GPS-SR-AMS-MS is composed of a three-dimensional control base system and a mining area three-dimensional deformation point system. See

Fig. 2. The three-dimensional control base system should be a triangular base control network far from the mining area and with good structure (the minimum distance from each control point to the mining area should be larger than 6.7km). One point should be used as the three-dimensional base origin and the other two only as the checkpoints of the three-dimensional base origin or as the backup points in case of mining area age to the three-dimensional base origin. The three-dimensional base origin should set up a compulsive centering observation pier, which can link the reflection prism, electronic total station, GPS receiving antenna and other measurement equipment, instrument or components through compulsive centering. The mutual distance between base points should be accurately measured in the precision no less than 1/700000 and the precision of three-dimensional point position of the base point should be better than ±0.3mm.

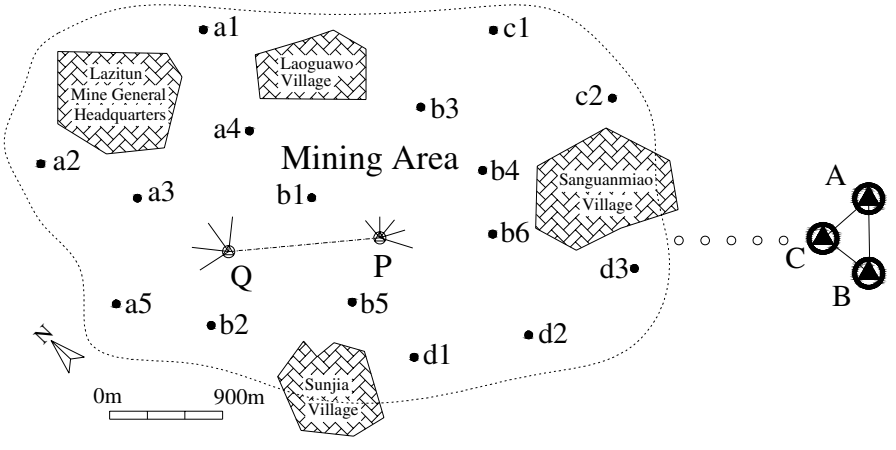


Fig. 2. Sketch of arrangement of Lazitun coal mine ground deformation monitor points

The control points of the base control network should adopt deeply covered steel pipe signs or be directly fixed on base rocks. It is appropriate to use simple network structure and high intensity graphs for base control network. Single triangle and geodetic quadrangle are recommended network structures. The monitor control network should be tested and verified regularly. Points A, B and C of Lazitun coal mine base control network are all directly filled and covered on the hard rock at the mountain peak and the base control network adopts GPS static measurement technology to observe. The observation lasts for 2 hours (in adoption of synchronic closed circuit observation mode by three GPS receivers). Select one or several points at the mining area or around the mining area as the monitor base stations (see Q and P in Fig. 2). It is required that all the three-dimensional deformation points set up on the mining area can be monitored through these monitor base stations and that each monitor base station must share visibility with at least another one monitor base station (the distance between the two monitor base stations sharing visibility should not exceed 90%

of the valid measurement distance of the surveying robot). The number of monitor base stations is decided by whether all the three-dimensional deformation points can be monitored. There is only one monitor three-dimensional base origin (i.e. B) in Lazitun coal mine ground deformation monitor system and points C and A are the checkpoints or backup points for the three-dimensional base origin B. There are only 2 (i.e. Q and P) monitor base stations, which are located at the mining area and share visibility (termed as mutual visibility double base station by the author). The three-dimensional deformation point system of the mining area consists of several deformation observation points. The deformation observation point should set up a compulsive centering observation pier, which can link the reflection prism through compulsive centering. The reflection prism linked on the observation pier should share visibility with the corresponding host of the GPS-SR-AMS-MS and its distance to the host of the GPS-SR-AMS-MS should not exceed 90% of the valid measurement distance of the surveying robot.

3.2 Automatic Monitor Process of GPS-SR-AMS-MS

When monitoring, place the base station GPS receiver compulsive centering on the three-dimensional base origin (i.e. B) and two hosts of the GPS-SR-AMS-MS on the two monitor base stations (i.e. Q and P) of the mining area (in case of more than one monitor base station, one host of the GPS-SR-AMS-MS should be placed on each monitor base station). The GPS receivers on two (or more) hosts of the GPS-SR-AMS-MS are mobile receivers, continuously receiving GPS satellite information and GPS base station information and getting the instant three-dimensional coordinate of monitor base station. Since the ground of mining area is in the movement of small breadth and low speed, the instant three-dimensional coordinate of monitor base stations (Q and P) at the mining area in the whole process of monitoring may have small changes.

Two hosts of the GPS-SR-AMS-MS provide rearview for each other. First, input the instant three-dimensional coordinate of the two monitor base stations in the two hosts of the GPS-SR-AMS-MS (make a distinction of lateral station coordinate and rearview coordinate). Second, aim at the reflection prisms of each other by hand to fix the rearview direction. Third, the surveying robot will aim automatically. After the rearview direction is fixed, observe the reflection prisms permanently placed at the deformation observation points each system host is responsible for (on account of the automatic aiming function) to get the initial three-dimensional coordinate (X, Y, Z) at each deformation point, which is used to train the surveying robot (observation program design, height of instrument, height of target, times of measurement, change face, time interval, communication parameter and so on). Then the surveying robot can make automatic cycle measurement of the three-dimensional coordinate according to the set time intervals for each deformation point.

Each monitoring process is described as follows: automatically input the two instant three-dimensional coordinates obtained from RTK-GPS system (minor change for each time); two hosts of the GPS-SR-AMS-MS automatically aim at the reflection prism of each other to get each other's three-dimensional coordinate; automatically

aim at the reflection prisms permanently placed at the deformation observation points each system host is responsible for to get the initial three-dimensional coordinate……The operation order and procedure for each time are completely the same and the monitor for each time is automatically carried out without interference from people. The difference lies in that the three-dimensional coordinates of observation points for each monitor are different. It is this difference that reflects the change in space and position of each deformation observation point. The change volume between the neighboring two monitor three-dimensional coordinate for the same deformation observation point is the relative change volume between the two observations at this point position. The difference between the initial three-dimensional coordinate and the three-dimensional coordinate for each monitor at the same deformation monitor point is the absolute change volume at the point position (or termed accumulative change volume). The host of GPS-SR-AMS-MS can automatically save all the monitor results in its own memory and transmit all the information and result to the monitor center office through data communication equipment. The monitor center office can control (send instructions, modify data, change the observation program and so on) the host of GPS-SR-AMS-MS at any time through data communication equipment. The computer in the monitor center office can analyze, draw pictures for or process the monitor data through special software, make long-distance diagnosis of the safety of the mining production.

4 Conclusion

The GPS-SR-AMS-MS has been tested on 9 different mining areas and good effects have been achieved. Its average absolute monitor error of three-dimensional deformation is better than ±1.37mm and the absolute monitor error of maximum three-dimensional deformation is 3.02mm, which can fully satisfy the precision requirement of security monitor of various mining areas. Since 2 (or more) hosts of GPS-SR-AMS-MS can provide rearview for each other, the monitor process can get ride of excessive reliance on the base point (that is, Q and P can be flying points without point position signs, which can be decided randomly and the host of GPS-SR-AMS-MS can be put anywhere so as to achieve free station monitor in real terms. But the origin B of three-dimensional base, checkpoint or backup point C and A are necessary and indispensable.) and the area and scope of monitor will be enlarged. The methods introduced in this article are all realistic in deformation monitor and security monitor of various mining areas, which can effectively reduce the monitor cost, ensure the monitor quality and guarantee the operation safety of mining areas.

Acknowledgment

The research project discussed in this paper has been funded by the National Natural Science Fund of China (No.52793096) and Jiangnan University 211 Construction Fund (No.2004012) . Sincere gratitude is presented to them.

References

1. Parkinson, W., Bradford, S.J., Enge, P.: Global Positioning System: Theory and Applications. AIAA Washington DC (1996)
2. Mining Manual Compilation Group. Mining Manual, pp. 36–47. Metallurgy Industry Publishing Company, Beijing (1992)
3. Chen-guang, J., Zhong-ping, J., Ji-ming, W., Ji-ren, P., Hong, Z.: Contributions of the Safety Surveying on the dam of Hydropower Station to the Sustainable Development of the Economy in the Reservoir Area. In: The Proceedings of the United Nations Symposium on Hydropower and Sustainable Development [CD-ROM] (2004)

Author Biographies

The first Author: Chen-guang Jiang, Male, Professor, Was born in Feb. 1964, Yueyang City, Hunan Province. He is an evaluation expert of National Scientific and Technological Prize of China, an evaluation expert of National Natural Science Fund of China, and an evaluation expert of Scientific and Technological Periodical of China. He has won the Forth Youth Science and Technology Prize of Shandong Province(1996), the First Youth Science and Technology Prize of Yantai City(1995), the Second Prize of Shandong Building Industry Scientific and Technical Advancement (the 1st one,1999).

Address: College of Environmental and Civil Engineering, Jiangnan University, Jiangsu Wuxi 214122, Jiangsu Province , P.R.China

E-mail: jiangcg999@yahoo.com.cn
Or jiangcg999@sina.com

Acquisition of Movement Pattern by Q-Learning in Peristaltic Crawling Robot

Norihiko Saga and Atsumasa Ikeda

Kwansei-Gkuin University, Department-of-Human-System-Interaction,
2-1-Gakuen, Sanda, Hyougo, Japan
cfp04548@ksc.kwansei.ac.jp, saga@kwansei.ac.jp

Abstract. A peristaltic-crawling robot composed of many servo motors was developed. This robot can move equivalently to an earthworm using peristalsis, with expansion and contraction of segments, and also using contact sensors. The movement pattern was designed based on the peristaltic motion of an earthworm. The optimum movement pattern for a peristaltic-crawling robot was acquired through Q-learning. Its effectiveness was confirmed by mounting it onto a real robot.

Keywords: A peristaltic-crawling robot, Q-learning.

1 Introduction

Peristaltic movement presents advantages—such as mobility that is flexibly adapted to complicated environments such as irregular places and with mobility in narrower spaces—over conventional wheels and meandering movement. [1]- [6] The need exists for development of peristaltic-crawling robots that are adapted to mobility in various environments such as irregular surfaces, rubble, and inside pipes, while taking advantage of these features. The movement pattern of robots must be developed concomitantly. With Q-learning, a reinforcement learning method, [7]-[12] the movement pattern of forward motion of a peristaltic crawling robot is acquired; its operation was verified using a real robot.

2 Q-Learning

A typical algorithm of reinforcement learning, Q-learning, is a learning algorithm that selects an action having the highest value of a behavioral assessment function (Q factor) from among possible actions on each state throughout repeated trial-and-error interaction with the environment. Figure 1 depicts an outline of it.

Boltzmann selection is an action selection method for Q-learning. It performs action selection at a fraction that is proportional to $\exp(Q(s,a)/T)$.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 1163–1169, 2009.

[1] Initialize matrix **Q** as a zero matrix.
[2] An agent observes state s_t of the environment in time step t.
[3] An agent performs actions according to arbitrary action selection methods.
[4] Get reward r_t from the environment.
[5] Observe state s_{t+1} after state transiting.
[6] The Q value is updated based on the Q-learning update.

$$Q(s_t, a_t) \leftarrow (1-\alpha)Q(s_t, a_t) + \alpha[\ r + \gamma \max_a Q(s_{t+1}, a)\ ] \qquad (1)$$

In that equation, r_t is the reward given at time t, $\alpha_t(s,a)$ $(0 < \alpha \leq 1)$ the learning rates, might be the same value for all pairs. The discount factor γ is such that $0 \leq \gamma < 1$.

[7] Time step t is advanced to time step t+1; it returns to step [3].

Fig. 1. Q-learning algorithm

3 Simulation Method

3.1 Peristaltic Crawling Robot

Figure 2 presents a peristaltic crawling robot used for verification of movement patterns acquired through simulation. This robot has segment units that are equivalent to earthworm segments, which carry out peristaltic movement with expansion and contraction. The number of segment units is determined as three, which is the minimum number necessary for peristaltic movement.

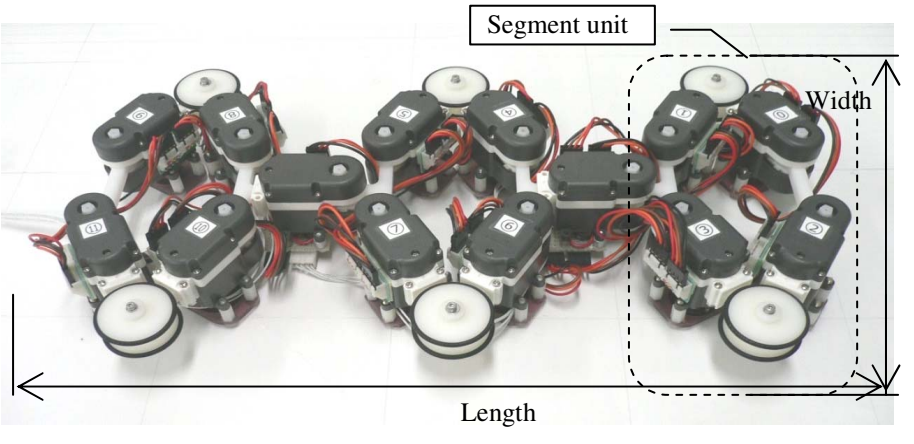


Fig. 2. Peristaltic crawing robot using servo motors

3.2 Simulation Model

The model of a peristaltic crawling robot is a simple model using an algorithm, not using a kinetic model. The robot is assumed to conduct advance operations within a 2D plane. The traveling plane is 160 mm wide, and the target position is 3000 mm distant from the start position.

Angles in the segment model are defined as presented in Fig. 3; each angle is constrained as

$$\begin{cases} \theta_3 = \theta_4 = \theta_5 = \theta_6 \\ \theta_1 = \theta_2 \end{cases}. \tag{2}$$

This constraint provides a unique form of a segment given θ_1. Therefore, θ_1 is represented hereinafter as Θ; the movement pattern of the robot is acquired by computing the pattern of segment angle Θ for every segment in the simulation.

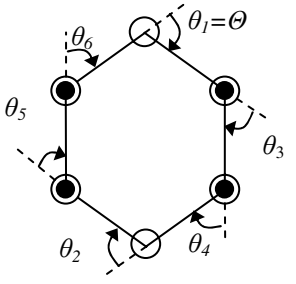


Fig. 3. Angle of segment model

The robot is presumed to operate according to the following conditions.

(A. 1) The robot's central line does not move. Its center shall not move unless a segment contacts the plane.
(A. 2) When a robot's segment contacts the plane, let the position of contact be a fixed position.
(A. 3) When multiple nonadjacent segments contact the plane, if a segment not contacted by the plane is going to operate, suspend it and return a penalty.

3.3 Application of Q-Learning

Segment angle Θ varies six ways: 0, 20, 40, 60, 80, and 100 deg. Because the robot has three segments, the segment form of the robot includes 6^3 combinations. Let this segment shape be state s, and let operations for transfer into state s be action a. A Q-table therefore has size of $6^3 \times 6^3$.

Let the distance from the robot's tail on time t [step] to the target position be d_t [mm]; reward r_t is defined as

$$r_t = d_{t-1} - d_t. \tag{3}$$

A penalty is given in the form of reward r_t = -50.

Action selection takes place according to Boltzmann selection. The temperature constant T is defined as

$$T = 150 \times 0.9934^{i-1}, \tag{4}$$

where i is the number of trials.

One trial comprises up to 40 steps. The simulation is repeated up to 1000 trials. When a penalty occurs, a trial is terminated even at fewer steps than 40. Let the learning rate α be 0.6, with discount rate γ 0.3, sampling frequency 20 Hz, and time interval 0.3 s.

4 Simulation Result

The sum of rewards for every trial on the simulation is presented in Fig. 4. The total reward increases concomitantly with the increasing number of trials, suggesting that learning is performed so that the sum of rewards increases.

Figure 5 depicts a movement pattern acquired from the simulation. This movement pattern represents an action selected from Q-table as having the largest Q factor after completion of 1000 trials. Clearly, two actions, actions A and B, are repeated after step 2, as portrayed in Fig. 6. Action A is a settling operation to fix the whole robot's position, whereas action B is an advancing operation to generate great distances and move forward. For that reason, the moving distance at a robot's tail becomes stepwise

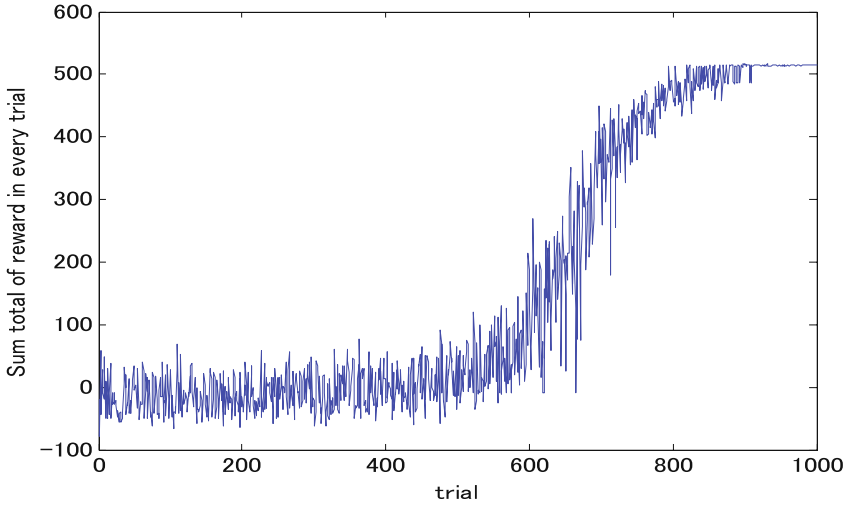


Fig. 4. Sum of rewards in every trial

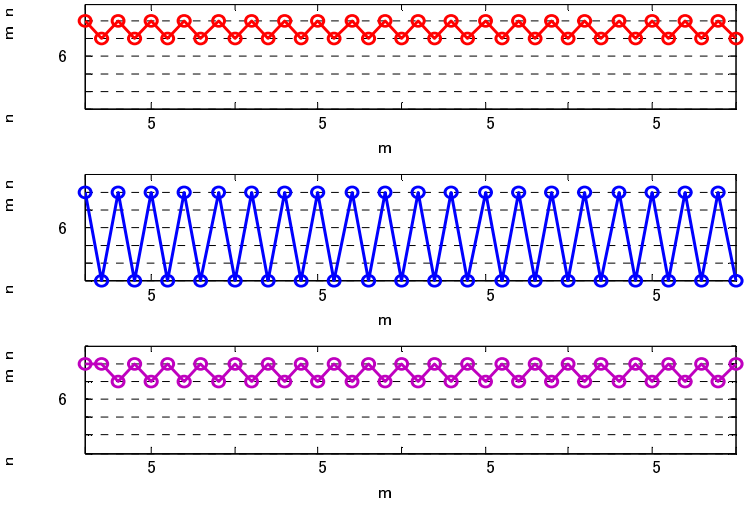


Fig. 5. Acquired motion pattern of the segment angle

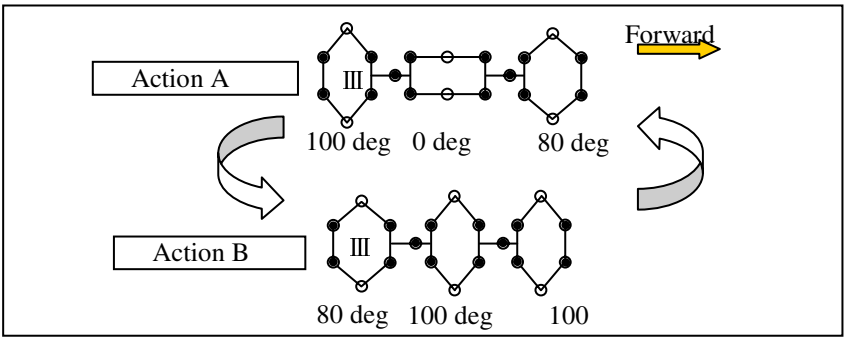


Fig. 6. Acquired motion pattern

as portrayed in Fig. 7. Because segments I and III are used for the whole robot's fixation, segment angles change little by little in actions A and B, whereas segment II generates distance, so that its segment angle changes greatly.

5 Mounting on Robot

Figure 7 presents the moving distance of the robot with the acquired movement pattern installed as the operation of the robot. This moving distance is determined with the robot's tail as the point of measurement demonstrating that the movement pattern acquired in the simulation can operate a robot. The moving distance of the robot is longer than the simulation result. This is considered to be true because the robot advanced more than anticipated because the segment was not fixed actually when it contacts a wall.

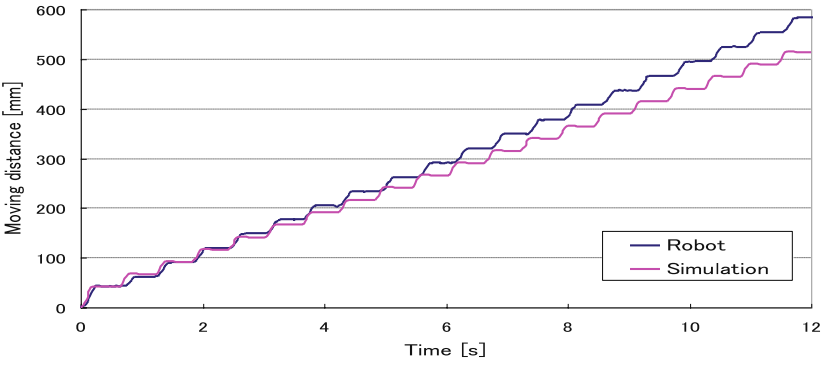


Fig. 7. Moving distance of acquired motion pattern

6 Conclusion

This study yielded the following conclusions.

(1) The movement pattern of a peristaltic crawling robot was acquired by the simulation using Q-learning.
(2) The acquired movement pattern requires fewer operations.
(3) The mobility of a robot with the acquired movement pattern was verified. The simulation validity was confirmed.

References

1. Long, G., Anderson, J., Borenstein, J.: The Kinematic Design of the OmniPede: A New Approach to Obstacle Traversion. In: Proceedings of the 2002 IEEE International Conference on Robotics and Automation, Washington DC, May 10-17, pp. 714–719 (2002)
2. Quillin, K.J.: Kinematic Scaling Of Locomotion By Hydrostatic Animals: Ontogeny Of Peristaltic Crawling By The Earthworm Lumbricus Terrestris. The Journal of Experimental Biology 202, 661–674 (1999)
3. Moon1, B.R., Gans, C.: Kinematics, Muscular Activity And Propulsion In Gopher Snakes. The Journal of Experimental Biology 201, 2669–2684 (1998)
4. Ono, G., Kato, S.: Research on air pressure drive type earthworm jurisdiction inspection robot. Precision engineering association science 2005A No. SPACE, 171–172 (2005)
5. Quillin, K.J.: Ontogenetic Scaling Of Hydrostatic Skeletons: Geometric, Static Stress And Dynamic Stress Scaling Of The Earthworm Lumbricus Terrestris. The Journal of Experimental Biology 201, 1871–1883 (1998)
6. Suleman, A., Burns, S., Waechter, D.: Design and modeling of an electrostrictive inchworm actuator. Mechatronics 14, 567–586 (2004)
7. Barto, A.G., Bradtke, S.J., Singh, S.P.: Real-time learning and control using asynchronous dynamic programming (COINS technical report 91-57). University of Massachusetts, Amherst (1991)

8. Mahadevan, Connell: Automatic programming of behavior-based robots using reinforcement learning. In: Proceedings of the 1991 National Conference on AI, pp. 768–773 (1991)
9. Werbos, P.J.: Advanced forecasting methods for global crisis warning and models of ntelligence. General,Systems Yearbook 22, 25–38 (1977)
10. Sato, M., Abe, K., Takeda, H.: Learning control of finite Markov chains with explicit trade-off between estimation and control. IEEE Transactions on Systems, Man and Cybernetics 18, 677–684 (1988)
11. Sutton, R.S.: Learning to predict by the methods of temporal difference. Machine Learning 3, 9–44 (1988)
12. Omori, H., Hayakawa, T., Nakamura, T.: Locomotion and Turning Patterns of a Peristaltic Crawling Earthworm Robot Composed of Flexible Units. In: 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems Acropolis Convention Center, Nice, France, September 22-26 (2008)

The Study on Optimal Gait for Five-Legged Robot with Reinforcement Learning

Adnan Rachmat Anom Besari[1], Ruzaidi Zamri[1], Anton Satria Prabuwono[2], and Son Kuswadi[3]

[1] Universiti Teknikal Malaysia Melaka,
Locked Bag 1200, Hang Tuah Jaya, 75450 Ayer Keroh, Melaka, Malaysia
[2] Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor D.E., Malaysia
[3] Electronic Engineering Polytechnic Institute of Surabaya (EEPIS), Institut Teknologi Sepuluh Nopember, Kampus ITS Keputih Sukolilo Surabaya 60111, East Java, Indonesia
{adnan_ra,ruzaidi}@utem.edu.my, antonsatria@ftsm.ukm.my, sonk@eepis-its.edu

Abstract. The research of legged robot was rapidly developed. It can be seen from recent ideas about new systems of robot movement that take ideas from nature, called biology inspired. This type of robot begins replacing wheeled robot with various functions and interesting maneuvers ability. However, designers should decide how many legs are required to realize the ideas. One of the ideas that are rarely developed is odd number of legs. This research focused on five legs robot that inspired from starfish. To realize the intelligent system in robot that does not depend on the model, this research used reinforcement learning algorithm to find the optimal gait when robot is walking. In order to achieve this goal, trial and error have been used to provide learning through an interaction between robot and environment based on a policy of reward and punishment. The algorithm is successfully implemented to get the optimal gait on a five-legged robot.

Keywords: Five-legged robot, gait optimizing, reinforcement learning.

1 Introduction

At the first stage in design of a walking machine, designers should decide how many legs are required to realize the ideas. Presently, biped, quadruped and hexapod machines occupy large volume in researches of practical walking robot. It is debatable which number of the legs is better, four or six. We should select the number of legs carefully by considering their locomotive environments, because each walking machine has peculiar merits. In case of heavy walking environments, hexapod walking may be more suitable than quadruped walking. If one of the six legs is broken down, continuation of the static walking may be ensured by the left five legs, while some parts of walking functions are reduced. In case of quadruped walking robot, it may not be possible to continue their static walking.

According to bionics, six legs/hexapods robot may be a better choice. On the other hand, almost all insects have six legs [1]. It easy to keep balanced with a redundant locomotion system, which increases reliability and workable even if one, two or three

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 1170–1175, 2009.

legs broken. Hexapod makes it possible for the robot to use one, two or three legs to work as hand and perform complex operations.

2 Related Work

As the reasons why the gait study of walking with odd number of legs is very few as compared with the gait study for even number of leg, the following matters are considered. Although we can easily find probable gait for the walking with even number of legs by observation of insects etc., this method is not available for the case of the walking with odd number of legs. In the fact, the wave gait which is for the walking with even number of legs is originated from the observation of insects' walking, and the above mentioned analytical proof is performed with an assumption that the wave gait is optimally stable. It is very difficult to come to the same conclusion by use of analytical method without the assumption of the wave gait [2].

Studies of the gait of the partially broken hexapod machine have been discussed from biological observations. Inagaki and Kobayashi [3] described in hexapod walking robot, it is considered that even if one of six legs is disabled, the remaining five legs may maintain static walking. However, to maintain the static stability at maximum, gait study for five-legged walking is necessary. They developed a method of gait study for five-legged walking, which is considered a case that hexapod walking robot damaged in one of six legs. The allowed leg arrangement, which means the arrangement of the center positions of stroke motion, to maintain the gait stability margin is only limited to the regular pentagonal shape. Thus, in the case that a hexapod walking robot which has general leg arrangement such as a rectangle type is damaged, the leg arrangement must be changed to the pentagonal shape. Thus, these gaits are enough useful as a gait for five-legged walking. Nowadays, the studies of five legs robot not just inspired from hexapod walking with disabled leg but inspired from nature that got the idea from starfish. Makoto et al. [4] focus Brittle Star realizes a flexible myriapod robot. Brittle Star is a kind of the Echinoderm has five long legs. They developed the five legs robot, and inquire forward motion it.

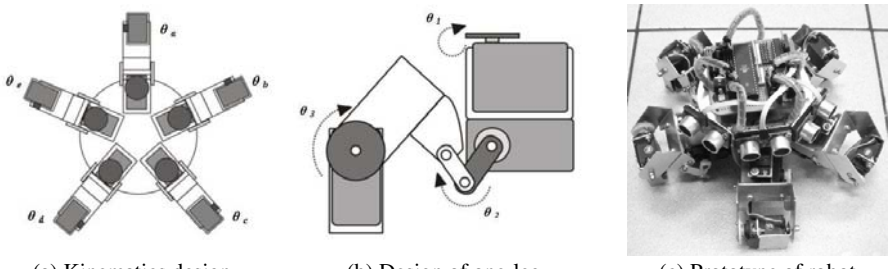
On the other hand, synchronized motion is very important for the multiple legged machines. This paper proposes a synchronized motion method that utilizes the notion of autonomous distributed systems. For the method, we have reviewed how quadrupeds [5], [6], [7], [8] and hexapods [9], [10], [11], [12], [13] on locomotion aspects and compared those methods. Most of the methods use degree of freedom (DOF) settings as the main parameter of walking robot. Finally, we show the total motion control can generate the gait automatically and change them smoothly by using reinforcement-learning algorithm.

3 System Design

Most of the systems use the natural movement of the whole number of feet (two, four, six and eight). Five-legged in the design of nature can be found on the starfish. The design is inspiring authors to develop this type of robots. The learning of five-legged walking is realized with the actual robot. This section introduces the system design of robot and application of reinforcement learning for this purpose.

3.1 Mechanical and System Design

Mechanical design of robot developed with five-legged actuator. It has five legs, and on every corner, there is three degree of freedom as a reference for the movement of robots. The kinematics system design and prototype of robot are shown in Figure 1.



(a) Kinematics design (b) Design of one-leg (c) Prototype of robot

Fig. 1. Design and prototype of five-legged robot

The system is divided into several sections: actuator and sensor selection, mechanical and hardware building, and next reinforcement learning algorithm implementation in robot. The type of actuator that selected is standard servo-motor (180 degree rotation) with a high enough torque (4.1kg.cm/56.93oz.in). Limit switch is used to detect position of the robot. Limit switch installed in each leg. When the limit switches are touching objects, it will send a data series (USART) to microcontroller to drive servo motor.

3.2 Learning Algorithm

The reinforcement learning makes the robot adapt to more stable gaits in normal conditions of no external effect of instability. As a result, the stability of walking with a commanded speed is continuously increased till a pattern having good stability margin with each state is achieved. When the speed command is changed, the robot performs a smooth passage with probably slowing down to the speed of 1 unit per step, and then adapting to the new speed in a few steps. This smooth passage between commanded speeds should be noted as a peculiar feature of the algorithm.

Among the possible gaits with a given speed there are some applications with more static stability, while there are some applications that hardly satisfy the stability condition. Improvement of stability can be achieved by learning with the best state transitions are memorized and utilized. In this application the robot is not aware of the cause, but experiences falls although it generates statically stable gaits. The falls result in negative reinforcement and the robot learns to avoid them by memorizing and making use of the stable states.

In this design, reinforcement learning discussed to determine the optimal movement in robot walking. Reinforcement learning algorithms that are generally used apply several parameters that change according to the learning process in robot. In each of parameters, there are functions that are defined first. Angle on hinge base and delay are

parameters that have been chosen. These parameters changed the established value. The program that created generally divided into walking description includes the determination of the motions that can be done by robot. It includes testing of learning function that gives the optimal results in walking in the field with several conditions.

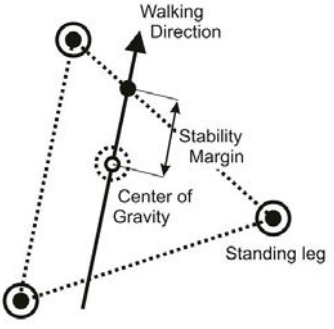


Fig. 2. Stability margin

Angle joint on hinge base of the legs, as shown in Figure 3, is referred to the point that can be reached by leg movement of joints that are closest to the body of robot (center of body). Because there are five legs, between the legs with one leg to the other leg is 72° away. So that each leg can move 72° (clockwise) and -72° (counter clockwise). But in reality, robot can only reach -50° to 50° because of mechanical robot legs require some space. Velocity of robot is resulted by time delay every movement joint on the robot legs. This parameter is given on the value of the function of delay in the microcontroller program. Before calling this function, interrupt will be turned off, and time delay will be longer than expected. Assessment is done by giving a "reward" when the movement of robot that have been produced in accordance with the environment condition, or by giving a "punishment" when the movement that produced robot are not appropriate. If "punishment" is received, means punishment point is added and the value of the parameters changed. If "reward" is received, means reward point is added and the value of parameters stored in memory. Every step described by the recursive functions.

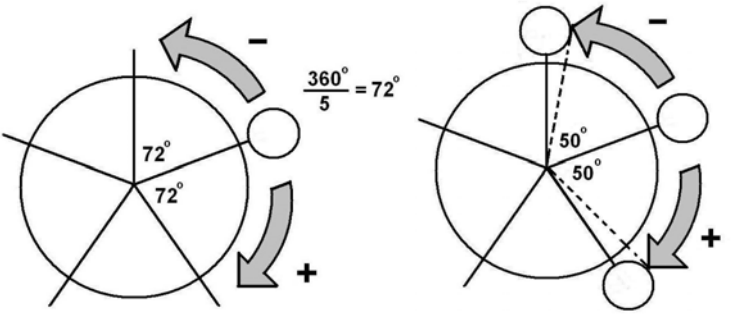


Fig. 3. Angle joints which formed the base of leg and the reduction

4 Result

Learning algorithm has been tested in the actual conditions. The robot is tested on several different levels of slope (0°, 10°). Error slope has been chosen as parameter that can be compared in measuring level of learning success. This parameter becomes the benchmark to get the most effective and efficient gait of robot. From here, we can conclude that the program algorithms can be created to run the system properly in accordance to criteria that are expected. Testing results in several different levels of slope that compared in error slope parameter are shown in Figure 4.

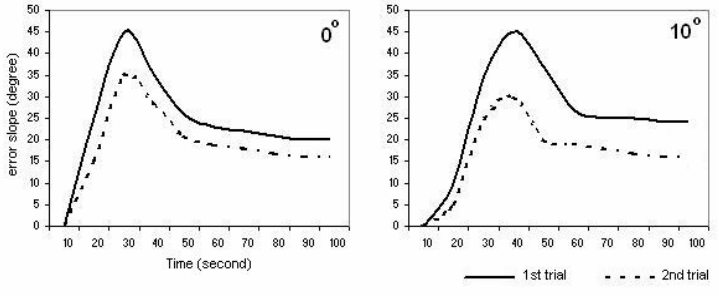


Fig. 4. Walking on several different levels of slope (0°, 10°)

5 Conclusion

This research focuses on five legs robot inspired by starfish. To realize the intelligent system in robot does not depend on the model, reinforcement learning algorithm used to find the optimized gait in walking robot. To achieve this goal, trial and error is used to provide learning through an interaction between robot and environment based on a policy of reward and punishment.

The speed on the walking robot depends on time delay, the greater delay reduces speed. Furthermore, error slope depends on the speed of movements. More rapid movement of servo motor causes greater error slope. Smaller angle that formed made greater error. In order to obtain the maximum results required a large torque in servo motor, so that the robot can hold the movement. Synchronization of movement between the legs on the robot can affect by considering time delay between gaps of movements. Locomotion of five-legged robot is shown in Figure 5.

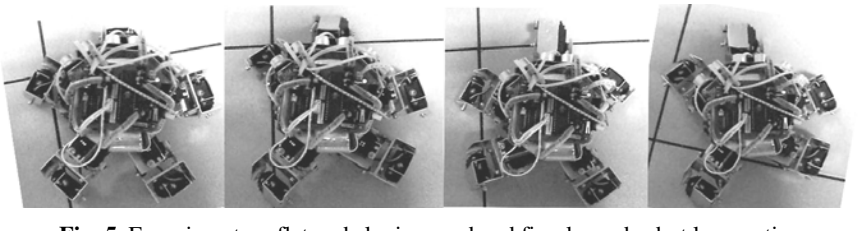


Fig. 5. Experiment on flat and sloping road and five-legged robot locomotion

Acknowledgments. The authors would like to thank Electronic Engineering Polytechnic Institute of Surabaya (EEPIS) supported by Japan International Cooperation Agency (JICA) under PREDICT Project for developing the five-leg robot, Universiti Teknikal Malaysia Melaka (UTeM) and Universiti Kebangsaan Malaysia (UKM) for financial support and providing facilities.

References

1. Wang, Z., Ding, X., Rovetta, A.: Structure Design and Locomotion Analysis of a Novel Robot for Lunar Exploration. In: 12th IFToMM World Congress, Besançon, France, June18-21 (2007)
2. Inagaki, K.: Gait study for hexapod walking with disabled leg. In: Proceedings of the 1997 IEEE/RSJ International Conference on Intelligent Robot and Systems, 1997. IROS 1997, September 7-11, vol. 1, pp. 408–413 (1997)
3. Inagaki, K., Kobayashi, H.: Adaptive wave gait for hexapod synchronized walking. In: Proceedings IEEE International Conference on Robotics and Automation, May 8-13, vol. 2, pp. 1326–1331 (1994)
4. Makoto, T., Koji, Y., Satoshi, E.: Studies on Forward Motion of Five Legs Robot (Nippon Kikai Gakkai Robotikusu, Mekatoronikusu Koenkai Koen Ronbunshu). Journal Code: L0318B 2005, 2P1-S-065 (2005)
5. Kimura, H., Yamashita, T., Kobayashi, S.: Reinforcement Learning of Walking Behavior for a Four-Legged Robot. In: Proceedings of the 40th IEEE, Conference on Decision and Control, Orlando, Florida USA (December 2001)
6. Lewis, M.A., Bekey, G.A.: Gait Adaptation in a Quadruped Robot. Journal Autonomous Robot 12(3), 301–312 (2002)
7. Chernova, S., Veloso, M.: An Evolutionary Approach to Gait Learning For Four-Legged Robot. In: Proceedings of International Conference on Intelligent Robot and Systems 2004 (IROS 2004), Sendai, Japan (September 2004)
8. Kohl, N., Stone, P.: Policy Gradient Reinforcement Learning for Fast Quadruped Locomotion. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), New Orleans, LA, May 2004, pp. 2619–2624 (2004)
9. Zennir, Y., Couturier, P., Bétemps, M.: Distributed Reinforcement Learning of Six-Legged Robot to Walk. In: The Fourth International Conference on Control and Automation (ICCA 2003), Montreal, Canada, June 10-12 (2003)
10. Zennir, Y., Couturier, P.: Multiactor approach and hexapod robot learning. In: Proceedings 2005 IEEE International Symposium on Computational Intelligence in Journal of Robotics and Automation, Espoo, Finland, June 27-30, pp. 665–671 (2005)
11. Barfoot, T.D., Earon, E.J.P., D'Eleuterio, G.M.T.: Experiments in learning distributed control for a hexapod robot. Journal of Robotics and Autonomous Systems 54, 864–872 (2006)
12. Haynes, G.C., Rizzi, A.A.: Gait Regulation and Feedback on a Robotic Climbing Hexapod. In: Proceedings of Robotics: Science and Systems (August 2006)
13. Erden, M.S., Leblebicioglu, K.: Free gait generation with reinforcement learning for a six-legged robot. Robotics and Autonomous Systems 56, 199–212 (2008)

Study on External Load Domain of Shield Machine Cutterhead

Weili Wen, Pingfa Feng, Zhijun Wu, and Jianhu Liu

Department of Precision Instrument and Mechanology,
Tsinghua University, 100084 Beijing, China
wlwen@mail.tsinghua.edu.cn, fengpf@mail.tsinghua.edu.cn,
wuzhijun@mail.tsinghua.edu.cn, liujh07@mails.tsinghua.edu.cn

Abstract. The shield machine cutterhead load domain in composite strata is an important and complex problem, which has significant impact on the design of the cutterhead. In this paper, the load model of the cutterhead including resistance and resistance moment is developed through the micro-unit method under composite strata, and responses of cutterheads with different open ratios are analyzed by numerical simulations with the developed load model.

Keywords: Cutterhead, Load model, Composite strata, Micro-unit, Numerical simulation.

1 Introduction

The shield machine is often required to work under the environment of mixed ground, and the research about the shield machine under composite strata has obtained more and more scholars' attentions [1-8]. The determination of the cutterhead load domain is the basis for cutterhead selection, design and check. Currently, the usual researches mainly focus on a single ground layer. Lu et al. [9, 10] researched the cutterhead system moment of the earth pressure balance shield machine; Guan [11] studied the shield machine front resistance, lateral resistance, and torque; Zhu et al. [12] studied the thrust force of the shield machine; Zhan et al. [13] studied the front and lateral resistance of the shield machine; Song et al. [14] studied the shield machine thrust and the cutterhead system moment in the sandy cobble layer without water. In practice, shield machines are often operated in composite strata, and the loads are various with different strata and cutterhead rotation angles. A more precise load model needs to be developed. Therefore, a load model under two ground layers is studied, according to the literatures [1, 6, 9, 11, 13], which can explain the problem clearly and be easily expanded to more ground layers conditions. Firstly, the load domain threshold is determined by the characterization of strata, and then a precise cutterhead load model is established. After that, numerical simulations are conducted by applying the load model developed above.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 1176–1182, 2009.

2 External Load Domain of Cutterhead

The load model is shown in Fig. 1, in which Cartesian coordinate system is used, the circular section represents the cutterhead, the working environment consists of Layer1 and Layer2, and the variety of the central angle θ represents the change of strata. When the two arc endpoints overlap in the positive y-axis, the angle θ is defined as the starting position (0°). The composition of strata changes with the θ, the cutterhead rotates anticlockwise with a certain speed, and earth pressure of soil cabin and water pressure are ignored. In the process of operating, the resistance and resistance moment of cutterhead front and lateral faces are associated with the composition of strata; the cutting resistance and resistance moment are determined by types and locations of the cutting tools.

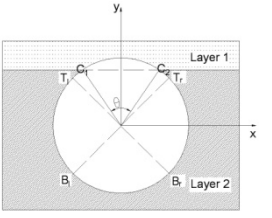


Fig. 1. External load model

2.1 Resistance of Cutterhead

The resistance of the cutterhead F_C [6, 15, 16] can be expressed as

$$F_C = F_1 + F_2 + F_3 \tag{1}$$

Where F_1 is the front resistance of the cutterhead; F_2 is the lateral resistance of the cutterhead; and F_3 is the cutting resistance of cutting tools.

For F_1, according to the literatures [11, 17], there is $\sigma_{nz} = K\sigma_{vz} = K\sum_{i=1}^{n}\gamma_i h_i$, where σ_{nz} is horizontal earth pressure at the depth of z of the cutterhead front face from ground surface; σ_{vz} is vertical earth pressure at the depth of z of the cutterhead front face from ground surface; K is the coefficient of side pressure; n is the number of layers; γ_i is volume weight of soil of layer i; h_i is the thickness of soil of layer i. It is assumed that the layers from ground surface to the top of the cutterhead are all Layer1, and the available area of the cutterhead distributes uniformly. By reference to literatures [11, 13], the micro-unit area dA_t of the cutterhead section in Layer1 is selected, the resistance of which is defined as $dF_{1t} = (1-\alpha_0)\sigma_{nz}dA_t$, where α_0 is the open ratio of the cutterhead. Then an integration is conducted to get F_{1t}.

Similarly, the resistance of the cutterhead section in Layer2, F_{1b}, can be obtained. Then, F_1 can be defined as $F_1 = F_{1t} + F_{1b}$.

For F_2 [6], it can be formulated as

$$F_2 = F_{2t} + 2F_{2s} + F_{2b} \tag{2}$$

Where F_{2t} is the friction between the top quadrant of the cutterhead lateral face (arc $\widehat{T_lT_r}$, see Fig. 1) and soil; F_{2s} is the friction between the side quadrant of the cutterhead lateral face (arc $\widehat{B_lT_l}$ or $\widehat{T_rB_r}$, see Fig. 1) and soil; and F_{2b} is the friction between the bottom quadrant of the cutterhead lateral face (arc $\widehat{B_rB_l}$, see Fig. 1) and soil.

For F_3 [8, 15, 16, 18, 19], it can be calculated as

$$F_3 = \sum_{i=1}^{n_h} F_{Vi} + \sum_{j=1}^{n_c} F_{Nj} \tag{3}$$

Where n_h, n_c are the numbers of disc cutters and cutters respectively.

2.2 Resistance Moment of Cutterhead

The resistance moment of the cutterhead T_C [6, 15, 16] can be expressed as

$$T_C = T_1 + T_2 + T_3 \tag{4}$$

Where T_1 is the front resistance moment of the cutterhead; T_2 is the lateral resistance moment of the cutterhead; and T_3 is the cutting resistance moment of the cutting tools.

For T_1, the hypothesis is the same as F_1. Referring to literatures [9, 11], the micro-unit area dA_t of the cutterhead section in Layer1 is selected, whose resistance moment is defined as $dT_{1t} = (1-\alpha_0)\mu_{ms1}\sigma_{nz} r dA_t$, where μ_{ms1} is the friction coefficient between soil of Layer1 and the cutterhead. Then T_{1t} can be obtained by carrying out an integration.

Similarly, the resistance moment of the cutterhead section in Layer2, T_{1b}, can be obtained. Hence, T_1 can be represented by $T_1 = T_{1t} + T_{1b}$.

For T_2, it is assumed that the layers from ground surface to the top of the cutterhead are all Layer1. According to literatures [9, 11, 20], the micro-unit area dA_t of the cutterhead section in Layer1 is selected, the resistance moment of which can be expressed as $dT_{2t} = \frac{D_C^2}{4}\mu_{ms1}\sigma_\alpha t_C d\alpha$, where σ_α is the earth pressure at the angle of α; t_C is the axial width of the cutterhead. Then T_{2t} can be gotten by proceeding an integration.

Similarly, the resistance moment of the cutterhead lateral face in Layer2, T_{2b}, can be obtained. Thus, $T_2 = T_{2t} + T_{2b}$.

T_3 [8, 10, 15, 16, 19] can be defined as

$$T_3 = \sum_{i=1}^{n_h} F_{Ri} \cdot r_i + \sum_{j=1}^{n_c} F_{Tj} \cdot r_j \tag{5}$$

Where r_i is the distance between disc cutter i and the center of the cutterhead; and r_j is the distance between cutter j and the center of the cutterhead.

3 Establishment of Simulation Model

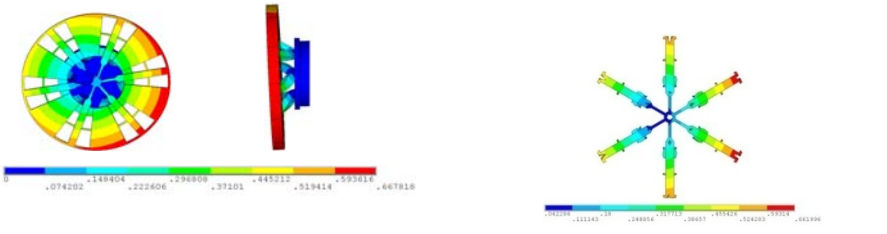
To study the load model further, the theoretical derived load model is used to analyze the panel-style cutterhead by the finite element method. As the cutter open ratio is an important parameter [12, 21], two cutterheads with different open ratios (36% and 25%) are selected according to open ratio ranges [8, 22].

The resistance and resistance moment are applied to the cutterheads, the end of the flanges are fixed, the material of cutterheads is steel, and the impact of disc cutters is ignored. Layer1 is clay, Layer2 is sandy gravel, and physical parameters of strata are cited from literatures [14] and [15].

4 Analysis of Simulation Results

4.1 Deformation of Cutterhead

The deformation of the cutterhead with an open ratio of 36% under θ=132° is shown in Fig. 2(a). It can be seen that the largest deformation, 0.67 mm, occurs at the right side of the cutterhead edge, as a result of the non-uniform load distribution, and the deformation increases radially. Because of the dead weight and the larger force at the bottom of the cutterhead, the whole cutterhead has a downward overturning trend. The resistance moment makes the connection parts of the flange have a torsion deformation. The deformation of cutterhead arms is shown in Fig. 2(b). It can be seen that the largest deformation occurs at the right side of cutterhead arms and the roots of left arms have smaller deformations.



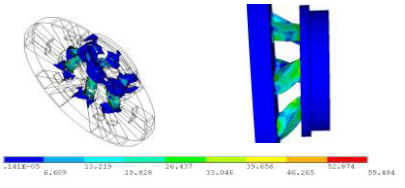
(a) Cutterhead deformation (b) Cutterhead arms deformation

Fig. 2. Distribution of deformation

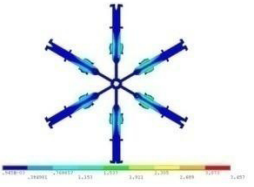
4.2 Intensity of Cutterhead

The equivalence stress distribution of cutterhead with an open ratio of 36% under θ=132° is shown in Fig. 3(a). It can be seen that the stress concentration mainly occurs in flange connections. The maximum equivalent stress, 59.484 MPa, is in the

flange connecting roots which are danger regions of strength failure. From Fig. 3(b), it can be seen that the maximum equivalence stress of cutterhead arms occurs in the inner plate, which effectively reduces the force acting on the roots of the arms. Choosing a suitable support plate can enhance the strength and stiffness of the cutterhead arms.



(a) Cutterhead equivalence stress

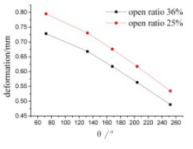


(b) Cutterhead arms equivalence stress

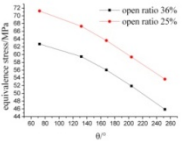
Fig. 3. Distribution of equivalence stress

4.3 Impact of Composite Strata and Open Ratios on Cutterhead Performance

Fig. 4 depicts the combined displacement and equivalence stress with the change of strata for two cutterheads of different open ratios. Because the responses of cutterheads are corresponding to the load changes, the simulations can reflect the load rules. It can be seen that in current case, with the clay increasing, the deformation and equivalent stress of the cutterheads decrease, and for the same formation of strata, the cutterhead with a smaller open ratio suffers a greater deformation and equivalent stress. The reason may be that with the increase of the effective area of the cutterhead, the resistance and resistance moment increase greatly. The load rules reflected above are consistent with the relevant opinions of literatures [12, 21].



(a) Distribution of diplacement



(b) Distribution of equivalence stress

Fig. 4. Cutterhead strength and stiffness vs. composite strata

5 Conclusions

1. A more comprehensive load model of the cutterhead has been established by micro-unit method, based on the strata formation and cutterhead rotation angles. The established load model has better precision and wider applicability than traditional models.
2. Numerical simulations of the cutterheads has been carried out by using the theoretical derived load model. The results show that with the increase of clay, the

deformation and equivalence stress both have downward trends, and the largest deformation of cutterheads occurs at the right side of the cutterhead edges; the greatest equivalence stress occurs in the flange connections which are the danger regions for the strength failure. The cutterhead strength and stiffness are impacted by the load distribution.

3. An analysis of the effect of open ratios on the cutterhead performance has been conducted under the composite strata with the load model. It is found that the larger deformation and equivalence stress happen for a smaller open ratio cutterhead under the same strata composition. This conclusion provides a theoretical basis for the optimization of cutterhead design under composite strata.

Acknowledgments. This research was supported by National Basic Research Program of China (973 Program) No. 2007CB714007.

References

1. Yin, L.C., Zhu, Z.H., Li, Y.Z., et al.: Japanese New Technology in Tunnel Shielding (in Chinese). Huazhong University of Science and Technology Press, Wuhan (1999)
2. Zhao, J., Gong, Q.M., Eisensten, Z.: Tunnelling through a frequently changing and mixed ground: A case history in Singapore. Tunnelling and Underground Space Technology 22(4), 388–400 (2007)
3. Zhu, W.B., Ju, S.J., Shi, H.O.: Guangzhou Metro Line 3 Shield Tunnelling Construction Technology (in Chinese). Jinan University Press, Guangzhou (2007)
4. Zeng, H.B.: Study on Application Technology of Earth Pressure Balance Shield Tunnelling in Complex Strata (in Chinese). Ms D Thesis. Hohai University, Nanjing (2006)
5. Xu, S.H., Zhu, W.B., Yuan, M.Z.: Exploration and Breakthrough of Mix Ground Shield Technology in Guangzhou Metro (in Chinese). In: Proceedings of 2005 Shanghai International Forum on Tunnelling, pp. 189–197. Tongji University Press, Shanghai (2005)
6. Zhang, F.X., Zhu, H.H., Fu, D.M.: Shield Tunnelling Method (in Chinese). China Communications Press, Beijing (2004)
7. Reilly, B.J.: EPBMs for the North East line project. Tunnelling and Underground Space Technology 14(4), 491–508 (1999)
8. Hu, X.P.: Study on Abrasion of Cutters during EPB Shield Tunneling in Sandy Cobble Stratum (in Chinese). Ms D Thesis. Beijing Jiaotong University, Beijing (2006)
9. Lu, Q., Wang, H.L., Fu, D.M.: Research on Torque of Cutterhead for Earth Pressure Balance Shield Machine (in Chinese). In: Proceedings of 2005 Shanghai International Forum on Tunnelling, pp. 452–460. Tongji University Press, Shanghai (2005)
10. Lu, Q., Fu, D.M.: Research on Torque of Cutterhead for Earth Pressure Balance Shield with Simulating Experimental (in Chinese). Chinese Journal of Rock Mechanics and Engineering 25(S1), 3137–3143 (2006)
11. Guan, H.S.: Study on the Calculating Models of Key Parameters & Mechanical Behavior of the EPB Shield Machine (in Chinese). PhD Thesis. Southwest Jiaotong University, Chengdu (2007)
12. Zhu, H.H., Xu, Q.W., Liao, S.M., et al.: Experimental Study on Thrust Force during Tunnel Excavation with Earth Pressure Balance Shield Machine (in Chinese). Rock and soil Mechanics 28(8), 587–1594 (2007)

13. Zhan, S.X., Kang, C.G., Cao, L.J., et al.: Computation of Earth Pressure on Working Face and Optimization of Cylinder Thrust Force for EPB Shield Machine (in Chinese). Science Technology and Engineering 8(23), 6278–6282 (2008)
14. Song, K.Z.: Calculation on Thrust and Cutter Disc Torque of Shield in Sandy Cobble with No Water (in Chinese). Construction Machinery 10, 58–60, 63 (2004)
15. Cui, G.H., Wang, G.Q., Wang, J.X., et al.: Cutting Dynamic Model for Cutters of Shield Machine under Soft Soil (in Chinese). Journal of Jilin University (Engineering and Technology Edition) 38(S2), 139–143 (2008)
16. Song, K.Z., Pan, A.G.: Operation Principle Analysis of Cutting Tools on Shield (in Chinese). Construction Machinery 3, 74–76 (2007)
17. Chen, X.Z.: Soil Mechanics and Geotechnical Engineering (in Chinese). Tsinghua University Press, Beijing (2004)
18. Song, K.Z., Li, C.M., Yuan, D.J., et al.: Study on Distribution Regularities of Disc Cutter Thrust Force for Shield Tunneling (in Chinese). Chinese Journal of Rock Mechanics and Engineering 27(S2), 3875–3881 (2008)
19. Zhang, Z.H., Ye, D.H., Mao, C.J.: Full Face Rock Tunneling Machine Tools and Force Analysis– the Second Technical lecture Full Face Rock Tunneling Machine (in Chinese). Construction Machinery 10, 28–32, 46 (1998)
20. Sugimoto, M., Sramoon, A.: Theoretical Model of Shield Behavior During Excavation. I: Theory. Journal of Geotechnical and Geoenvironmental Engineering 128(2), 138–155 (2002)
21. Yang, H.J., Fu, D.M., Ge, X.R.: Experimental Study and Numerical Simulation of Earth Pressure around Shield Machine (in Chinese). Chinese Journal of Rock Mechanics and Engineering 25(8), 1652–1657 (2006)
22. Wang, H.X., Fu, D.M., Li, X.H.: Statistical Analysis of EPB Shield Main Parameters (in Chinese). In: Proceedings of 2005 Shanghai International Forum on Tunnelling, pp. 248–255. Tongji University Press, Shanghai (2005)

Obstacle-Climbing Capability Analysis of Six-Wheeled Rocker-Bogie Lunar Rover on Loose Soil

Xiaoliu Yu[1,3], Yongming Wang[1,2], Meiling Wang[1], and Lifang Wang[1]

[1] Anhui University of Technology, Ma'anshan, Anhui Province, China
[2] Southeast University, Nanjing, Jiangsu Province, China
[3] State Key Laboratory of Robotics, Shenyang, Liaoning Province, China
wanxiyu@ahut.edu.cn

Abstract. Taking six-wheeled rocker-bogie lunar rover as an object, on the basis of force analysis between the wheels and lunar soil, its obstacle-climbing force model on loose soil was established in this paper, and the wheel sinkages were obtained. Based on the method of solving the wheel's driving torque solution space feasible regions, this paper analyzed the forward obstacle-climbing capability of six-wheeled rocker-bogie lunar rover on loose soil, including single-wheel obstacle-climbing and two wheels obstacle-climbing simultaneously. Simulations show that under the loose soil environment, the wheels have different obstacle-climbing capability, i.e. the rear wheel is the best, the middle one is the worst and the front one is moderate.

Keywords: Obstacle-climbing Capability Analysis, Lunar Rover, Rocker-bogie, Force Analysis, Solution Space Feasible Region.

1 Introduction

Lunar surface is rather complex and rough, which is almost completely covered by a thick loose bed (i.e. lunar soil). Therefore, to complete the exploration mission, it is essential for lunar rover to have a good obstacle-climbing capability on loose soil.

Taking the six-wheeled rocker-bogie lunar rover as an object [1-2], and based on establishing the force model between the wheels and lunar soil, its obstacle-climbing capability on loose soil was analyzed in this paper.

2 Force Model of Lunar Rover on Loose Soil

2.1 Force Analysis of Rigid Wheel on Loose Soil

When the six-wheeled rocker-bogie lunar rover travels on loose soil, the rigid wheels will have a small sinkage. As shown in Fig.1, a vertical load W and a driving force P are applied to the wheel; a torque T is applied at the wheel rotation axis by the motor, h is the wheel sinkage, as in [3]. θ_1 is the approach angle, which is the corresponding central angle of the wheel from the vertical position to the first touchdown position; θ_2 is the departure angle, which is the corresponding central angle of the wheel from the vertical

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 1183–1189, 2009.

position to the lift-off position. For the rigid wheel, θ_2 is usually very small, and can be regarded as zero. Their interaction region between the wheel and loose soil can be divided into two zones: [θ_m, θ_1] and [θ_2, θ_m], where θ_m is the corresponding central angle of the wheel from the vertical position to the maximum normal stress position, as shown in Fig.1. And τ_1, τ_2 are the shear stresses of the two zones, σ_1, σ_2 are the normal stress of the two zones.

Based on Karl Iagnemma's study [4], there are two assumptions. The first assumption is that the angle θ_m is the average of θ_1 and θ_m, as written in (1).

$$\theta_m = \frac{\theta_1 + \theta_2}{2} \quad . \tag{1}$$

The second assumption is that the maximum shear stress and the maximum normal stress occur at the same position corresponding to θ_m, which has been confirmed by simulation. Based on the above-mentioned assumptions, the maximum normal stress σ_m can be written as in (2) and the shear stress is described as in (3).

$$\sigma_m(\theta_m) = \left(k_c + k_\phi b\right)\left(\frac{r}{b}\right)^n \left(\cos\theta_m - \cos\theta_1\right)^n \quad . \tag{2}$$

$$\tau_m(\theta_m) = \left(c + \sigma_m \tan\phi\right)\left(1 - e^{-\frac{r}{k}(\theta_1 - \theta_m - (1-\lambda)(\sin\theta_1 - \sin\theta_m))}\right) \quad . \tag{3}$$

Where, k_c is the cohesive modulus of lunar soil deformation, k_φ is the frictional modulus of lunar soil, n is the sinkage coefficient, b is the wheel width, and r is the wheel radius. c is the cohesion of lunar soil, k is the shear deformation modulus of lunar soil, ϕ is the internal friction angle of lunar soil, and λ is the slip ratio of wheel.

According to the force equilibrium relationship, we can establish the dynamics equations of single wheel, and written in (4) and (5). And as in [5-6], the wheel sinkage h can be calculated in (7).

$$\begin{aligned} W &= rb\left(\int_{\theta_1}^{\theta_2} \sigma(\theta)\cos\theta \cdot d\theta + \int_{\theta_1}^{\theta_2} \tau(\theta)\sin\theta \cdot d\theta\right) \\ &= \frac{rb}{\theta_m(\theta_1 - \theta_m)}\left(\sigma_m(\theta_1\cos\theta_m - \theta_m\cos\theta_1 - \theta_1 + \theta_m) + \tau_m(\theta_1\sin\theta_m - \theta_m\sin\theta_1)\right) \end{aligned} \quad . \tag{4}$$

$$\begin{aligned} P &= rb\left(\int_{\theta_1}^{\theta_2} \sigma(\theta)\sin\theta \cdot d\theta - \int_{\theta_1}^{\theta_2} \tau(\theta)\cos\theta \cdot d\theta\right) \\ &= \frac{rb}{\theta_m(\theta_1 - \theta_m)}\left[\sigma_m(\theta_m\sin\theta_1 - \theta_1\sin\theta_m) + \tau_m(\theta_1\cos\theta_m - \theta_m\cos\theta_1 - \theta_1 + \theta_m)\right] \end{aligned} \quad . \tag{5}$$

$$T = \frac{1}{2} r^2 b \tau_m \theta_1 \quad . \tag{6}$$

$$h = r(1 - \cos\theta_1) \quad . \tag{7}$$

Given that $\theta_m = \theta_1/2$, and the vertical load W, the driving force P are known, according to (2)-(6), we can calculate the variables θ_1 ,σ_m, τ_m, λ, T.

2.2 Force Analysis of Rocker-Bogie Mechanism on Loose Soil

When the lunar rover travels on loose soil, the rigid wheels will have a sinkage h, therefore the kinematics of rocker-bogie mechanism will also be changed. Without regard to the normal friction forces, the force analysis is based on the rocker-bogie planes. The two side rocker-bogie mechanisms are selfsame except the displacements and forces at the joints, so their mechanical models are selfsame too. This paper only analyzes the right side rocker-bogie mechanism.

For convenience, it is assumed that they are point-contact between the wheels and lunar soil. The angles α_1, α_2 and α_3 represent the angle between the horizontal plane and the wheel–terrain contact plane. r is the wheel radius, h_1, h_2 and h_3 are sinkages of the rear, middle and front wheels. There is a free pivoting joint at point B, and there is a differential joint at point A, as in [7]. F_x, F_z and M_y are the forces and torque at point A. W_1, W_2, W_3 are the normal forces of the wheels. P_1, P_2, P_3 are the driving forces of the wheels, as shown in Fig.2.

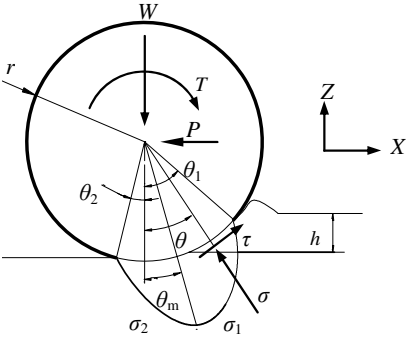


Fig. 1. Rigid wheel on loose soil

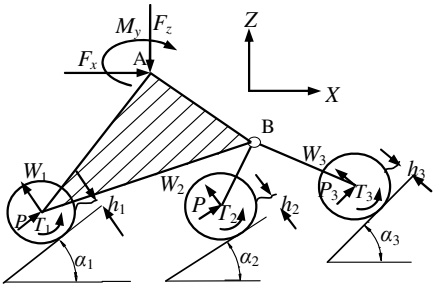


Fig. 2. Force analysis of rocker-bogie mechanism

According to X and Z direction force equilibrium, the following force equations can be established. According to Y direction torque balance of the rocker-bogie mechanism, (10) can be obtained.

$$F_x - W_1 \sin\alpha_1 - W_2 \sin\alpha_2 - W_3 \sin\alpha_3 + P_1 \cos\alpha_1 + P_2 \cos\alpha_2 + P_3 \cos\alpha_3 = 0 \quad . \tag{8}$$

$$-F_z + W_1 \cos\alpha_1 + W_2 \cos\alpha_2 + W_3 \cos\alpha_3 + P_1 \sin\alpha_1 + P_2 \sin\alpha_2 + P_3 \sin\alpha_3 = 0 \quad . \tag{9}$$

$$\begin{aligned} &M_y + (x_A - x_a)(W_1\cos\alpha_1 + P_1\sin\alpha_1) + (z_A - z_a)(W_1\sin\alpha_1 - P_1\cos\alpha_1) - T_1 + \\ &(x_A - x_m)(W_2\cos\alpha_2 + P_2\sin\alpha_2) + (z_A - z_m)(W_2\sin\alpha_2 - P_2\cos\alpha_2) - T_2 + \\ &(x_A - x_f)(W_3\cos\alpha_3 + P_3\sin\alpha_3) + (z_A - z_f)(W_3\sin\alpha_3 - P_3\cos\alpha_3) - T_3 = 0 \end{aligned} \quad . \tag{10}$$

Where, x_A, z_A are X and Z coordinates of the differential joint at point A, $x_f, z_f, x_m, z_m, x_a, z_a$ are X and Z centre coordinates of the front, middle and rear wheels.

The torque conservation equation at point B can be written as in (11).

$$\begin{aligned}&(x_B - x_m)(W_2\cos\alpha_2 + P_2\sin\alpha_2) + (z_B - z_m)(W_2\sin\alpha_2 - P_2\cos\alpha_2) - T_2 + \\ &(x_B - x_f)(W_3\cos\alpha_3 + P_3\sin\alpha_3) + (z_B - z_f)(W_3\sin\alpha_3 - P_3\cos\alpha_3) - T_3 = 0\end{aligned} \quad (11)$$

Where, x_B z_B are the X and Z coordinates of the free pivot at point B.

According to (8)-(11), there are six unknowns implied in the four equilibrium equations, so there are two variable parameters of the rocker-bogie mechanism. It is noted that we can choose the reasonable driving torques to control the whole rocker-bogie mechanism for obtaining a better obstacle-climbing capability.

3 Analysis of the Driving Force and Wheel Sinkages

3.1 Analysis of the Driving Force

Given the vertical load W, the wheel will have the maximum driving force P_{max} when the slip ratio λ=1, and at that time the actual wheel center velocity is zero, namely the wheel will get stuck in loose soil, the motor will do useless work. In practice, the slip ratio λ< 1, and λ_{max} is needed to be given. The wheel will have the minimum driving force when the driving torque T is dropped to zero, denoted as P_{min}. Under the loose soil environment, When W and P are given and λ_{max} =0.9, the driving torque T is determined according to (1)-(7), and the relationship curve of T, P, and W is shown in Fig.3.

3.2 Analysis of the Wheel Sinkages

When the normal forces W and the driving force P are given, the wheel sinkage h is determined according to (1)-(7), and their relationship curve is shown in Fig.4.

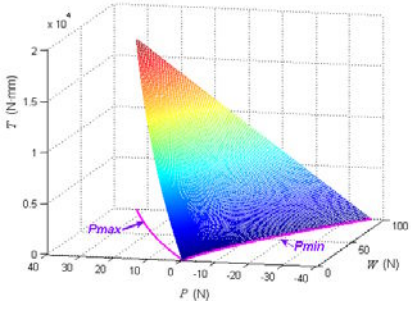


Fig. 3. Relationship curve of T, P, and W

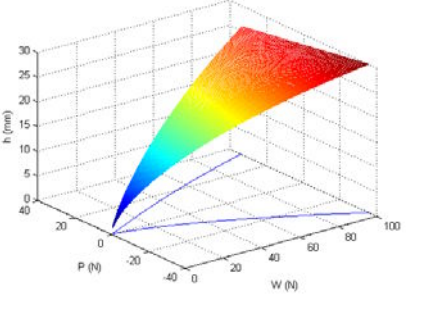


Fig. 4. Relationship curve of h, P, and W

4 Obstacle-Climbing Capability Analysis

Whether the lunar rover travels forward or backward, their obstacle-climbing capabilities are almost consistent. This paper only analyzes the forward obstacle-climbing capability. The wheels and lunar soil characteristic parameters are given in Table 1, as in [8].

Table 1. Simulation parameters

Wheel radius r(mm)	Wheel width b(mm)	Frictional modulus k_φ(N·m$^{-(n+2)}$)	Cohesive modulus k_c(N·m$^{(n+1)}$)	Shear deformation modulus k (m)	Cohesion c (kPa)	Internal friction angle ϕ(°)	Sinkage coefficient n
130	130	814370	1379	0.01778	0.172	40	1.0

4.1 Analysis of Single-Wheel Obstacle-Climbing

Taking the right wheel 3 for example, given that the obstacle is a 40° slope, the height of obstacle is 100-200mm, i.e. α= [0 0 40° 0 0 0], Z= [0 0 100-200 0 0 0]. When giving the obstacle heights of 100mm, 150mm and 200mm, the solution space feasible regions of the driving torque T_1, T_2 and T_3 are shown in Fig.5. Simulation shows that as the obstacle height increasing, the right solution space feasible regions have the gradually decreasing trend while the left solution space feasible regions changed little.

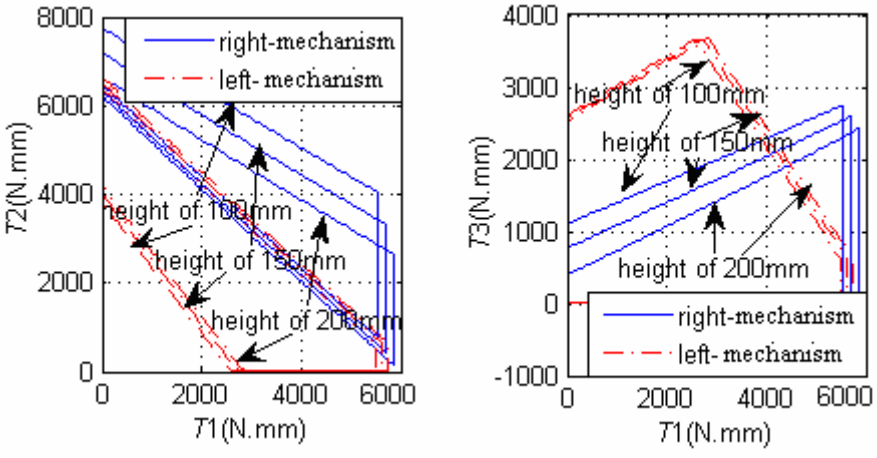


Fig. 5. The solution space feasible regions of the right wheel 3

4.2 Analysis of Two Wheels Obstacle-Climbing Simultaneously

Wheel 3 and wheel 6 obstacle-climbing simultaneously. Given that the obstacle is a 40° slope, the height of obstacle is 100-200mm, i.e. α= [0 0 40° 0 0 40°], Z= [0 0 100-200 0 0 00-200]. When giving the obstacle heights of 100mm, 150mm and 200mm, the solution space feasible regions of driving torque T_1, T_2 and T_3 are shown in Fig.6.

Wheel 2 and wheel 5 obstacle-climbing simultaneously. Given that the obstacle is a 40° slope, the height of obstacle is 0mm, i.e. α= [0 40° 0 0 40° 0], Z= [0 0 0 0 0 0], there is no solution by simulation. Fig.7 shows the solution space feasible regions of wheel 2 and wheel 5 when giving α= [0 30° 0 0 30° 0], Z = [0 0-100 0 0 0-100 0].

The Fig.6 and Fig.7 shows that with the increase of obstacle height, the size of the feasible region has no obvious change except its shape and position.

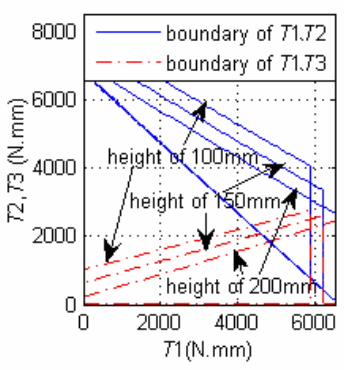


Fig. 6. The solution space feasible regions of wheel 3 and wheel 6

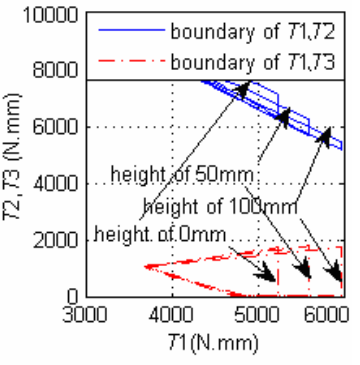


Fig. 7. The solution space feasible regions of wheel 2 and wheel 5

5 Conclusions

On the basis of force analysis between the wheels and lunar soil, this paper established the obstacle-climbing force model of six-wheeled rocker-bogie lunar rover on loose soil. From the above analysis, we may draw the conclusion that when the lunar rover goes forward, with the same obstacle heights and slope angles, the obstacle-climbing capability of the wheels are different, namely the rear wheel is the best, the middle one is the worst, and the front one is moderate.

Acknowledgments. This work is supported by the key innovation project granted by the Ministry of Education in China (No.708054), the fund of the Chinese Academy of Sciences State Key Laboratory of Robotics (No.RLO200816) and the key project of Anhui Education Department in China (No.KJ2009A015Z).

References

1. Kubota, T., Kuroda, Y., Kunii, N., et al.: Small, Light-weight rover Micro5 for lunar exploration. Acta Astronautica 52(2-6), 447–453 (2003)
2. Hayati, S., et al.: The Rocky 7 Rover: A Mars Sciencecraft Prototype. In: Proceedings of the 1997 IEEE International Conference on Robotics and Automation, 1997, pp. 58–64 (1997)
3. Shibly, H., Iagnemma, K., Dubowsky, S.: An equivalent soil mechanics formulation for rigid wheels in deformable terrain with application to planetary exploration rovers. Journal of Terramechanics 42, 1–13 (2005)
4. Iagnemma, K., Shibly, H., Dubowsky, S.: On-line terrain parameter estimation for planetary rovers. In: Proceedings of IEEE international conference on robotics and automation, pp. 3–4 (2002)

5. Yoshida, K., Hamano, H.: Motion Dynamics and Control of a Planetary Rover with Slip-Based Traction Model. In: Proc. SPIE, vol. 4715, pp. 275–286 (2002)
6. Yoshida, K., Hamano, H.: Motion Dynamics of a Rover with Slip-Based Traction Model. In: IEEE International Conference on Robotics and Automation, Washington, DC, pp. 3155–3160 (2002)
7. Wang, Y., Yu, X., Tang, W.: Analysis of Obstacle-climbing Capability of Planetary Exploration Rover with Rocker-bogie Structure. In: 2009 International Conference on Information Technology and Computer Science, pp. 329–332. Kiev, Ukraine (2009)
8. National Aeronautics and Space Administration. Nasa Space Vehicle Design Criteria (Environment): Lunar Surface Models. Nasa, pp. 27–30 (1969)

Optimal Design on Cutterhead Supporting Structure of Tunnel Boring Machine Based on Continuum Topology Optimization Method

Zhen Li[1], Wei Sun[1], Junzhou Huo[1], Pengcheng Su[2], and Liying Deng[2]

[1] School of Mechanical & Engineering, Dalian University of Technology, Dalian 116024, P.R. China
lizhen_dlut@163.com, sunwei@dlut.edu.cn, huojunzhou@gmail.com
[2] Northern Heavy Shenyang Heavy Machinery Group Co., Ltd., Shenyang 110025, P.R. China
supc2005@126.com, shdly0123@163.com

Abstract. A continuum topology optimization method is introduced to the conceptual design of a cutterhead supporting structure of tunnel boring machine to avoid numerous empirical solutions. The conceptual design of the supporting structure is determined in the guidance of topology optimization combined with given working conditions. A finite element analysis indicates that the deformation and stress of the optimized cutterhead are reduced greatly as compared with those of the original design. Furthermore, the deformation distribution of the optimization is more homogeneous than that of the original one. The present work indicates that the continuum topology optimization plays an effective role in guiding the supporting structure design of cutterhead.

Keywords: Supporting structure of cutterhead, Topology optimization, Conceptual design, Finite element analysis.

1 Introduction

Full-face rock tunnel boring machine (TBM) has been applied widely in underground tunneling projects and municipal establishment constructions due to its high tunneling speed, high degree of mechanization and onetime tunnel formation and etc. The cutterhead equipped with disc cutters is a key component of TBM, which functions as excavating rocks, stabilizing the working face, collecting the excavated mucks and so on. The TBM is generally encountered various stratums from the mud, clay, sand to soft and hard rock during excavating tunnels, which makes the cutterhead bear complex loadings. Therefore, the optimal design of the cutterhead plays an important role in improving the rock-breaking capacity, tunneling efficiency and properties[1,2].

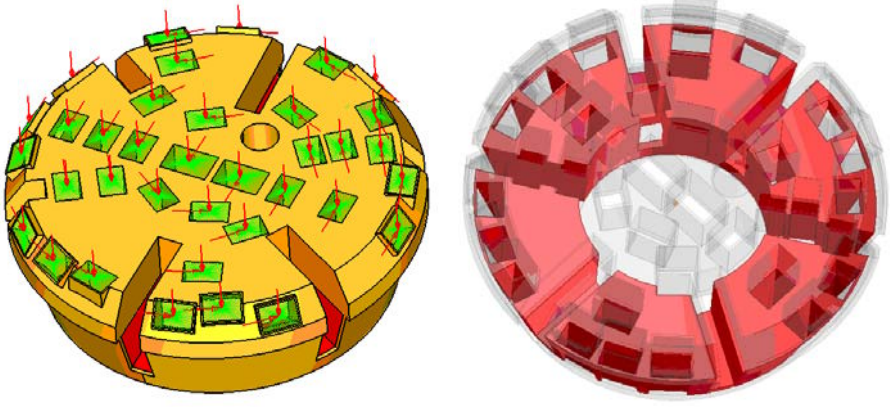
As the cutterhead structure is confined greatly by the geological conditions, the cutterhead structure designs include supporting structure design, structure topology parameters design and muck transfer system design, etc.[3,4] Among

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 1190–1195, 2009.

them, the supporting structure design determines the whole distributions of deformation and stress of the cutterhead. However, the designs on supporting structure have been mostly based on trial-and-error methods up to now. It is very difficult to find an optimal supporting structure that makes the deformation and stress of cutterhead distribute uniformly during the tunnel excavation. Therefore, in this paper, the continuum topology optimization method is introduced to research the optimal design of supporting structure, and its final configuration is determined by the conceptual design combined with given working conditions. The deformation and stress distributions of both optimized and original cutterhead are analyzed by using the finite element method.

2 The TBM Cutterhead Modeling

In this paper, the cutterhead model is designed reference to that used in the Datie projects. In order to simulate the stress state of cutterhead really, this model includes all of the necessary parts, i.e. 8 center cutters, 21 face cutters, 13 gauge cutters, 6 mucking holes and 1 person hole, as shown in Fig.1 (a). It is noted that the disc cutters are excluded in this model to simplify the finite element analysis. Therefore, the normal load of 500 kN and circumferential friction force of 60 kN acting on disc cutters are now applied to the cutter saddle directly. In addition, the disc cutters need to be replaced due to abrasion in TBM tunneling process, it does not set material in corresponding disc cutters place of the topology optimization domain. The mucking channels for mucking holes are also reserved in the topology optimization domain (Fig.1 (b)).



(a) Cutterhead model

(b) Topology optimization domain of the supporting structure

Fig. 1. Cutterhead model

3 Continuum Topology Optimization Method

The objective of structural optimization is to find the "best possible" or "optimal" structure which meets all the multidisciplinary requirements such as those imposed by functionality and manufacturing conditions. Topology optimization is one of the most challenging research topics in the field of structural optimization. Since continuum topology optimization can give the optimal structure configuration at the conceptual design stage, the topology optimization method has been attracted more and more attentions in many industrial fields[5].

A variable density method is used in optimize the topology structure of the supporting structure. The relative density of material at an arbitrary point in the design domain is taken as the design variable. Thus the material properties of every point in the design domain are determined by the relative densities of materials. The method is implemented based on the solid isotropic material with penalization (SIMP) model[6] where relative densities are interpolated at each finite element, providing a continuous material distribution in the domain. At each point of the design domain, the local effective elasticity modulus of the material $E^p(\rho)$ is defined as:

$$E^p(\rho) = \rho^p E \tag{1}$$

where E is the elasticity modulus of basic material, ρ is a pseudo-density describing the amount of material at each point of the design domain and the value between 0 and 1, and $p \in [1, 4]$ is a penalization factor to recover the discrete design. $\rho = 0$ indicates the material void and $\rho = 1$ the material solid.

For the stiffness design of structures, it usually adopts the structural compliance as the objective function of optimization problem. Therefore the mathematic model of topology optimization can be formulated as follows:

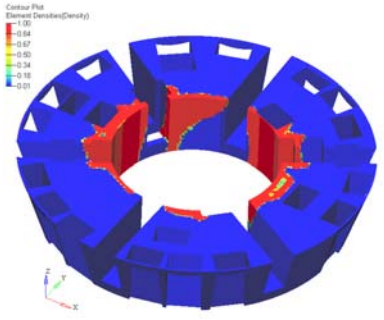
$$\begin{aligned} \min \quad & f(\boldsymbol{\rho}) = U^T K U \\ \text{subject to:} \quad & K(\boldsymbol{\rho}) U = F \\ & V(\boldsymbol{\rho}) = \sum_{e=1}^{M} \int_{\Omega_e} \rho_e \, d\Omega \leq V^* \\ & 0 < \boldsymbol{\rho}_{min} \leq \boldsymbol{\rho} \leq 1 \end{aligned} \tag{2}$$

where $\boldsymbol{\rho}$ is the design vector that varies from 0 (void) to 1 (solid) continually, ρ_e the material density of element e, $f(\boldsymbol{\rho})$ the objective function, $V(\boldsymbol{\rho})$ the sum of element volume, V^* the constraint on material volume, K the global stiffness matrix, U the displacement vector, F the force vector. Since all practical problems are solved numerically, the density variables must be given a lower limit value $\boldsymbol{\rho}_{min}$ to ensure the stiffness matrix of the finite element problem does not become singular. The problems of topology optimization must be solved with theoretically well-founded mathematical programming methods, such as sequential quadratic programming (SQP) and the method of moving asymptotes (MMA)[7]. Here, MMA is adopted to solve the topology optimization problem of compliant mechanism.

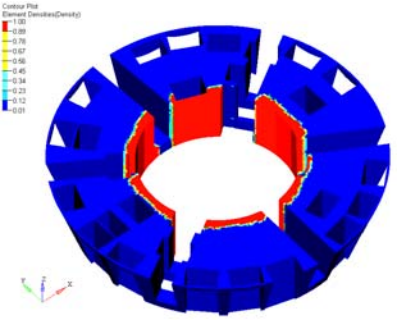
4 Optimal Design on Supporting Structure

4.1 Topology Optimization of Supporting Structure

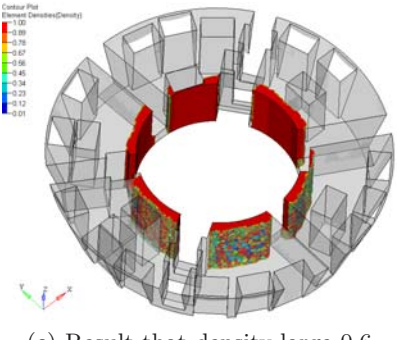
During the tunneling process, the deformation of the cutterhead should be small as possible to ensure abrasion consistency of disc cutters and reduce stress



(a) Original result without symmetrical constraint

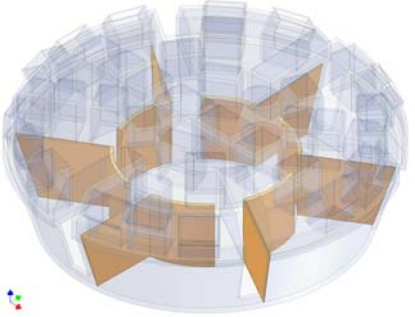


(b) Result applied symmetrical constraint

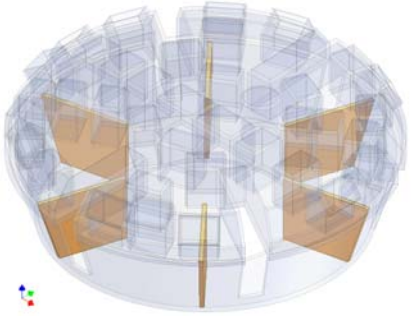


(c) Result that density large 0.6

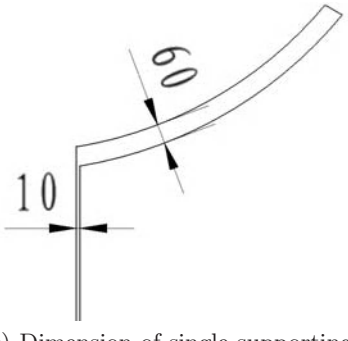
Fig. 2. The result of topology optimization



(a) Conceptual design of supporting structure



(b) Original design of supporting structure



(c) Dimension of single supporting rib

Fig. 3. The result of topology optimization

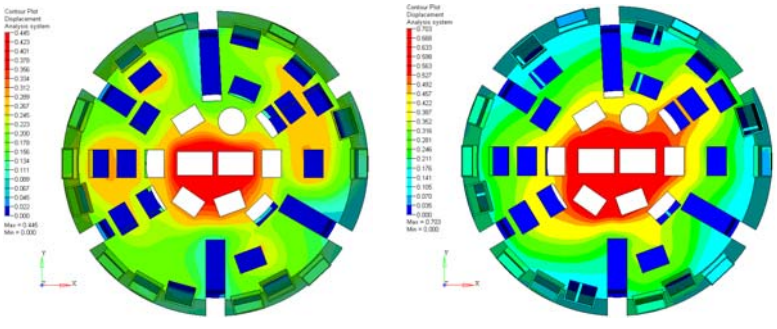
concentration of cutterhead. Therefore, the objective function for the topology optimization of supporting structure is chosen as the entire cutterhead compliance. The volume constraint is set as less than 10 % of the design domain. For the solid material, the elasticity modulus and Poisson's ratio are 200 GPa and 0.3 respectively.

No constraints are first applied to the structure due to the complicated design domain. The optimized result is shown in Fig.2 (a). It can be seen that material distributes inhomogenously in the region near the inner circle of the design domain. Such structure has poor manufacturability. Then, a compulsory symmetrical constraint is applied to the design domain according to the distributed characteristics of mucking holes. The results (Fig.2 (b)) show that 6 substructures distribute uniformly in the region near the inner circle of the designs domain after applying constraints, and this kind of structure has well manufacturability, superior to that without constraints. Figure 2 (c) is the result that removes the domain ($\boldsymbol{\rho} < 0.6$) to show entity supporting structure.

4.2 Conceptual Design of Supporting Structure

The final configuration of the supporting structure (Fig.3 (a)) is obtained through the conceptual design based on the result topology optimization (Fig.2 (c)). This supporting structure is composed of 6 supporting ribs with same shape. Figure 3 (c) gives the dimensions of the supporting ribs, in which the ribs with the thickness of 60 mm play the major supporting function, while the ribs with the thickness of 10 mm transfer mucks. Furthermore, the gross weight of the optimized supporting structure is 3220.874 kg, which reduced 7.44 % compared with that of the original design (Fig.3 (b)).

The deformations and stresses of the optimized and original cutterheads are calculated by finite element method. Figure 4 gives respectively the deformations of the optimized and original cutterhead under loadings. The maximum deformation and the maximum equivalent stress of the optimized cutterhead are



(a) Result of conceptual design (b) Result of original design

Fig. 4. Deformation of cutterhead for different supporting structure

0.445 mm and 178.66 MPa respectively, while those of the original one are respectively 0.703 mm and 250.725 MPa. Topology optimization makes the maximum deformation and equivalent stress decrease 37 % and 29 %, respectively. Furthermore, the deformation of optimized cutterhead is more homogeneous than the original one. Therefore, it can reduce the stress concentration of cutterhead and extend the life of the disc cutter using the conceptual design.

5 Conclusion

The optimal configuration of cutterhead supporting structure of tunnel boring machine is obtained through the conceptual design where a continuum topology optimization method is introduced. The finite element analysis indicates that compared with the original design, the maximum deformation and equivalent stress of the optimized cutterhead decrease 37 % and 29 %, respectively. Furthermore, the deformation distribution of the optimization becomes more homogeneous. The present work shows that the continuum topology optimization is an effective method to guide the optimal design of cutterhead supporting structure.

Acknowledgements

This work was supported by the Major State Basic Research Development Program of China (973 Program) (Granted No.2007CB714006), Notional Key Technology R&D Program (Granted No.2007BAF09B01), Liaoning key scientific and technological project (Granted No.2008220017) and China Postdoctoral Science Foundation (Granted No.20080431136).

References

1. Maidl, B., Schmid, L., Ritz, W., Herrenknecht, M.: Hardrock Tunnel Boring Machines. Ernst & Sohn Verlag fr Architektur und technische Wissenschaften GmbH & Co. KG, Berlin (2008)
2. Kolymbas, D.: Tunnelling and Tunnel Mechanics: A Rational Approach to Tunnelling. Springer, Heidelberg (2008)
3. Zhijie, L., Hongfei, T., Yanjun, S., Guanghui, Y.: Cutterhead design key issues of a full face rock tunnel boring machine. China Mechanical Engineering 19, 1980–1985 (2008)
4. Weibin, Z., Shijian, J., Haiou, S.: Guangzhou Metro Line 3 Shield Tunnelling Construction Technology. Jinan University Press (2007)
5. Eschenauer, H., Olhoff, N.: Topology optimization of continuum structures: A review. Applied Mechanics Reviews 54, 331–389 (2001)
6. Bendsøe, M.P., Sigmund, O.: Topology Optimization — Theory, Methods, and Applications. Springer, Heidelberg (2003)
7. Bruyneel, M., Duysinx, P., Fleury, C.: A famility of mma approximations for structural optimization. Structural and Multidisciplinary Optimization 24, 263–276 (2002)

The Flexible Two-Wheeled Self-balancing Robot Based on Hopfield

Xian-gang Ruan and Jianwei Zhao*

Institute of Artificial Intelligence and Robots, Beijing University of Technology, 100124, Beijing, China
zhaojianwei@emails.bjut.edu.cn

Abstract. The linear model of flexible two-wheeled self-balancing robot that has been proved is used in the paper. The lumbar of robot is the spring. The discrete Hopfield neural network is used for controlling robot's posture based on association study. Study of biological control method to realize flexible robot posture control in the adaptive, self-organizing capacity, the reasonable energy function is defined for the non-linear and strong coupling robot, then design the Hopfield controller, used the dynamic function of flexible robot, which satisfied the need of robot's dynamic process. Analyze the convergence of the controller. Simulation experiments show the effectiveness of the method and proved the stability of system. Validity and rationality of the system controller designed are verified through the performance experiments of the prototype and analyzed the experiment result in detail. The robot like the human being not only structure, but also physiological intelligence.

Keywords: Flexible robot, lumbar, Hopfield network.

1 Introduction

Two-wheeled upright robot, also called two-wheeled balancing robot or two-wheeled self-balancing robot, which is the balance controlling system like human being to keep the upright posture and its main is the controlling of movement balance. Around the main problem of motion balance, both at home and abroad have a lot of researches of two-wheeled upright balancing robot control. American Boskovich designs and manufacture a two-wheeled upright robot, its control system exercise two-wheeled robot balance control in 1995 [1]. Kim et al suggest that an improved input-output linearization control method for two-wheeled upright robot position and attitude tracking control problem in 1999 [2]. Urakubo et al, Lyapunov control method introduced to control the position and rotation of two-wheeled robot in order to overcome the problem of dynamics model inaccurate, and the physical experiment has been done after the simulation experiment in 2001 [3]. Amir Karniel et al based on neurophysiology of learning mechanism, construct a simulated lamprey brainstem artificial neural circuits, and composed a closed-loop system to control the exercise of phototaxis in 2002 [4]. Trong et al based on Lyapunov stability theory, and design nonlinear control strategy for the two-wheeled of welding robot to control its tracking the

* Corresponding author.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 1196–1204, 2009.

desired trajectory welding in 2003 [5]. Kim et al analyzed the dynamic balance of two-wheeled upright robot, and explained the characteristics of two-wheeled upright robot balance control and rotary motion control characteristics in 2005 [6]. Kozlowski et al, time differentiable smooth control law is introduced to the two-wheeled upright robot, and realize the robot's balance control and achieved good results in 2006 [7]. There are several classic entities in recent years, professor Anderson of Southern Methodist University American designed the two-wheeled upright robot called nBot, which is characterized by robust balance control of motion performance in the rugged road [8]. The uBot-4 of Massachusetts USA laboratory scientific, while maintaining a balanced posture, but also hold the ball completed, open the drawer, and shovel debris mission on the ground [9]. Segway Inc. they developed the two-wheeled manned car as airport police officers means of transport. Quasimoro [10] and Joe [11] et al machine entities.

20^{th} century 80's, Hopfield had issued a series of important scientific papers [12-14]. The Hopfield neural network, which has some similarities to the human nervous system, has been widely adopted by researchers in Artificial Intelligence: in robotics it has been mainly utilized in path planning (Fan [15], Sadati [16]), and robot navigation (Jun [17] and Lendaris [18]). Jiang [19] has used Hopfield neural networks to create an autonomous robot which can dynamically adapt to avoid obstacles. However, there has been little research published on flexible two-wheeled self- balancing robots.

To sum up, no two-wheeled flexible about robot research reports. This paper uses springs imitating human lumbar spine and designs the real flexible humanoid robots. The linear model of flexible two-wheeled self-balancing robot that has been proved is used in the paper. The robot's balance controller is considered and successfully implemented using the Hopfield network methods. With the study of function, self-organizing and adaptive, the robot can be called truly intelligent robot. The paper provides a theoretical basis for the control research of flexible two-wheeled self-balancing robot, which will be a meaningful and value of scientific research platform.

2 Hopfield Network Optional Control

2.1 The Structure and Working Methods of Discrete Hopfield Network

Hopfield neural network are first proposed by American physicist JJ Hopfield in 1982, is mainly used to simulate the memory mechanism of biological neural network and is a fully connected neural network. Because of its discrete time, Hopfield network is called discrete Hopfield neural network (DHNN). Hopfield neural networks are the evolution of the state of a non-linear dynamic system and can be a group of non-linear differential equations to describe. Stability of the system can be analyzed with "energy function" (Lyapunov function or Hamiltonian). If certain conditions are met, the energy function in the network is running constantly decreasing and finally arrive the stable state. The discrete Hopfield neuron network (DHNN) is implemented on the robot's single board computer.

The DHNN has a single-layer - its input and output values for the second feedback network; in total there are n neurons per node. The time domain is divided into six

steps using 64 neurons to express a state of volume. The robot has six state variables that need be expressed, so 384 neurons node are needed. The output of each node is connected to every other neuron's input, while all other output neurons are connected to the input neurons.

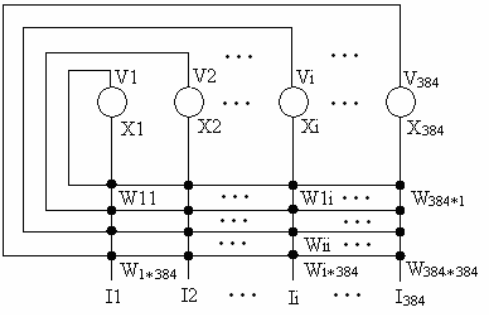


Fig. 1. The structure of the discrete Hopfield network implemented by the robot

The working method of a neuron is

$$\begin{cases} s_i = \sum_{j=1}^{n} \omega_{ij} V_j - I_i \\ V_i = f(s_i) \end{cases} \quad i, j = 1,2,...,n \tag{1}$$

where ω_{ij} represent the weights, I_i is the threshold, $f(\bullet)$ is a transformation function. Energy changes occur asynchronously rather than synchronously so asynchronous updates are used in the entire network. Used randomly selected by means of adjust.

2.2 Energy Function and Stability of Networks

Hopfield network energy function has specific physical meaning. Its discrete form is a Hamiltonian energy, at the same time it is also an extended Lyapunov function and the generalized Lyapunov. State equations for the output:

$$c_i \frac{dx_i}{dt} = -\frac{x_i}{R_i} + \sum_{j=1}^{n} \omega_{ij} V_j + I_i \tag{2}$$

where: $i, j = 1,2,...,n$; $V_i = f(x_i)$. Hopfield energy functions is

$$E = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \omega_{ij} V_i V_j + \sum_{i=1}^{n} \frac{1}{R_i} \int_0^{V1} f^{-1}(\eta) d\eta = -\frac{1}{2} V^T W V - V^T I + \sum_{i=1}^{n} \frac{1}{\tau_i} \int_0^{V1} f^{-1}(\eta) d\eta \tag{3}$$

where: $1/R = 1/r + \sum_j \omega_{ij}$, r is input resistance of neurons operation amplifier, ω_{ij} represent the weights, c_i is the total input capacitance of the i operation

amplifier, x_i is the input state of the i operation amplifier, V_i is its output, I_i is impressed current. For the DHNN, since $f(\bullet)$ is transformed function that it is two-valued function, the integral term is 0, when V_i is 0, while it is positive in other cases. If the τ_i was very great, the third can be ignored. Since $V_i \in (-1,1)$ or $V_i \in (0,1)$, the above definition of energy function E is bounded. Only need to prove $dE/dt \le 0$ that the system is stable.

$$\frac{dE}{dt} = \sum_{i=1}^{n}(-\frac{dx_i}{dt}\frac{dV_i}{dt}) = -\sum_{i=1}^{n}(\frac{dx_i}{dV_i}\frac{dV_i}{dt}\frac{dV_i}{dt}) = -\sum_{i=1}^{n}(\frac{dx_i}{dV_i}(\frac{dV_i}{dt})^2) \tag{4}$$

Earlier assumption that $V_i = f(x_i)$ is the increased function of Monotony, so $x_i = f^{-1}(V_i)$ is also the increased function of Monotony. For $dx_i/dV_i \ge 0$, so $(dx_i/dV_i)^2 \ge 0$. And then $dE/dt \le 0$. Only if all the constants are only admitted to equate 0.According to Lyapunov stability theory, the system must be gradual and stable. In other side, the evolution of the network is always moving in the direction of E reducing exercise, has been achieved to a very small E, then all of V_i is a constant, also the network converges to the steady state.

2.3 Application DHNN for Solving Optimal Control Problem

Definition DHNN energy functions and the linear quadratic performance index:

$$E = -\frac{1}{2}V^T WV + V^T I \tag{5}$$

$$J = X^T(N)HX(N) + \sum_{k=0}^{N-1}(X(k)Q(k)X(k) + U(k)R(k)U(k)) \tag{6}$$

If the linear quadratic performance index of would been transformed the form of DHNN energy function, the optimal solution of linear quadratic problem could be obtained through the neural network. Since the outputs of the DHNN are in the form of a state (0 or 1). A real number can be expressed by using a group of neurons in a state of volume. For Hopfield neural network, if using M neurons expressed a control value u, exist matrices $D \in r*N*L$.

Where: $F \in rN*rN$, $G \in rN*1$, $\tilde{R} = diag[R(0), R(1), \cdots, R(N-1)] \in R^{rN*rN}$, and they met bellow equations as follows:

$$J'' = -\frac{1}{2}V^T(diag(D^T FD) - D^T FD)V + V^T D^T G \tag{7}$$

$$\begin{cases} F = \tilde{R} + \Psi^T(N)H\Psi(N) + \sum_{k=1}^{N-1}(\Psi^T(k)Q(k)\Psi(k)) \\ G = (\Psi^T(N)H\Phi(N) + \sum_{k=1}^{N-1}(\Psi(k)Q(k)\Phi(k)))X(0) \end{cases} \tag{8}$$

Where: V is the states of DHNN $V \in [-1,1]$, $L = r * N * M$. The DHNN has $r * N * M$ neurons. By comparing (6) and (7), the weights matrix and the domain value vector of DHNN can be expressed:

$$\begin{cases} W = diag(D^T FD) - D^T FD \\ I = D^T D \end{cases} \tag{9}$$

The performance index of system is equivalent to the energy function of Hopfield neural network. It is obvious weight matrices are symmetric and diagonal elements equal to zero and also non-negative, so the stability conditions of DHNN is satisfied, the requirements of the problem solution can be achieved through DHNN.

3 The Hopfield Network of Flexible Robot Realizing

$$\begin{bmatrix} \dot{x} \\ \dot{\theta}_1 \\ \dot{\theta}_2 \\ \ddot{x} \\ \ddot{\theta}_1 \\ \ddot{\theta}_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & K_{32} & K_{33} & 0 & 0 & 0 \\ 0 & K_{12} & K_{13} & 0 & 0 & 0 \\ 0 & K_{22} & K_{23} & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ \theta_1 \\ \theta_2 \\ \dot{x} \\ \dot{\theta}_1 \\ \dot{\theta}_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ K_{37} \\ K_{17} \\ K_{27} \end{bmatrix} F \tag{10}$$

$$y = \begin{bmatrix} x \\ \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ \theta_1 \\ \theta_2 \\ \dot{x} \\ \dot{\theta}_1 \\ \dot{\theta}_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} F \tag{11}$$

In which: $K_{12}, K_{13}, K_{17}, K_{21}, K_{23}, K_{27}, K_{31}, K_{32}, K_{37}$ (shown in [20]).

The formulas (10) and (11) are the state-space equations of simplified model of the flexible two-wheeled upright self-balance robot. Those are a group of simple equations and describe the characteristics of robot's input and output.

In the practical system, m_1 is the quality of the robot base, which is 7kg. m_3 is the quality of the lumbar spine of flexible robot, which is 1kg. m_2 is the quality of the upper body of flexible robot, which is 10kg. l_1 is the distance from the centroid of robot's equivalence base part to the central of rotational axis, which is 0.1m. l_2 is the distance from the centroid of upper body to the top of lumbar spine, which is 0.3m. In this paper, selects: Q=[1 0 0 0 0 0; 0 100 0 0 0 0; 0 0 100 0 0 0; 0 0 0 100 0 0; 0 0 0 0 1 0; 0 0 0 0 0 1], R=[1]. Using MATLAB order LQR(),the controllability rank criterion $rank(B\ AB\ A^2B\ A^3B\ A^4B\ A^5B) = 6$ and the observability rank criterion

and $rank(C\ CA\ C^2A\ C^3A\ C^4A\ C^5A) = 6$, so the system is completely controllable and observable and meets the conditions required for optimal control. The performance index of the system is equivalent to the energy function of the Hopfield neural network. It is obvious that the weight matrices are symmetric and the diagonal elements are equal to zero and also non-negative. The stability conditions of the DHNN are satisfied so the robot's balance control problem can be solved using the DHNN. M = 64, so the control value $u(k)$ used 64 neurons to express it in the k th moment. Since L is 384 ($L = r * N * M = 1 * 6 * 64 = 384$), there are 384 neurons in the network. The DHNN is a 384-order network of 384 neurons composed of the six-tuple $ncm_{DHN}^{(N)} = \langle V_G, A_G, IF, OF, WA, OA \rangle$,

Where: (1) the node set V_G is $\{V_i | i = 1,2,...,384\}$. V_i represent the neurons of the DHNN, I_i is the threshold of V_i , $u_i(t)$ is the volume of V_i in the t th moment, $o(t) = (o_1(t), o_2(t), ..., o_N(t))^T$ are the states of $ncm_{DHN}^{(N)}$ in the t th moment($t = 0,1,2,...$). (2) The link relation matrix A_G is equal to $W = \{D_{ij}\}_{384*384}$. The selection matrix D is

$$D = \begin{bmatrix} D1 \\ D2 \\ D3 \\ D4 \\ D5 \\ D6 \end{bmatrix} \begin{cases} Di = (Di1, Di2, \cdots, Di*64) \in R^{1*384} \\ Dij = \begin{cases} (2^{-1}, 2^{-2}, \cdots, 2^{-64}) \in R^{1*64} (1 \le i, j \le 64) \\ 0_{1*64} \quad i \ne j i = j \end{cases} \end{cases} \tag{12}$$

where W_{ij} is the connected weight from V_i to V_j .

(3) Input domain ($IF \subseteq V_G$).

(4) Output domain ($OF \subseteq V_G$).

(5) Working algorithm ($WA : IF \to OF$; $DHNN\_WA(o(0), o(t))$).

Step 1 (initialization): let t be zero, and determine the subset of working $V_G^{(t)}$.

$\{W_{ij}\}$, $\{I\}$ can be obtained from equation (12) below where ($W_{ii} = 0 (i, j = 1,2,...,384)$).

Step 2 (integration): calculate the integrated input of a neuron

$u_j(t+1) = \sum_{i=1}^{N} D_{ij} o_i(t) - I_i \quad (v_j \in V_G^{(t)})$, is the state inside u_i .

Step 3 (excitation): update the state of neuron i

($o_j(t+1) = \begin{cases} o_j(t) & v_j \notin V_G^{(t)} \\ -1 & v_j \in V_G^{(t)} \ and \ u_j(t) < I_j \\ +1 & v_j \in V_G^{(t)} \ and \ u_j(t) \ge I_j \end{cases}$). The output state of the other neurons remains unchanged ($X_j(t+1) = X_j(t) \quad (j = 1,2,...,384; j \ne i)$). Then select another neuron from the 384 neurons to update its state.

Step 4 (stop conditions): if $ncm_{DHNN}^{(N)}$ is stable then stop; or if the energy E is optimal or less than the initial value given (0.5), stop.
Step 5 (unconditional transfer): let $t = t + 1$, determine the new subset of working $V_G^{(t)}$ and transfer to Step2.

In fact, a 384 order DHNN network is uniquely determined through its connected relation matrix W and threshold vector $I = (I_1, I_2, ..., I_{384})^T$, and can be written as $ncm_{DHNN}^{(384)} = \langle W, I \rangle$ .The weights and threshold of the network can be obtained from the above equations and the network given a random initial value. When the system reaches stability the network operator is used to calculate the state variables and control values. The experimental results are as follows: V = [1 -1 -1 -1 1 -1 1 1 1 -1 -1 1 1 -1 1 1 -1 1 1 1 1 -1 1 1 -1 1 -1 1 1 -1 1 -1 -1 1 -1 1 1 1 -1 1 1 -1 1 -1 1 1 1 1 -1 1 -1 -1 -1 1 1 -1 -1 1 -1 -1 1 -1 1 -1 -1 1 1 1 -1 -1 -1 1 1 -1 -1 -1 -1 1 1 -1 -1 -1 -1 1 -1 -1 -1 -1 1 -1 -1 -1 1 -1 -1-1 -1 1 -1 1 1 -1 1 -1 -1 1 1 1 -1 1 1 1 -1 -1 -1 1 1 1 1 1 -1 1 1 -1 -1 -1 -1 -1 1 -1 1 -1 -1 1 1 1 -1 -1 1 -1 -1 1 11 1 -1 1 -1 1 -1 1 -1 1 -1 1 1 -1 -1 -1 1 -1 1 1 1 -1 1 -1 -1 1 -1 -1 -1 1 -1 -1 -1 1 -1 1 -1 1 -1 -1 -1 1 -1 1 -1 1 11 1 1 1 -1 -1 1 1 1 1 -1 -1 1 -1 1 1 -1 1 1 1 1 -1 -1 -1 1 1 1 -1 -1 -1 -1 1 1 1 -1 -1 -1 1 1 -1 1 -1 1 -1 1 1 1 1 1 -1 -1 -1 -1 -1 1 -1 -1 -1 1 1 -1 -1 -1 1 1 1 1 1 1 1 -1 -1 1 1 1 -1 -1 1 1 1 1 1 -1 1 1 1 -1 1 -1 1 -1 1 1 -1 1 1 1 -1 -1 -1 -1 1 -1 1 -1 1 -1 1 -1 -1 1 1 1 -1 -1 -1 -1 -1 -1 1 -1 -1 1 1 1 1 -1 1 -1 -1 1 1 1 -1 -1 -1 -1 -1 1 -1 1 1 -1 -1 1 -1 -1 -1 1 1 1 -1 1 -1 1 -1 1 -1 1 -1 -1 1 -1 -1 -1 1 1 -1 -1 1 1 1 -1 1 1 1 1 -1 -1 1 1 -1 -1 -1 1 1 -1 -1 1 -1 1 1 1 -1 1 -1 -1 -1 1]'

Table 1. State variables and control values of flexible robot

state / vary	$\dot{x}$	$\dot{\theta}_1$	$\dot{\theta}_2$	$\ddot{x}$	$\ddot{\theta}_1$	$\ddot{\theta}_2$
U	1.0e-013* (0.0000)	1.0e-013* (-0.0000)	1.0e-013* (-0.0000)	1.0e-013* (0.0000)	1.0e-013* (0.0000)	1.0e-013* (0.1421)
x1	1.0e-019* (0.0000)	1.0e-019* (0.0000)	1.0e-019* (0.0000)	1.0e-019* (0.0911)	1.0e-019* (0.6213)	1.0e-019* (-0.0030)
x2	1.0e-018* (0.0091)	1.0e-018* (0.0621)	1.0e-018* (-0.0003)	1.0e-018* (-0.0456)	1.0e-018* (-0.3106)	1.0e-018* (0.0015)
x3	1.0e-016* (-0.0005)	1.0e-016* (-0.0031)	1.0e-016* (0.0000)	1.0e-016* (-0.0162)	1.0e-016* (-0.1412)	1.0e-016* (0.0124)
x4	1.0e-016* (-0.0162)	1.0e-016* (-0.1412)	1.0e-016* (0.0124)	1.0e-016* (0.0124)	1.0e-016* (0.7036)	1.0e-016* (-0.0618)
x5	1.0e-014* (0.0008)	1.0e-014* (0.0070)	1.0e-014* (-0.0006)	1.0e-014* (0.0418)	1.0e-014* (0.3694)	1.0e-014* (-0.0305)
x6	1.0e-013* (0.0418)	1.0e-013* (0.3694)	1.0e-013* (-0.0305)	1.0e-013* (0.0303)	1.0e-013* (-0.2125)	1.0e-013* (0.1441)

Where U represents the robot's control value, x_i ($i = 1,2,...,6$) is the state volume of the i th step of the robot's final six steps (the infinite time domain is divided into

six steps numerous limited time domain), the energy value of the network is $E = -3.3263e + 009$, and the performance index is $J = 2.5431e - 028$.

The energy values of the several groups are same from the above tables, E=-3.3263e+009. The performance index is also the order of magnitude on e-028. The control values U of the several groups are the order of magnitude on e-013 from the above tables, have a very small amount of control exist too. As the balance control of biological and people is a continuous adjustment and maintain process. Instead of standing motionless, but kept control the balance. Therefore the control variables of the robot can not completely are zero, there should be a very small control variable.

The upper body angle, upper body angle speed, base angle, base angle speed, speed and acceleration of robot are the order of magnitude on e-013 from the above tables, have a very small amount of control exist too, the robot is in the balance. As the balance control of biological the control balance of robot is a dynamic process. If all of the variables were completely zero and standing motionless, it is inconsistent with the practical and biological characteristics. So there should be very small variables with the ideal. Since the other variables are the ideal, the control variables of the robot can not completely are zero. The robot's balancing is a continuous adjustment and maintain process. Neurons went to find the optimal balance point through association and memory method. Because the robot's balancing is a dynamic process not a static point, the state values of each group neurons are different.

4 Conclusion

The paper researches and analyzes the flexible two-wheel self-balance humanoid robot in physiology. The robot's balance controller is considered and successfully implemented using the Hopfield network methods. The reasonable energy function is defined for the non-linear and strong coupling robot. Analyze the convergence of the controller. Simulation experiments show the effectiveness of the method and proved the stability of system. The paper provides a theoretical basis for the control research of flexible two-wheeled upright robot, which will be a meaningful and value of scientific research platform. As a complex non-balancing system, the flexible two-wheeled self-balancing robot provides an object for research robot system kinematics and dynamics analysis, at same time, also provides a good experimental subject for studying the control of research, especially self-organizing balance control methods and techniques.

References

1. Urakubo, T., et al.: Motion control of a two-wheeled mobile robot. Advanced Robotics 15(7) (2001)
2. Karniel, A., et al.: Dynamic properties of the lamprey's neuronal circuit as it drives a two-wheeled robot. In: The SAB 2002 Workshop on Motor Control in Human and Robots: on the interplay of real brain and artificial devices, Edinburgh, Scotland (2002)
3. Bui, T.H., et al.: A simple nonlinear control of a two-wheeled welding mobile robot. International Journal of Control, Automation, and Systems 1(1) (2003)

4. Kim, Y.-H., et al.: Dynamic analysis of a nonholonomic two-wheeled inverted pendulum robot. Journal of Intelligent and Robotic Systems: Theory and Applications 44(1) (2005)
5. Kozlowski, K., Pazderski, D.: Stabilization of two-wheeled mobile robot using smooth control law: experiment study. In: The 2006 IEEE International Conference on Robotics and Automation, Orlando, Florida (May 2006)
6. David, P.A.: NBot Balancing Robot, a two wheel balancing robot. [EB/OL] (2007,2,1), http://www.geology.smu.edu/~dpa-www/robo/nbot/index.html (May 2007)
7. Thibodeau, B., Deegan, P., Grupen, R.: Static Analysis of Contact Forces with a Mobile Manipulator. In: IEEE International Conference on Robotics and Automation, Orlando, Florida, May 2006, pp. 4007–4012 (2006)
8. Salerno, A.: On the nonlinear controllability of a quasiholonomic mobile robot. In: Proc. IEEE ICRA, Taiwan, pp. 3379–3384 (2003)
9. Grasser, F., D'Arrigo, A., Colombi, S., Rufer, A.: Joe: A mobile, inverted pendulum. IEEE Trans. Ind. Electron. 49(1), 107–114 (2002)
10. Hopfield, J.J.: Neural networks and physical systems with emergent collective computational abilities. Proceedings of National Academic Science 79, 2554–2558 (1982)
11. Hopfield, J.J.: Neurons with graded response have collective computational properties like those of two-state neurons. Proceedings of National Academic Science 81, 3088–3092 (1984)
12. Hopfield, J.J., Tank, D.W.: "Neural" computation of decisions in optimization problems. Biological Cybernetics 52, 141–152 (1985)
13. Hopfield, J.J., Tank, D.W.: Computing with neural circuits: a model. Science 233, 625–633 (1986)
14. Hopfield, J.J.: Artificial neural networks. IEEE Circuits and Devices Magazine 4, 2–10 (1988)
15. Chang-Hong, F., Wei-Dong, C., Yu-Geng, X.: A neural networks-based approach to safe path planning of mobile robot in unknown environment. Acta Automatica Sinica 30(6), 816–823 (2004)
16. Lendaris, G.G., Mathia, K., Saeks, R.: Linear Hopfield networks and constrained optimization. IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics) 29(1), 114–118 (1999)
17. Sadati, N., Taheri, J.: Solving robot motion planning problem using Hopfield Neural Network in a fuzzified environment. In: Proceedings of The 2002 IEEE International Conference on Fuzzy Systems, vol. 1 & 2, pp. 1144–1149 (2002)
18. Tani, J.: Visual attention and learning of a cognitive robot. In: Gerstner, W., Hasler, M., Germond, A., Nicoud, J.-D. (eds.) ICANN 1997. LNCS, vol. 1327, pp. 697–702. Springer, Heidelberg (1997)
19. Ming-Ai, L., Xiao-Gang, R.: Optimal control with continuous Hopfield neural network Proceedings. In: 2003 IEEE International Conference on Robotics, Intelligent Systems and Signal Processing (IEEE Cat. No.03EX707), vol. 2, pp. 758–762 (2003)
20. Jianwei, Z., Xiaogang, R.: Dynamic Modelling and Control of flexible Two-wheel Upright Self-balance Humanoid Robot. Robot 31(2), 179–186 (2009)

Real-Time Simulation System of Virtual Picking Manipulator Based on Parametric Design

Haixin Zou, Xiangjun Zou*, Yinle Chen, Yan Chen, Quan Sun, Hongjun Wang, and Tianhu Liu

Key Laboratory of Key Technology on Agricultural Machine and Equipment (South China Agricultural University), Ministry of Education, Guangzhou 510642, China
Tel.: 13433955703
{zouhaixin123,xjzou1,poloyinle,xdf1239}@163.com, 420655308@qq.com, xtwhj@tom.com, liuparalake@scau.edu.cn

Abstract. Aiming at the complexity of operation and control of the agricultural picking manipulator, a simulation system of the manipulator based on virtual design was developed. Firstly, the mechanism models and mathematical models of the manipulator were constructed. The simulation platform of virtual picking manipulator was developed. The designers can make parametric design of the manipulator real-timely; secondly, constraint and routing design for the motion trace of the virtual manipulator were made to realize real-time simulation of the manipulator; thirdly, the positive and inverse solutions of the kinematical equation of the manipulator was solved in the system, obtained from which the parameters were transferred to the motion node of the manipulator in the virtual scene to control its motion; finally, the visual and interactive dynamic simulation of the manipulator's operation behaviors and controls was realized. The results shows that the system is practical and can realize optimized design and motion control.

Keywords: Lichi picking manipulator, parametric design, real-time interaction, motion control, simulation system.

1 Introduction

Since the original tomato harvesting robot was born in the United States in 1980s, studies on agriculture picking manipulator (PM), such as cucumber and tomato fruit PM the technology of agriculture fruit PM has made a great progress, and studies on different species of fruit PM have achieved preparatory results, such as the success rate of eggplant PM has reached 89%, the average time for picking was 37.4s [1]; Scarfe A.J etc developed a automatic PM with a intelligent Kiwi vision system, and its picking speed was 1/s [2]; Muscato. G. etc developed a prototype of orange PM, the finding and picking speed was 8.7 seconds, but it was not good enough to replace the manual completely after testing [3].

* Corresponding author.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 1205–1210, 2009.

Agricultural fruits are not only various but have unique biological characteristics living in uncertain and complex growing environment. For automatic agricultural fruits PM, it has to deal with problems such as the uncertainty of fruits location, different of their sizes and colors, vulnerability and flexibility, furthermore, including the unknown and poor weather conditions, irregular road scene and many other complex issues, which make the target location, the optimum design and control of the end-effector difficult, and slow the studies relatively. Therefore, in order to solve these problems, many researchers have made improvement on research methods. For example, three-dimension (3D) dynamic simulation was adopted to verify the mechanism's rationality and feasibility to the apple positioning [4]. The target location method of interactive virtual manipulator based on stereo vision was proposed [5]. Netherlandish used Mitsubishi RV-E2 for reference, and the degree of freedom of manipulator is 7 after refit, including mechanic paw and cutter [6]. If industrial machine was used to refit, the structure would be too big to suit fruit picking.

The emphasis of the studies above is not concerned much with parametric design of the PM based on real-time interactive simulation. Recently, modern design methods like parametric design and its modeling and simulation technology have become effective methods in the field of agricultural biological harvesting. This paper put forward the research of the real-time simulation system of the virtual PM. With the simulation system, parametric method was adopted to design the PM, and to make real-time simulation in virtual environment (VE). Finally the test platform was used to check the motion control algorithm and its effectiveness. The results show that the PM based on parametric design technology and virtual reality technology has great advantages, and people can make simulation and optimal mechanism design for the picking behavior of the manipulator in VE, which has a great practical value.

2 Function and Framework of the Interactive Simulation System

The main functions of the interactive simulation system for PM are: according to the user's requirements and commands, a 5-DOF joint-type manipulator is designed based on parametric design to generate a new model, and a real-time 3D virtual scene is generated to simulate the picking scene of virtual PM.

The simulation system essentially consists of two parts: the first part is parametric design which is based on 5-DOF joint-type manipulator. For the characteristics of different fruits, growth environment and users' requirements, design parameters are input to the design interface, and then the system will transfer these parameters into the simulation control solving module to solve the geometry and location information of the manipulator. Finally, a new model of the manipulator is reconstructed for the information. The second part is an interactive 3D visual simulation. The simulation process is: firstly, import the need model into the -dimensional virtual scene; secondly, input the motion parameters of manipulator, such as the amount of the rotations of each joint, the coordinates of the end-effector to achieve, etc. The system will call the simulation control solving module to calculate the manipulator's motor angle, the steering and other information; thirdly, the joint solution is combined with path planning and others to get the movement simulation of the PM in VE. This

simulation movement is a dynamic change. With the change of the input parameters of the manipulator or/and the motion parameters, the virtual manipulator (VM) makes corresponding changes. According to the function and design principles, the system architecture of the simulation system of the virtual PM based on the parametric design is shown in Figure 1. Among them, the main function of the user interface is to achieve the interactive communication between the users and the simulation system and assist users to input of users' command and parameters.

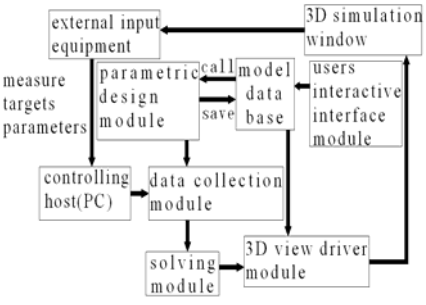


Fig. 1. Function and Framework of the Interactive Simulation System [7]

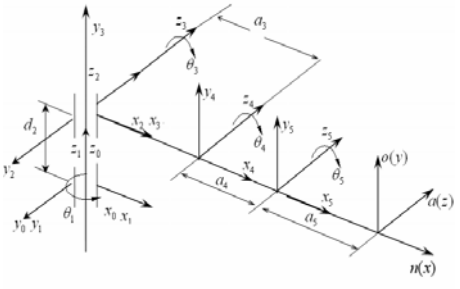


Fig. 2. the Coordinate System of the Manipulator's Link

There are three functions of data acquisition module: (1) To accept the input of users' parameters and to complete the maintenance and update of data; (2) To collect the data in real time, and to transform users' configuration parameters and real-time control commands to variable value required, and then to transfer the variable value to the solving module; (3) To transform the data of the PM's position and posture input by the host to the needed format for motion equation, and pass it to solving module.

Solving module is the most important part of the simulation system, and its functions are as follows: (1) To call the corresponding database of 3D model according to user's choices; (2) To control the behavior and process the simulation system in real time; (3) To get the location and the posture vector of the manipulation through solving equations of motion manipulator by the manipulator's state parameters; (4) The manipulator's relevant parameters solved are exported to 3D scene to drive module to show the visual scene. It will assist users to optimize the manipulator's parameters and make a motion control strategy.

According to the model of user's calls and the derived parameters from the solving module, 3D scenes drive module completes the show of the3D scenes.

3 Platform and Interface Technology of the Simulation System

The simulation system of PM adopts succession environment of the VC++ as the development platform, takes full advantage of simulation functions of the EON studio and its SDK, and achieves the control of 3D simulation through calling the API function library of the EON in VC++. EON Studio is used to develop real-time 3D for multimedia applications. After models are imported into the EON, the behaviors of 3D model can be driven by the route and event mechanism.

4 Virtual PM Modeling and Simulation of 5-DOF

An important function of picking simulation is the real-time control of the simulation. In this simulation system, in accordance with an external equipment including the vision system, other sensors, and the data users input directly, the manipulator's movement equation is solved in real-time to get the its position and orientation. The virtual scenes are driven and the picking motion of the VM is simulated.

4.1 Manipulator Picking Kinematics Model with 5-DOF [8, 9]

Kinematics control is based on kinematics model. Therefore, many researchers have conducted manipulator kinematics model modeling then analysis based on it. In this paper, the manipulator's mathematical model is described based on a PM with 5-DOF.

According to joint's coordinate system and the four characteristic parameters (d, a, α, θ) of the generalized link, the coordinate system is shown in Figure 2. Transformation orientation matrix of the link *i* relative to link *i-1* is shown as follows:

$$^{i-1}T_i = Rot(x_i, \alpha_{i-1})Trans(x_i, a_{i-1})Rot(z_i, \theta_i)Trans(z_i, d_i) \tag{1}$$

Where $Rot(x_i, \alpha_{i-1})_{i-1}$ is notated the rotation angle α_{i-1} around the axis x_i; $Trans\,(x_i, a_{i-1})$ is notate the distance a_{i-1} along the axis-translation; the rest of analogy. Multiplied by the transformation matrix connecting rod, the transformation matrix is as follows:

$$^{0}T_5 = {}^{0}T_1(\theta_1)^{1}T_2(h_2)^{2}T_3(\theta_3)^{3}T_4(\theta_4)^{4}T_5(\theta_5) \tag{2}$$

To solve the joint variables, it also needs to be multiplied transformation matrix which can be solved by MATLAB to obtain manipulator's position and orientation.

The positive solutions are that: to get the position and orientation of the end-effector relative to the base coordinate system while given the known link parameters and joint angle vector.

4.2 Simulation Control Flow of the Manipulator

When the data of virtual robot model and others is read, they have the attributes like position and orientation, which can be set by designing the EON node routing, and then realized the virtual picking motion of virtual manipulator (VM). In the main driver cycle of EON view, real-time response functions are invoked to the virtual robot, and the VM's position and orientation are set according to the calculations. The simulation control flow chart of the virtual PM is demonstrated in Figure 3.

In the virtual simulation system, the position of viewpoint can be fixed or changed by the mouse and keyboard commands. In the concrete realization of the process, the point of view needs control commands to control the viewpoint position.

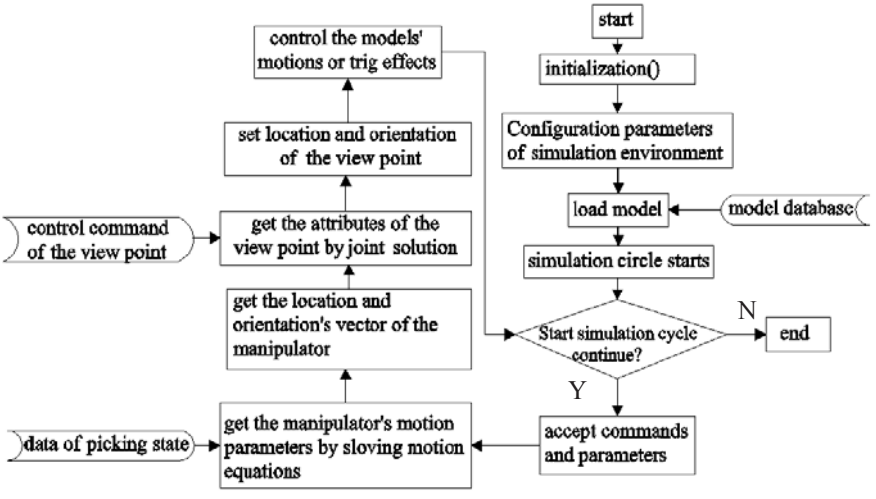


Fig. 3. Control Flow of the virtual manipulator Simulation [7]

Fig. 4. VM and the Meter Display

Fig. 5. Prototype of the Manipulator

4.3 Virtual Manipulator and Interaction of Manipulator Prototype

In order to simulate the situation of virtual manipulator tracking the picking target better, and assist users to test the function of the external input device, the simulation system was designed to use meter to display the parameters of PM and picking target real-timely and their relationship, as shown in Figure 4.

The data of external input device is transferred by a self-developed data collection chip and USB2.0 interface procedure. The interface connects external robot testing hardware and the simulation system of interactive VE in the computer (as showed in figure 5, and motion data is collected and input to the simulation system real-timely. Finally, the communication between the manipulator prototype and the simulation system is realized and the motion of the manipulator is driven.

5 Conclusions

Based on the virtual PM model of 5-DOF and interactive manipulator parametric design, a interactive virtual PM simulation system was developed by using the real-time and authenticity of EON simulation software whose common modularize system architecture and the control process mechanism graph of the VM simulation in EON/C++ have the attribution of generality. The simulation system can form a control closed-loop with control engine and external input device, and is suitable for simulating the situation of the manipulator tracing the picking target, and assist users to test the performance of the external input device, and the principle and simulation of aided testing also can be used for field robots.

Acknowledgments. This work was supported by the National Natural Science Foundation of China with Research Grant, 50775079, Natural Science Foundation of Guangdong, China, 925106420100009 07006692, 05006661, 9151064201000030, and Ph.D. Program Foundation of Ministry of Education of China, 200805640009.

References

1. Song, J., Sun, X.Y., et al.: Design and Experiment of Opening Picking Robot for Eggplant. J. Nongye Jixie Xuebao/Transactions of the Chinese Society of Agricultural Machinery 40(1), 143–147 (2009)
2. Scarfe, A.J., Flemmer, R.C., et al.: Development of an Autonomous Kiwifruit Picking Robot. In: 2009 4th International Conference on Autonomous Robots and Agents, pp. 380–384 (2009)
3. Muscato, G., Prestifilippo, et al.: A Prototype of an Orange Picking Robot: Past History, the New Robot and Experimental Results. Industrial Robot 32(2), 128–138 (2005)
4. Baeten, J., Donne, K., et al.: Autonomous Fruit Picking Machine: A Robotic Apple Harvester. In: 6th International Conference on Field and Service Robotics-Chamonix, France, pp. 531–539 (2007)
5. Zou, X.J., Zou, H.X., et al.: Research on Manipulator Positioning Based on Stereo Vision in Virtual Environment. J. Key Engineering Materials 392-394, 200–204 (2009)
6. Van Henten, E.J., Van Tuijl, B.A.J., et al.: Field Test of an Autonomous Cucumber Picking Robot. J. Biosystems Engineering. 86(3), 305–331 (2003)
7. Wang, L., Meng, X.F., et al.: Design and Realization of Flight Scene Simulation System. J. Journal of System Simulation. 20(1), 73–75 (2008)
8. Xiong, Y.L.: Foundation of Robot Technology/ 术 . Wu Han, China (1996)
9. Jun, L.: The Association Positioning of Picking-Machine and Vision based on Virtual Design, pp. 8–13. South China Agriculture University, D. Guangzhou, China (2009)

Study of Modeling and Simulating for Picking Manipulator Based on Modelica

Yan Chen[1,*], Shuang Jin[1], Xiang-jun Zou[1], Dong-fen Xu[2], and Wei-liang Cai[1]

[1] Key Laboratory of Key Technology on Agricultural Machine and Equipment (South China Agricultural University), Ministry of Education, P.R. China, 510642
[2] Research Center of Human Computer Interaction, College of Informatics, South China Agricultural University

Abstract. In order to process the study of Picking Manipulator for litchi, multi-field unified modeling language Modelica is used to model and simulate Picking Manipulator. In this paper, the structure of Picking Manipulator is analyzed and the mathematical model of manipulator control system is built by adopting multi-cooperating controlling model. More, the process of modeling Picking Manipulator under Modelica/Dymola is introduced, and the model of Picking Manipulator system is built while the extensible modeling base is established. Finally, the model will be simulated and the validity of models will be validated. Through the thesis, the basic for the further study of Litchi Picking Manipulator will be provided.

Keywords: Picking Manipulator, Modelica, Modeling, Simulating.

1 Introduction

Fruit picking, which cost too much and need a lot of work, is an important link in agricultural production. The using of harvesting robot can reduce picking cost and improve working conditions. Harvesting robot research and development have already experienced nearly 30 years. Recently, many colleges and universities have begun the study of modeling and simulating harvesting robot in china and made process [1-2].

The system of robot deals with many fields, such as machine, electronic, control. The whole capability need unite simulating of many fields. There are two simulating methods: directly modeling and indirectly modeling [3]. The indirectly modeling need simulate among many fields. There are some errors in translating data. Now, many models for robot systems use this method. The directly modeling do not depend on professional software in each fields. It adopts uniform modeling language to constitute models of different fields and can achieve the seamless integration and data changes, which can improve the efficient of modeling [4].

Modelica is a kind of unite modeling language among many fields. Since 1977, it has been frequently used in other countries [5]. Many colleges have begun the study

* Yan Chen (1964-), Female, Born in Guangdong Province, Associate Professor, Research interests: Virtual Reality Intelligent Design and Manufacturing.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 1211–1216, 2009.

of using Modelica now, for example, modeling and simulating of cars with many axes [6], motor-linkage [7], steamer [8], capability of cars [4] and so on. The study shows that Modelica fits models among many fields. It not only can provide accurately simulating but also can improve the efficiency of modeling.

Picking manipulator is a complicated electromechanical system, which is coupling. To study the whole capability of picking manipulator, the litchi picking manipulator is simulated by using Modelica [9]. The model of picking manipulator system is constructed and simulated in Dymola through analyzing the structure of picking manipulator and building the mathematical model of control system.

2 The Mechanically Ontology Structure of Picking Manipulators

Because of the complexity of picking objects and picking environment, the litchi picking manipulator adopts joint type structure,and it has 5-DOF. Joint 1 is the moving joint and others are rotating joints. The five joints are driven by five DC Motor with different power and torsion, and slowdowned by gear reducers. The diagram is shown in Fig. 1.

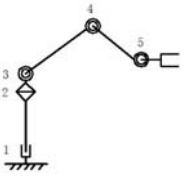


Fig. 1. Constructional diagram of the picking manipulator

3 The Mathematical Model of Manipulator Control System

The design of manipulator usually adopts simplified methods,which ignores coupling effect between manipulator's joints. Then the entire control system will be divided into several independent single-input-to-output system, and each independent system will use the same bottom control algorithm [10].

3.1 The Model of Transfer Function

According to physics and electromagnetic theory, DC motor's mathematical model can be gained. And the DC motor's transfer function between input-voltage and output-speed is gained by Laplace transform which eliminate intermediate variables, given by Eq. (1) [9].

$$\frac{\Omega(s)}{U_a(s)} = \frac{K_t}{L_a J s^2 + (L_a B + R_a J)s + R_a B + K_e K_t} \tag{1}$$

Where K_t and K_e are constant, representing the motor's torque constant and induced electromotive force's constant. Vectors $U_a(s)$ and$\Omega(s)$ represent the armature voltage and motor speed, and L_a and R_a represent the armature circuit inductance and resistance, and J and B represent the motor rotational shaft's equivalent inertia and damping.

If the armature circuit inductance L_a and the motor rotational shaft's equivalent damping B are neglected, Eq.(1)can be transformed into Eq.(2). The transfer function of DC motor speed control system will be approximated as a habitual links:

$$\frac{\Omega(s)}{U_a(s)} = \frac{1/K_e}{\tau_m s+1} \tag{2}$$

Where the time constant τ_m represent $\tau_m = R_a J/K_t K_e$.

3.2 The Mathematical Model of Single Axis Controlling System

To improve the stability and precision of the controlling system, the multi-control is adopted, which adds forward feedback in the negative feedback PI and uses the angle of optical encoder as gain output variables, the picking manipulator's single axis multi-control system is shown in Fig.2.

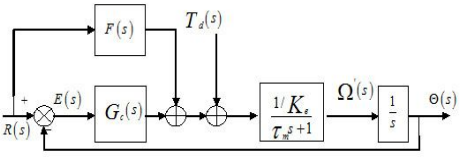


Fig. 2. Diagram of the multi-control system

Where K_P and K_I are constant, representing the ratio and integral coefficient. Vectors $F(s)$ and $G_c(s)$ represent motor's forward feed controller and $G_c(s) = K_P + K_I /s$, and $G_P(s)$ and $E(s)$represent the object's transfer function and system error, where $R(s)$ and $\Theta(s)$ represent reference input location and optical encoder position, $G_P(s)=1/sK_e(\tau_m s+1)$ and $E(s)= R(s)- \Theta(s)$,and $T_d(s)$ represent the motor rotational shaft's equivalent total torque (non-load torque and load torque).

Based on above, the system output can be expressed by Eq. (3):

$$\Theta(s) = G_p(s)\left[G_c(s)E(s) + F(s)R(s) + T_d(s)\right] \tag{3}$$

The error of this system is:

$$E(s) = \frac{1-F(s)G_p(s)}{1+G_c(s)G_p(s)}R(s) - \frac{G_p(s)}{1+G_c(s)G_p(s)}T_d(s) \tag{4}$$

The first item of the error function is aroused by the reference input, while the second is aroused by the anti-jamming moment. If we assume $F(s) = G_P^{-1}(s)$, then the forward feedback control can transfer as the inverse function. So the error aroused by reference input is zero.

4 Modeling the Picking Manipulator System

The model of picking manipulating system based on Modelica and dymola is composed by the ontic model of manipulator, motor model, single axis control model, many axis control models and so on.

4.1 The Ontic Model of Picking Manipulator

As mentioned in the frontal parts, picking manipulator is composed by joints and member bars. When building the model, the models in multibody are called and the parameters of models are setuped. The rotating joints can use Modelica.Actratedrevolute modules while moving joints adopt screw-drived. Many models from Modelica bank are called to build screw-drived model with self-lock. Finally according to the Modelica and signal standard, the ontic model of picking manipulator is built, shown in Fig. 3. At last, models are put into the picking manipulator bank.

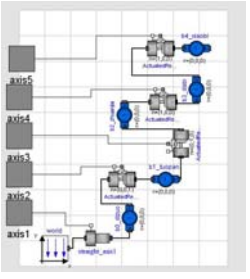


Fig. 3. The ontic model of picking manipulator

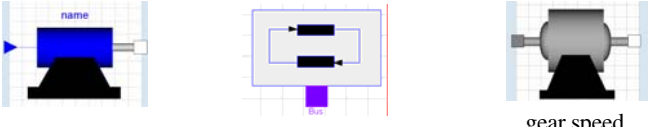
4.2 The Model of DC Servo Motor

According to the equivalent circuit of DC servo motor [9], the packaged components of DC servo motor model has been built, shown in Fig. 4(a).

4.3 The Model of Picking Manipulator Control System

The multi-control was adopted in the model of picking manipulator, while the multi-control models was built by calling the Modelica standard banks. The package of the component in this system is shown in Fig. 4(b).

The joint of picking manipulator and motor are connected by the gearbox, so that the gear speed transfer devices model is essential. The packaged components are shown in Fig. 4(c).

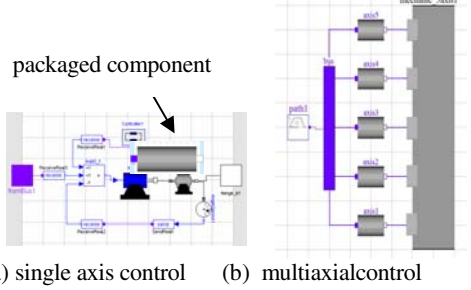


(a) motor model (b) multi-control model gear speed transfer model

Fig. 4. The packaged components of model

After the information collected by the computer from the video gathering system, the controller transport signals to the DC servo motor, and the motor joints the picking manipulator by gear reducer. The single axis control model was build according to the front principle and Modelica model banks, shown in Fig. 5(a).

When building multiaxial control picking manipulator system



(a) single axis control (b) multiaxialcontrol

Fig. 5. control system model

(Fig. 5(b)) by using the single axis control model, the construction of path planning model and the signal switches are need to build.

5 Analysis of the Simulation

To make sure the validity of the model, a simulation is done in Dymola by taken the single axis control manipulator for example. To compare the transforming curves of output imports under different parameters, the positional stepped singles are inputted while simulating.The simulating results are shown in Fig. 6.

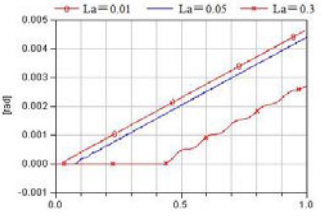


Fig. 6. Effect of La on curves of the output position

The curves shows that: The curves of output position are lines and there is little delay phenomenon while the control system was directly driven by the positional stepped singles. It explains that the precision of control system is very high and verifys the validity of the mathematical models and simulation. Curves also show that different inductance values can affect the response time, output speed (displacement) significantly. With the increase of La, the system's time-delay increase, the output speed decrease, and the noise can be improved. It shows that the system can be simulated by changing models' parameters. It can help us to optimize the manipulating system.

6 Conclusions

In this paper, the unite modeling language Modelica for multi-fields was used to build the picking manipulator system model and the open, reusable model banks. And its validity was verified through simulating. The farther study of system's simulation can instruct the optimization and development of manipulator, and moreover, it is convenient to construct new simulating models for simulation by using the models. It is of great significance and practical value.

Funding Projects

National Science Foundation of China (50775079)
National Science Foundation of China of Guangdong Province (07006692)
Principal Foundation of South China Agricultural University (2009X004)

References

[1] Zou, X.-j., Luo, X.-w., Lu, J., et al.: Modeling for Behavior and Simulation of Agriculture Mobile Robot in Virtual Environment. Journal of System Simulation 18(8) sup. 2, 551–554 (2006)

[2] Cui, Y.-j., Zhang, Z.-l., Fan, L., et al.: Study on the Kinematics Simulation of Fruit Picking Manipulator Based on ADAMS. Journal of Agricultural Mechanization Research 4, 59–61 (2008)

[3] Zhao, J.-j., Ding, J.-w., Zhou, F.-L., et al.: Modelica language and Its Unified Modeling and SimulationJournal of System Simulation 18(S2), 570–573 (2006)
[4] Ren, W.-q., Song, J.: Application of Multi-domain Unified Modeling Method in Vehicle Performance Simulation. Computer Simulation 25(11), 274–277 (2008)
[5] Ferretti, G., Magnani, G., Rocco, P., et al.: Modeling and Simulation of a gripper with Dymola. Mathematical and Computer Modeling of Dynamical Systems 12(1), 89–102 (2006)
[6] Zhang, G.-h., Yang, S.-w., Wang, Z.-h.: Object-Oriented Modeling and Simulation of Multi-axis Vehicle Dynamics. Tractor & Farm Transporter 5, 73–78 (2007)
[7] Zhao, Y.-x., Deng, Y.-j., Chen, X.: Dymola-Based Motor-linkage Mechanism Modeling and Simulation. Machine Development 8(35), 60–61 (2006)
[8] He, J.-h.: Modeling and Simulation of Steam Turbine System Based on Modelica. North China Electric Power University, Baoding (2008)
[9] Jin, S.: The Modeling and Simulation of the Robot's Control System in the Multi-domain Physical System. South China Agricultural University, Guangdong (2008)
[10] Xiong, Y.-l., Tang, L.-x., Ding, H., et al.: The Base of Robot Technology. Huazhong University of Science Publisher, Hubei (1999)

Automatic Reconstruction of Unknown 3D Objects Based on the Limit Visual Surface and Trend Surface

Xiaolong Zhou[1], Bingwei He[1,2], and Y.F. Li[3]

[1] School of Mechanical Engineering & Automation Fuzhou University, Fuzhou, China
[2] State key laboratory of precision measuring technology and instruments Tianjin University, Tianjin, China
[3] Department of Manufacturing Engineering and Engineering Management City University of Hong Kong, Hong Kong
mexlzhou@yahoo.com.cn, bw_he@yahoo.com.cn, meyfli@cityu.edu.hk

Abstract. This paper presents a new planning approach of generating unknown 3-D models automatically. The new algorithm incorporates the limit visual surfaces with the trend surfaces and selects the suitability of viewpoints as the NBV on scanning coverage. The limit visual surfaces and trend surfaces are modelled by means of the known boundary region data obtained from initial view. The optimal design method is used to obtain the maximal visible area of next viewpoint and correspondding pose parameters in left and right planning process respectively. And the position that can obtain the maximal visible area is defined as the next best view position. The reconstrcted result of real model showes that the method is effective in practical implementation.

Keywords: View Planning, Limit visual surface, Trend surface, Next best view, Three-dimensional reconstruction.

1 Introduction

Automatic reconstruction of unknown 3-D objects has been of great importance role in the areas of machine vision, object recognition, and automatic modeling. The planning for automatic measurement of 3-D object is to search an optimal space pose with a sequence of viewpoints. According to the sensor configuration and task specifications, it determines a least viewpoint number and best spatial distribution so that the measurement task can achieve the highest efficiency. Recently, many methods on determining the location of the best next viewpoint have been proposed for different kinds of sensors [1-7]. For example, Bottino and Laurentini [3] present a general approach to interactive, object-specific volumetric algorithm, based on a necessary condition for the best possible reconstruction to have been performed. The proposed algorithms just only suit to simple convex polyhedra but not for general polyhedra, especially nonconvex polyhedra. Tarbox and Gottschlich [4] propose a model based approach. In which the measurability matrix is computed and the next view is based on glancing angles and the part of "difficulty to view." And the strategy proposed in [5] is similar but adds an incremental process and a constraint on sensor measurement error. Recently, Scott [6] presents a new view planning algorithm based

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 1217–1223, 2009.

on the modified measurability matrix (3M) which is an enhancement and extension of Tarbox's measurability matrix (2M) concept. Sablatnig et al. [7] present an approach to next view planning for shape from silhouette for 3D shape reconstruction with minimal different views. The next view position is determined by comparing the difference between current image and acquired image. However it leads to a larger computing time with a larger number of viewpoints.

The goal of this paper is to develop a new strategy of automatic viewpoint planning for unknown 3-D object measurement and reconstruction.

2 The Visual Region of the Vision System

The developed laser line scanning vision system is shown in Fig.1: it consists of an analogue camera, laser line generator and stepper motor driven linear slide providing X, Z scanning motions for the camera and laser. The part is placed on a rotary table that can rotate 360 degrees with Z and linear slide providing Y motions. The camera and laser line generator are mounted on the stepper motor driven linear slide, with the laser line perpendicular to the rotary workstation and an angle between the camera's optical axis and the laser line. The objective of the vision system is to provide a 3D profile of the object based on a coordinate system defined on the rotary table.

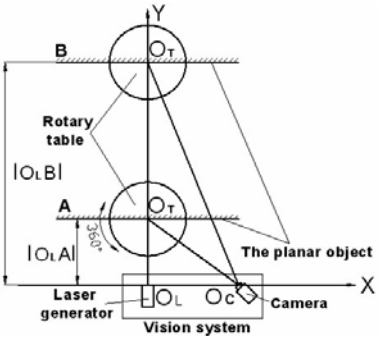


Fig. 1. The laser line scanning vision system. A (B) denotes the nearest (farthest) measurement position of the vision system. O_T is the rotary table center.

2.1 Determination the Limit Visual Region of the Vision System

A 3D object surface is measured by the vision system with both them at a certain distance, and 3D surface is fitted simultaneously. If the fitting accuracy belongs to the allowable errors range, then this distance is regarded as effective focus distance, or else as invalid. According to this method, it is easy to obtain the effective depth of field (DOF) range of the vision systems by moving workstation along the Y direction (as shown in Fig.1). In order to simplify the fitting process of the measurement data, here the planar object is used and the fitting accuracy is set to 0.02mm. From the experiments, the nearest and farthest measurement distance is 64mm and 186mm respectively. So the effective DOF of the system is 122mm.

If a surface point beyond the field of view (FOV), it will not be detectable to a common CCD camera. But in this vision system, the FOV is satisfied two constraints: 1) the laser line is not occluded by the object surface in order to be detected by CCD camera; 2) CCD camera is not occluded by the object surface. So, FOV is determined by elimination of these two blocks. Assume the angle between the normal vector of the object surface and the laser plane is defined as the left (right) visual angle. Then the left (right) limit visual angle is defined as the maximal rotary angle which the projection of the laser on the plane could be detected by CCD camera at one certain position. From the experiments, the relationship between left (right) limit visual angle θ_{il} (θ_{ir}) and the measurement distance d are obtained as (1), (2) respectively:

$$\theta_{il} = \begin{cases} 0.0053d + 0.2731, & 64 \le d \le 160 \\ 1.1211, & 160 < d \le 186 \end{cases}. \tag{1}$$

$$\theta_{ir} = \begin{cases} 0.0103d + 0.2278, & 64 \le d \le 116 \\ 1.4226, & 116 < d \le 186 \end{cases}. \tag{2}$$

Where the units of θ_{il} and θ_{ir} are radian.

Assume that the point (x, y, z) is in effective focus depth, a limit visual curve on XY plane is obtained according to above presents. When z is varied, the surface is constructed from a series of the limit visual curves, which is denoted as the limit visual surface. It represents the limit visual position of the object surface which is even visible to our vision system. And the equations of the left and right limit visual curves are obtained based on equations (1) and (2).

$$\begin{cases} x = -188.7\ln(\sin(\dfrac{y + d_0 + 51.5}{188.7})) + C_1, & y_0 \le y \le 160 - d_0 \\ y = -2.1x + C_2, & y > 160 - d_0 \end{cases}. \tag{3}$$

$$\begin{cases} x = 97.1\ln(\sin(\dfrac{y + d_0 + 22.1}{97.1})) + C_3, & y_0 \le y \le 116 - d_0 \\ y = 6.7x + C_4, & y > 116 - d_0 \end{cases}. \tag{4}$$

Therefore, if the prior 3D data cloud of the object surface are obtained, the limit visual surface of the vision system is obtained from a series of the limit visual curves acquired according to the equations (3) and (4).

2.2 Obtaining the Trend Surface of the Object

To an unknown object with complex surface shape, the limit visual surface model constructed only to reflect the maximal information of the unknown object. So, here we incorporate the trend surface to predict the unknown object surface.

The trend surface is a function of two orthogonal coordinate axes which can be represented by

$$z = f(x, y) + e\ . \tag{5}$$

in which the variable z at the point (x, y) is a function of the coordinate axes, plus the error term e. This expression is the generalized form of the General Linear Model (GLM), which is the basis of most trend methods.

The function $f(x, y)$ is usually expanded or approximated by various terms to generate polynomial equations. For an norder three-dimensional surface, the form of the power series is given by

$$f(u,v) = \sum_{i=0}^{n} \sum_{j=0}^{i} b_{ij} u^j v^{i-j} \quad . \tag{6}$$

where u and v are the coordinates on an arbitrary orthogonal reference system, b_{ij} is the constant coefficient of the surface (b_{00} is the surface base).

3 The View Planning Method

Suppose the initial knowledge of the model has been got at initial view (shown as the solid line $A_1B_1C_1D_1$ in Fig.2). Then the view planning strategy presented here is:

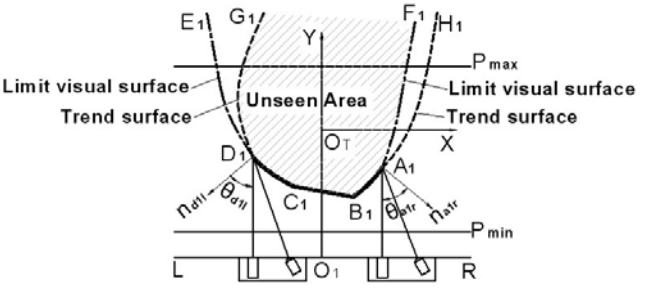


Fig. 2. The knowledge of initial view and its limit visual surfaces and trend surfaces

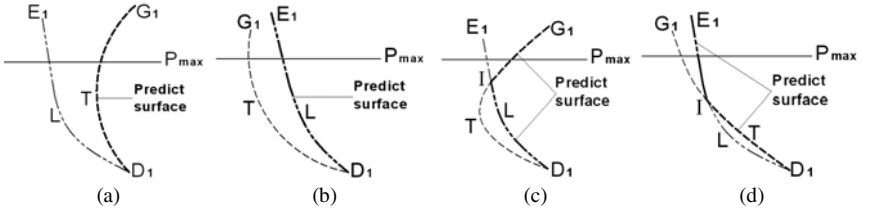


Fig. 3. The relationship with the limit visual surface and the trend surface. T denotes the trend surface, L denotes the limit visual surface, and $P_{\max}$ is the farthest measurement position of the vision system. I is the intersection point of the limit visual surface with the trend surface.

1) Firstly, construct the left (right) limit visual surface D_1E_1 (A_1F_1) and trend surface D_1G_1 (A_1H_1) according to the partially acquired object surface (shown as Fig.2). And determine the predict surface to plan the next view according to the relationship between the limit visual surface and the trend surface (shown as Fig.3). In Fig.2, the left (right) predict surface is D_1G_1 (A_1F_1). Besides, during the process of the determination of the next viewpoint, not only the maximal predict surface should be detected, but also the boundary data of known model should be seen in order to easily register data points.

2) Secondly, regard the visible area of left (right) predict surface of next viewpoint as the object function, the rotation angle θ and translation distance d as the variables, the visibility of known boundary data as the constraint condition. Then adopt optimal design method to obtain the maximal visible area S_l (S_r) and corresponding position parameters (θ_l (θ_r) and d_l (d_r)) in left (right) planning process. And define this position as the NBV candidate position.

3) Finally, compare the visible area got from above two NBV candidate positions, and select the larger one as the final NBV position.

4 Experiment

The experiment was carried in our laboratory for automatic reconstruction of the part model. The initial knowledge of part model was acquired from first view in random position (shown as Fig.4(a)). Then the left and right limit visual surfaces and trend surfaces were constructed (shown as Fig.4(b)). And the experimental data of NBV candidate positions of viewpoint 1 were shown as table 1. (Suppose the rotary table counterclockwise rotating was positive, otherwise was negative)

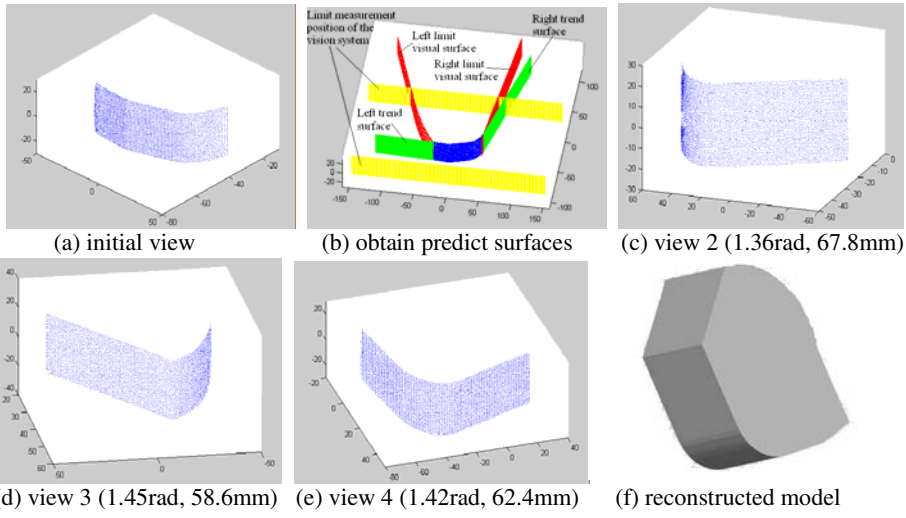
Table 1. The NBV candidate positions data of initial viewpoint 1

d_0 (mm)	θ_{1l} (rad)	d_{1l} (mm)	S_{1l} (mm^2)	θ_{1r} (rad)	d_{1r} (mm)	S_{1r} (mm^2)
160	1.36	67.8	1.93×10^4	1.69	63.0	1.83×10^4

In Table1, For $S_{1l} > S_{1r}$, so the next best view was the position that rotary table counterclockwise rotated 1.36rad and translated 67.8mm to. Then we could get the viewpoint 2 (shown as Fig.4(c)). Similarly, the residual process of the reconstruction object were shown as Fig.4(d-e). The final reconstructed model was shown as Fig.4(f). The whole reconstruction result was shown as Table 2.

Table 2. The reconstruction result analysis

View numbers	Actual volume (mm^3)	Reconstructed volume (mm^3)	Volume error (%)	Reconstruction presicion
4	332767	335248	+0.75	0.032



(a) initial view (b) obtain predict surfaces (c) view 2 (1.36rad, 67.8mm)

(d) view 3 (1.45rad, 58.6mm) (e) view 4 (1.42rad, 62.4mm) (f) reconstructed model

Fig. 4. The reconstruction process of the Part model

5 Conclusions

In this paper, we presented a new approach of generating 3-D models automatically, putting emphasis on planning of NBV. The proposed algorithm incorporated the limit visual surfaces with the trend surfaces obtained according to the partial known object knowledge. The final NBV was the position that obtained the largest visible space area among viewing points by using the optimal design method. And the experimental result showed that the method was effective in practical implementation.

Acknowledgments. The work was supported by a grant from the National Natural Science Foundation of China (No.50605007) and Program for New Century Excellent Talents in Fujian Province University (Project No. XSJRC2007-07) and State key laboratory of precision measuring technology and instruments and a grant from the Research Grants Council of Hong Kong (Project No. CityU117507)

References

[1] Larsson, S., Kjellander, J.A.P.: Path planning for laser scanning with an industrial robot. Robotics and Autonomous Systems 56, 615–624 (2008)
[2] Li, Y.F., Liu, Z.G.: Information entropy-based viewpoint planning for 3-D object reconstruction. IEEE Trans. on Robotics 21(3), 324–337 (2005)
[3] Bottino, A., Laurentini, A.: What's NEXT? An interactive next best view approach. Pattern Recognition 39, 126–132 (2006)
[4] Tarbox, G.H.: Planning for Complete Sensor Coverage in Inspection. Computer Vision and Image Understanding 61(1), 84–111 (1995)

[5] Scott, W.R., Roth, G., Rivest, J.F.: View planning for automated three-dimensional object reconstruction and inspection. ACM Computer Surveys 35(1), 64–96 (2003)
[6] Scott, W.R.: Model-based view planning. Machine Visionand Applications 20(1), 47–69 (2009)
[7] Sablatnig, R., Tosovic, S., Kampel, M.: Next view planning for shape from silhouette. In: Computer Vision (CVWW 2003), Czech Pattern Recognition Society, pp. 77–82 (2003)

Hyper-chaotic Newton-Downhill Method and Its Application to Mechanism Forward Kinematics Analysis of Parallel Robot

Youxin Luo

College of Mechanical Engineering, Hunan University of Arts and Science Changde, 415000, P.R. China
LLYX123@126.com

Abstract. Aimed to the nonlinear equations-finding problem that only one solution is found with the Newton and quasi-Newton methods and sometime no solution is found when the iteration diverges, combining Newton-downhill method with hyper-chaotic system, hyper-chaotic Newton-downhill method based on utilizing hyper-chaotic discrete system to obtain locate initial points to find all real solutions of the nonlinear questions was proposed. Using cosine matrix method and link length constraint we establish nine variable equations of the general 6-SPS mechanism. The numerical example in forward kinematics analysis of the general 6-SPS parallel platform shows that all real solutions have been quickly obtained, and it proves the correctness and validity of the proposed method.

Keywords: Hyper-chaotic system, Newton-downhill method, 6-sps mechanism, nonlinear equations.

1 Introduction

Kinematics analysis is fundamental for the parallel mechanism. Based on the kinematics analysis researcher can analyze velocity, acceleration, mechanism dynamics and mechanism synthesis. The process of kinematics analysis is attributed to find the solutions of the nonlinear equations and this process is very difficult. Analysis of forward kinematics for the 6-SPS platform is a difficult problem following the kinematics analysis of space 6R serial manipulator [1]. Many scholars have done much valuable research and work [1-7]. Raghavan investigated 960 paths by homotopy continuation method and firstly proved the 40 solutions of forward kinematics for 6-SPS platform [2], Sreenivasan etc [3] induced to 768 Homotopy paths through homogeneous method, in which the calculation task is huge. Liu etc [4] found all the 40 solutions of forward kinematics through quartic Homogeneous method and coefficient Homotopy continuation method. Wampler [5] proved we can get 40 solutions of forward kinematics by using Homotopy continuations method to track 40 paths but he did not give definite sample. Lee [6] solved forward kinematics of general 6-6 Stewart platform using algebraic elimination. Luo [7] has found forward kinematics of 6-SPS mechanism in the real numbers field, in which he considered Julia

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 1224–1229, 2009.

set point appears in the neighbor space of solution equations' Jacobian matrix determinant whose value is zero. But assumption is not proved. And more, to the multi-variables Jacobian matrix determinant, we must first find symbol expression, given other variable's values besides a variable, then find the chaotic fields of the un-given variable. The processing is very complex and difficult. Chaotic sequence method is a new method of using chaotic and hyper-chaotic system to generate initial value of Newton iterative method, we can find all real solutions [8] in the mechanism synthesis. Henon hyper-chaotic Newton iterative method can not solve the 6-SPS mechanism synthesis problem. When iterating divergence in adopting Newton and quasi-Newton method, we use mathematical program method. Chaotic and hyper-chaotic mathematical program can find all real solutions but the efficiency needs to improve [9].

Literature [10] investigated position symbol solutions of some kinds of unique 6-SPS parallel mechanism. By quaternion to describe the platform rotation we can avoid singularity in the formula and realize decoupling of position and stance of the platform. It provides a new method for the general 6-SPS mechanism. In the literature [4,7,9] the equations are created by Euler and universal transformation formula, singularity will happen, so it can not solve by Newton method. In this paper using cosine matrix method and link length constraint we establish nine variable equations of the general 6-SPS mechanism. Combined with Newton downhill method and hyper-chaotic method, it provides a new method to find the forward kinematics of the general 6-SPS mechanism.

2 Hénon Hyper-chaotic System

Lyapunov exponent is one of the efficient methods to describe the chaotic characteristic of nonlinear system. The number of Lyapunov exponent is equal to the state space dimension n of system. If one of Lyapunov exponent is positive, the system is chaotic; if at least two positive, the system is hyper-chaotic; the more number positive Lyapunov exponents exists the more unstable system it is. Generally more number of status parameters (such as high dimension system or for the discrete system, n>2), more unstable it probably is.

Literature [9] designed a general Hénon mapping:

$$\begin{Bmatrix} x_{1,k+1} \\ x_{i,k+1} \end{Bmatrix} = \begin{Bmatrix} a - x_{n-1,k}^2 - bx_{n,k} \\ x_{i-1,k} \end{Bmatrix} \tag{1}$$

Where $i = 2,3,\cdots,n$ is the number of system dimension; k is the discrete time; a and b are the adjusting parameters. When $i = 2$, it is the famous Hénon mapping. Literature [9] point out when given $a = 1.76, b = 0.1$, analyzing Lyapunov exponent in the system for $n > 2$, we found with the number n increased, the number n_1 of positive Lyapunov exponents has some relation with the system dimension n i.e. $n_1 = n - 1$.

3 Newton Downhill Method

To find solutions of nonlinear equations:

$$f(\mathbf{x}) = 0 \tag{2}$$

Where $\mathbf{x} = [x_1, x_2, \cdots, x_n]^{\mathbf{T}}$

Iterative formula of Newton downhill method is as follows:

$$\mathbf{x}_{n+1} = \mathbf{x}_n - \omega[\mathbf{F}'(\mathbf{x}_n)]^{-1}\mathbf{F}(\mathbf{x}_n) \tag{3}$$

Where, $0 < \omega \leq 1$.

When $\omega = 1$, it is Newton iterative method. It requires the value of ω to satisfy $\|\mathbf{F}(\mathbf{x}_{n+1})\| < \|\mathbf{F}(\mathbf{x}_n)\|$ for convergence, so take gradually halve method to confirm ω[11].

4 Hyper-chaotic Newton Down-Hill Method

The steps of finding all solutions of non-linear equations based on hyper-chaotic Newton downhill is as follows:

Step 1: Construct the chaotic set $\mathbf{x}_0(i+1, j)$ ($i = 1,2,\cdots,n$), $n+1$ is the variable number in the hyper-chaotic system, n is the number of positive Lyapunov exponent i.e., the variable number of solving equations, $j = 1,2,\cdots,N$, N is the length of chaotic set), finding $\mathbf{x}_0(i, j)$;

Step 2: Take the j[th] initial value $\mathbf{x}_0(:, j)$ as the initial value in Newton downhill method, after running iterative formula (3) j times, we can get all real solutions $\mathbf{x}^*$ in equations (2).

5 6-SPS Forward Kinematics of Parallel Mechanism

Figure1 is the general 6-SPS parallel robotic mechanism, when given every link length A_iB_i, forward kinematics is that find the up platform stance compared to down platform. The coordinate A_i in the fixed coordinate system is A_{ix}, A_{iy}, A_{iz}, the coordinate B_i in the dynamic coordinate system is B_{ix}, B_{iy}, B_{iz}, A_iB_i length is l_i, supposing the coordinate B_1 in the fixed coordinate system is (x_7, x_8, x_9), we can get transformation matrix which dynamic coordinate $X_2Y_2Z_2$ compared to the fixed coordinate $X_1Y_1Z_1$.

$$\mathbf{T}=\begin{bmatrix}\mathbf{R} & \mathbf{P}\\ 0 & 1\end{bmatrix}=\begin{bmatrix}l_x & m_x & n_x & x_7\\ l_y & m_y & n_y & x_8\\ l_z & m_z & n_z & x_9\\ 0 & 0 & 0 & 1\end{bmatrix} \tag{4}$$

Where R is 3×3 cosine matrix, every row is X_2, Y_2, Z_2 direction cosine in the coordinate $X_1Y_1Z_1$.

Let $\mathbf{x}=[x_1,x_2,x_3,x_4,x_5,x_6]=[l_x,l_y,l_z,m_x,m_y,m_z]$,According to the property of the cosine matrix, we have

$$\begin{cases}x_1^2+x_2^2+x_3^2=1\\ x_4^2+x_5^2+x_6^2=1\\ x_1x_4+x_2x_5+x_3x_6=0\end{cases} \tag{5}$$

$$n_x=x_2x_6-x_3x_5, n_y=x_3x_4-x_1x_6, n_z=x_1x_5-x_2x_4 \tag{6}$$

On the constraint of link length

$$\left\|\mathbf{T}[B_{ix}\quad B_{iy}\quad B_{iz}\quad 1]^{\mathbf{T}}-[A_{ix}\quad A_{iY}\quad A_{iz}\quad 1]^{\mathbf{T}}\right\|=l_i^2 (i=1,2,\cdots,6) \tag{7}$$

Substituting formula (4) and (6) into (7), with formula (5), they consist 9 variables forward kinematics equations of the general 6-SPS parallel robotic mechanism.

In Figure1, the coordinate A_i in the fixed coordinate $X_1Y_1Z_1$ are $A_1(0,0,0)$, $A_2(2,0,0)$, $A_3(4,1,0)$, $A_4(5,3,1)$, $A_5(3,5,-1)$, $A_6(-1,2,2)$, B_i in the dynamic coordinate system are $B_1(0,0,0)$, $B_2(1,0,0)$, $B_3(2,1,0)$, $B_4(4,3,2)$,

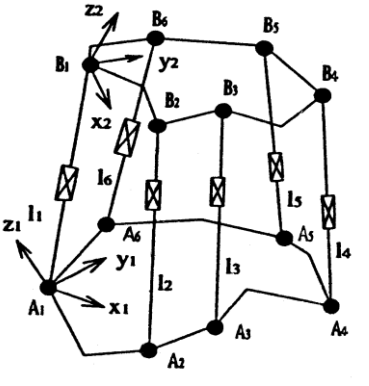


Fig. 1. The diagram of general 6-SPS parallel robot mechanism

$B_5(3,4,-1)$, $B_6(-2,-1,3)$, $l_1=11$, $l_2=12$, $l_3=13$, $l_4=15$, $l_5=14$, $l_6=10$, finding all the forward kinematics in the mechanism [9].

On the basis of Hénon hyper-chaotic mapping Newton downhill method, taking the random numbers as the initial points in hyper-chaotic set, the general Henon hyper-chaotic with the variables number $n_1=10$ in formula (1), generates the positive Lyapunov exponents hyper-chaotic sequences with variable number $n=9$ (Given $N=100$. if N is too large, the calculation time is too long and if N is too small, all real solutions can not be found, so in this kind of mechanism synthesis, we firstly use less value in the trying process and then use larger value, usually adopting $(5-10)\times$ variables number), by iterative process in 1.28 seconds, we can get all real solutions in formula (8) as Table 1. The corresponding initial values for the four solutions are shown as Table 2 and the corresponding four solutions for Euler angle are also shown as Table 3 which is the same to literature [9], but Newton method can not solve the equation (8).

Table 1. Results of Synthesis

No	l_x	l_y	l_z	m_x	m_y	m_z	x_7	x_8	x_9
1	-0.6539	0.2181	0.7245	-0.6314	-0.6849	-0.3637	0.2301	3.8573	10.2989
2	-0.5958	0.4756	0.6472	-0.6076	-0.7939	0.0240	0.3682	5.3866	9.5838
3	-0.5410	-0.7491	-0.3823	0.6784	-0.6574	0.3280	-1.8184	1.1055	-10.7922
4	0.0291	-0.8060	-0.5912	0.9820	-0.0872	0.1673	-1.1889	-0.3137	-10.9311

Table 2. Corresponding initial values for the four solutions

o	$\mathbf{x}_0(1)$	$\mathbf{x}_0(2)$	$\mathbf{x}_0(3)$	$\mathbf{x}_0(4)$	$\mathbf{x}_0(5)$	$\mathbf{x}_0(6)$	$\mathbf{x}_0(7)$	$\mathbf{x}_0(8)$	$\mathbf{x}_0(9)$
	1.7151	0.6063	1.3626	-1.0638	-0.7199	1.5724	1.4914	1.8983	1.6809
2	1.4026	0.6350	-0.3936	-1.3483	-0.6499	1.1813	1.6878	-0.3005	0.4151
	1.3717	-1.2466	-0.7422	-1.8116	1.6029	-1.1244	1.7612	1.6092	-1.2500
	1.0629	-0.3340	-0.7141	0.9300	0.9689	0.8010	1.6996	1.5346	0.5182

Table 3. Corresponding four solutions for Euler angle

No.	x_1	x_2	x_3	x_4	x_5	x_6
1	0.2768	0.5113	-0.6170	0.2301	3.8573	10.2989
2	0.5116	0.3675	0.9636	0.3682	5.3866	9.5838
3	-0.8487	0.2703	-0.4596	-1.8184	1.1055	-10.7922
4	-0.1554	0.3434	-0.7563	-1.1889	-0.3137	-10.9311

6 Conclusions

Using coordinate transformation cosine matrix and link length constraint the author creates the general 6-SPS equations with 9 variables. Combined Newton downhill method and hyper-chaotic system, taking the hyper-chaotic serials generated by hyper-chaotic system as the initial value in Newton downhill method, the paper provides the method of finding all real solutions in nonlinear equations which is based on hyper-chaotic mapping Newton downhill method and give steps. The method solves the divergence in the Newton iterative and quasi-Newton method. The sample of forward kinematics of the general 6-SPS platform mechanism proves it is correct and effective.

Acknowledgments. This research is supported by the grant of the 11th Five-Year Plan for the construct program of the key discipline (Mechanical Design and Theory) in Hunan province (XJT2006180), National Science Foundation of China (50845038).

References

1. Wen, F.A., Liang, C.G., Liao, Q.Z.: The forward displacement analysis of parallel robotic mechanisms. China Mechanical Engineering 10(9), 1011–1013 (1999) (in Chinese)
2. Raghavan, M.: The Stewart platform of general geometry has 40 configurations. ASME J. Mech. Des. 1l5(1), 227–282 (1993)
3. Sreenivasans Nanua, P.: Solution of the direct position kinematics problem of the general Stewart platform using advanced polynomial continuation. In: The 22nd Biennial Mechanisms Conference, Scottsdale, Scottsdale, pp. 99–106. ASME, New York (1992)
4. Liu, A.X., Yang, T.L.: Finding all solutions to forward displacement analysis problem of 6-SPS parallel robot mechanism. Chinese Mechanical Science and Technology l5(4), 543–546 (1996) (in Chinese)
5. Wampler, C.W.: Forward displacement analysis of general six-in parallel SPS (Stewart) platform manipulators using SOMA coordinates. Mechanism and Machine Theory 31(3), 331–337 (1996)
6. Lee, T.Y., Shim, J.K.: Forward kinematics of the general 6-6 Stewart platform using algebraic elimination. Mechanism and Machine Theory 36, 1073–1085 (2001)
7. Luo, Y.X., Li, D.Z.: Finding all solutions to forward displacement analysis problem of 6-SPS parallel robot mechanism with chaos-iteration method. Journal of Chinese Engineering Design 10(2), 95–101 (2003) (in Chinese)
8. Luo, Y.X., Li, D.Z., Fan, X.F., et al.: Hyper-chaotic Mapping Newton Iterative Method to Mechanism Synthesis. Journal of Mechanical engineering 54(5), 372–378 (2008)
9. Luo, Y.X.: Hyper-chaotic Mathematical Programming Method and its Application to Mechanism Synthesis of Parallel Robot. Transactions of the Chinese Society for Agricultural Machinery 39(5), 133–136 (2008) (in Chinese)
10. Li, L.Y., Wu, H.T.: A symbolic solution of forward kinematics analysis of a 6-SPS parallel mechanism. Journal of Yangzhou University 9(3), 42–45 (2006) (in Chinese)
11. Cun, G., Wang, Z.L.: MATLAB program assembly. Electronic Industrial Press, Beijing (2008) (in Chinese)

Inverted Pendulum System Control by Using Modified Iterative Learning Control[1]

Hongxia Gao[1], Yun Lu[2], Qian Mai[1], and Yueming Hu[1]

[1] Engineering Research Center for Precision
Electronic Manufacturing Equipments of Ministry of Education,
College of Automation, South China University of Technology, Guangzhou, China
[2] South Building, No.5 Keyuan Road, Central Zone, Science & Technology Park,
Nanshan District, Shenzhen, Guangdong, P.R. China
hxgao@scut.edu.cn, romantikc@qq.com, mqpig@163.com,
auymhu@scut.edu.cn

Abstract. The inverted pendulum is a typical multi-variable and non-linear system. It is hard to improve its performance because of modeling difficulty, parametric uncertainties during actual system operation. Iterative Learning Control (ILC) can achieve perfect tracking or performance when there is model uncertainty or when we have a "blind" system. This paper proposed a modified iterative learning control method to improve the dynamics of the car with an inverted pendulum. Experiment results demonstrated that the proposed method can improve the dynamic performance of the car with an inverted pendulum.

Keywords: Iterative Learning Control, Inverted Pendulum System.

1 Introduction

The inverted pendulum is a typical multi-variable and non-linear system, which is seriously instable and strong coupling. It can effectively reflect the key points of the control problem, such as non-linear problem, the robustness problem, the dynamic problem, the stabilization problem and tracking problem. The experiences can be applied to the actual projects. For example, the experiment results of inverted pendulum can afford good reference for the balance problem of the walking robot, and the verticality control problem of the rocket. So, there are many researches interested in the control of the inverted pendulum and varieties of new control method are applied to achieve perfect performance of it.

Iterative Learning Control (ILC), as an intelligent control methodology, is able to achieve perfect tracking or performance when there is model uncertainty or when we have a "blind" system. So, it is effective to control the cart with inverted pendulum which is non-linear.

Based on in-depth analysis on modeling the angle control problem of the car with an inverted pendulum, this paper proposed a modified Iterative Learning Control

[1] This work was supported by National Natural Science Foundation of China No. 60835001.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 1230–1236, 2009.

method to improve the dynamics of inverted pendulum. Experiment results demonstrated that the proposed method can improve dynamic performance of the inverted pendulum.

2 Inverted Pendulum Modeling[1]

2.1 Problem Setup and Design Requirement

The car with an inverted pendulum, shown below, will be controlled by an impulse force F. It is need to determine the dynamic equations of motion for the system, linearize about the pendulum's angle θ, and find a controller to satisfy all of the design requirements given below.

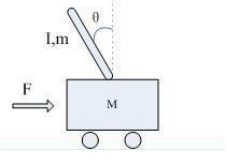


Fig. 1. The Cart with an Inverted Pendulum

First, let's assume that

M	mass of the cart=1.32kg	I	inertia of the pendulum
m	mass of the pendulum=0.07kg	F	force applied to the cart
b	friction of the cart=1N/m/se	x	cart position coordinate
l	length to pendulum center of mass=0.2m	θ	pendulum angle from vertical

In this paper, we will be only interested in the control of the pendulum position. This is because the techniques used in the paper can only be applied for a single-input-single-output system. Therefore, none of the design criteria deal with the cart's position. We will assume that the system starts at equilibrium.

So, the design requirements for this system can be:

- Settling time of the pendulum less than 5 seconds.
- Pendulum angle never more than 0.05 radians from the vertical.

2.2 Force Analysis and System Equations

Below are the two Free Body Diagrams of the system.

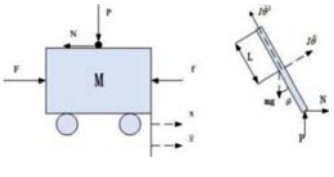


Fig. 2. Free Body Diagrams of the System

Analyzing the forces in the Free Body Diagram of the cart and of the pendulum you can get the following equations of motion[1]:

$$(M+m)\ddot{x} + b\dot{x} + ml\ddot{\theta}\cos\theta - ml\dot{\theta}^2\sin\theta = F \tag{1}$$

$$(I+ml^2)\ddot{\theta} + mgl\sin\theta = -ml\ddot{x}\cos\theta \tag{2}$$

Since Matlab can only work with linear functions, this set of equations should be linearized. After linearization the two equations of motion become (where u represents the input):

$$\begin{cases} (I+ml^2)\ddot{\phi} - mgl\phi = ml\ddot{x} \\ (M+m)\ddot{x} + b\dot{x} - ml\ddot{\phi} = u \end{cases} \tag{3}$$

After the Laplace transform and a little algebra, the linearized system equations can also be represented in the following state-space form:

$$\begin{bmatrix} \dot{\phi} \\ \ddot{\phi} \\ \dot{x} \\ \ddot{x} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{mgl(M+m)}{I(M+m)+Mml^2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ \frac{m^2gl^2}{I(M+m)+Mml^2} & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \phi \\ \dot{\phi} \\ x \\ \dot{x} \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{ml}{I(M+m)+Mml^2} \\ 0 \\ \frac{(I+ml^2)}{I(M+m)+Mml^2} \end{bmatrix} \tag{4}$$

$$y = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \phi \\ \dot{\phi} \\ x \\ \dot{x} \end{bmatrix} + [0]u \tag{5}$$

3 Iterative Learning Control

ILC is a "... controller that learns to produce zero tracking error during repetitions of a command, or learns to eliminate the effects of a repeating disturbance on a control system output" [2].

In many definitions about ILC, almost all of them stressed the importance of "repetition". Learning through a predetermined hardware repetition is the key idea of ILC [3]. Considering a fixed-length input signal is applied to a system in an initial state, the system is returned to its initial state after the complete input has been applied, and the output that resulted from the applied input is compared to a desired target. The error is used to construct a new input signal to be applied for the next time the system operates. This process is then repeated.

In this paper, we consider to apply the general "PID-type" ILC algorithms [4, 5]. Its basic idea is illustrated in Figure 3.

Here,

$$u_{k+1}(t) = u_k(t) + \Gamma_p e_{k+1}(t) + \Gamma_i \int_0^t e_{k+1}(\tau)\,d\tau + \Gamma_d \dot{e}_{k+1}(t) \tag{6}$$

Where $u_{k+1}(t)$ is the input of the system and $e_{k+1}(t) = y_d(t) - y_{k+1}(t)$ is the error on trial $K+1$, with $y_{k+1}(t)$ the system's output and $y_d(t)$ the desired response.

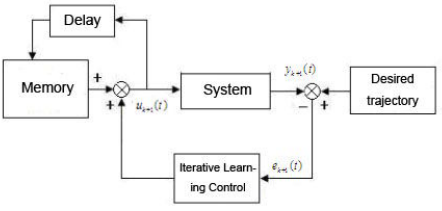


Fig. 3. Basic Idea of "PID-type" ILC

4 Simulation

4.1 Iterative Learning Control Simulation and Its Modification

The inverted pendulum's State-space form has been obtained. According to "PID-type" ILC theory, the simulation is shown as Fig 4. Assume that the system starts at equilibrium, so that theta $\theta = 0°$. Disturbance that applied to the cart can be triggered by a Pulse Generator. In Fig.4, the parameters *rk* represents the Force F. The parameters *rk*1 represents the stored data. The output that resulted from the applied input will be compared with the desired output. After that, the error will send to ILC. After studying, the result will be stored in MATLAB workspace in order to be called [6, 7].

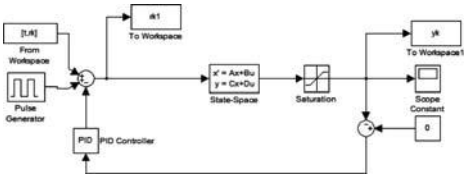


Fig. 4. Iterative Learning Control Simulation

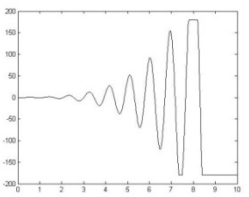
First, we will arbitrarily assign values to three PID parameters of ILC controller. For example, let's assume that the parameters of "PID-type" ILC controller are given in Table 1.

Table 1. The Parameters of "PID-type" ILC Controller

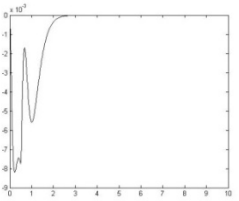
Type of ILC controller	Γ_P	Γ_i	Γ_d
PI	30	8	0
PD	30	0	5
PID	25	5	8

So, we can get the output trajectories of the system as shown in Fig. 5.

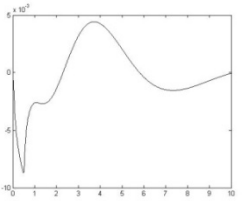
From the three plots, the system outputs with PI and PID-type ILC controller are not satisfactory. Especially PI-type ILC controller, it may lead the system to unstable. Above all, PD-type ILC controller is better than the others at both speed and accuracy. So, we will select PD-type ILC Controller to improve the system output.



PI Controller



PD Controller



PID Controller

Fig. 5. The Output Trajectories of the system with PI, PD and PID-type ILC Controller

After many times' experiments, when the parameters Γ_P and Γ_d is 40 and 15, we can achieve a rather good effect at speed and accuracy. Meanwhile, the result also meets the design requirements. The output trajectory is shown as Fig 6.

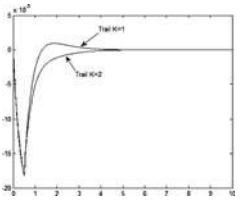


Fig. 6. The System Output with PD-type ILC Controller

From Fig. 6, we can find that the actual output is already close to the desired target when the trail K is 2. It means that PD-type ILC controller can improve the output performance of the car with inverted pendulum efficiently.

4.2 Performance Contrast between ILC and Full State Feedback Control

To illustrate the benefits of ILC controller, we will also apply the full state feedback control method [8] to demonstrate it.

Now we have already known the state-space form, it is easy to get the open-loop poles P of the system: P=[0, 0, 6.18, -6.18].

Based on the result of LQR design: the state feedback matrix K=[20.2846 2.4316 0.9099 1.9018]], the system output with full state feedback is shown as Fig. 7.

Obviously, the system output with full state feedback can't meet the design requirements. The pendulum's overshoot is too large and the settling time isn't satisfactory too. The performance contrast between the system with ILC and full state feedback

control is given in Table 2 below. From it, we can see the performance of system output with ILC controller is much better than it with full state feedback.

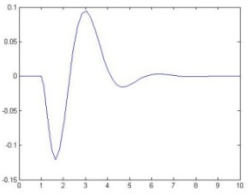


Fig. 7. The System Output with Full State Feedback

Table 2. The Performance Contrast between ILC and Full State Feedback Control

Type	The maximum offset angel	The settling time
Full State Feedback	0.12°	5s
ILC	0.018°	2.3s

5 Conclusion

The control problem of the car with inverted pendulum is among the typical control difficulties. Unmodeled dynamics and parametric uncertainties make it impossible to achieve desired performance. In this paper, the proposed control method based on ILC improved the speed and accuracy of the one-stage inverted pendulum when tracking the desired trajectory. And the experiences can be extended to the real objects, such as the rocket verticality control, etc. The future work is ILC' application on the second-stage or three-stage inverted pendulum.

Acknowledgments. I would like to express my sincere gratitude to Yachen Zhang, Wei Wei and Jinlong Cai for their support.

References

1. Bill, M., Dawn, T.: Control Tutorials for Matlab, `http://www.engin.umich.edu/group/ctm/examples/pend/invpen.html`
2. Phan, M.Q., Longman, R.W., Moore, K.L.: Unified Formulation of Linear Iterative Learning Control. In: Proceedings of 2000 AAS/AIAA Space Flight Mechanics Meeting, Florida, pp.AAA 00–106 (2000)
3. Ahn, H.-S., Moore, K.L., Chen, Y.: Iterative Learning Control: Robustness and Monotonic Convergence for Interval Systems. Springer, London (2007)
4. Arimoto, S., Kawamura, S.F.: Miyazaki: Bettering Operation of Dynamic Systems by Learning: A New Control Theory for Servomechanism or Mechatronic Systems. In: Proceedings of 23rd Conference on Decision and Control, Las Vegas, Nevada, pp. 1064–1069 (1984)

5. Arimoto, S., Kawamura, S., Miyazaki, F.: Bettering Operation of Robots by Learning. J. of Robotic Systems 1(2), 123–140 (1984)
6. Guan, W.X., Chen, W., Luo, J.Q.: Model Building, Simulation and Real-time Control of an Inverted Pendulum. Journal of Shenzhen Institute of Information Technology 5(2), 67–69, 89 (2007)
7. He, F.P., Dai, S.W.: Modeling and Simulation of Nonlinear Control System in MATLAB Environment. Information Technology 9, 94–96 (2007)
8. Ding, H.G., Zhao, J.Y., Zhang, D.S.: Study on Closed-loop Control System of Car-inverted Pendulum Based on LQR. Coal Mine Machinery 29(11), 124–126 (2008)

Dynamic Simulation of Passive Walker Based on Virtual Gravity Theory

Heng Cao, Yu Wang, Jun Zhu, and Zhengyang Ling

School of Mechanical and Power Engineering,
East China University of Science and Technology
200237 Shanghai, China
hengcao@ecust.edu.cn

Abstract. Passive dynamic walking is a high efficiency way for bipedal robots to reduce energy cost. To study the nature of passive dynamic walking, we propose a new dynamic simulation method and build a virtual passive walker prototype based on virtual gravity theory using ADAMS. The results of the simulation are believable by comparing with the former research results, and the influences on gait stability of three critical physical parameters are obtained by using the parameterization analysis tools. These simulation results show that the suitable initial parameter is a critical factor for a simplest passive walker's stability on slope ground and the dynamic simulation is an available way to study the bipedal robot's walking.

Keywords: Passive walker, virtual gravity, dynamic simulation, ADAMS.

1 Introduction

Since McGeer raised the method of passive dynamic walking [1, 2], many researchers used this way to increase the bipedal robots' walking efficiency [3]. These robots use less energy and less control than other traditional powered robots, but their gait is more natural.

Many researchers studied the various bipedal passive walking models by numerical simulation [4, 5] and experimental study [6, 7], because the highly nonlinear robot system is difficult to be solved by usual mathematics tools. However, the numerical simulation is usually to code a large number of programs, meanwhile the experimental method uses a lot of time and materials to prepare and operate. ADAMS is a multi-rigid body dynamic analysis software, which is based on Lagrange equation of motion [8]. That studying the nature of passive dynamic walker by ADAMS is a more simple way than numerical simulation and experimental study.

In this paper, we will build a general planar model, composed of straight legs and arc-feet, walking stably on slope based on the virtual gravity theory using ADAMS. The main results of simulation are discussed and the influences on gait stability of three critical physical parameters are obtained by using the parameterization analysis tools.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 1237–1245, 2009.

2 Model Analysis

Fig.1(a) shows a simple planar passive dynamic walker composed of straight legs and arc-feet walking on a small downhill angle slope. The model is symmetrical, which means the rigid legs (note the rigid legs 1 and 2 respectively) have the same physical parameters. Assume that a pair of rigid legs with arc-feet interconnected through a frictionless hinge in hip and there is no slip between the foot of the stance leg and the ground. The impact between the swing leg and the ground is perfectly inelastic impact and each exchange between the stance leg and the swing leg is instantaneous. The important parameters of this model are listed in Table 1.

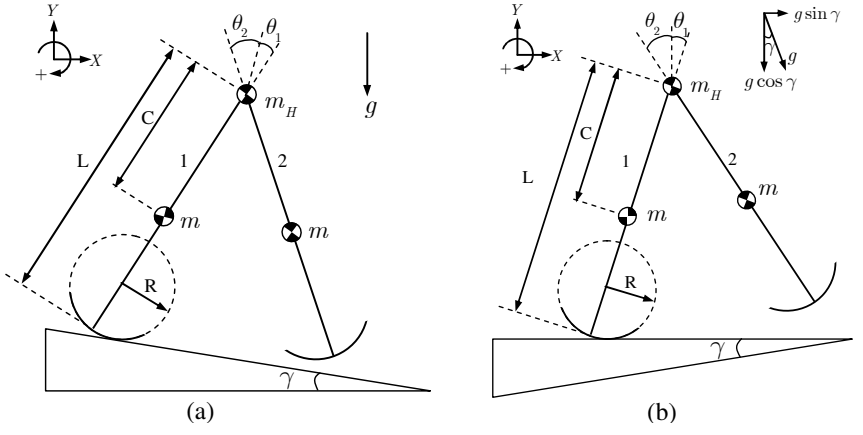


Fig. 1. (a) A simpler planar passive dynamic walker composed of straight legs and arc-feet is walking on a small downhill angle slope; (b) the state of changed model which is equivalent with the former model based on virtual gravity theory.

Table 1. Parameters of the virtual passive walker prototype

Symbol	Description	Value	Unit
L	Length of leg	0.1	m
R	Radius of arc-feet	0.033	m
C	Distance of hip and leg's center of mass	0.05	m
m_H	Mass of hip	10	Kg
m	Mass of leg	0.01	Kg
g	Gravity acceleration	9.810	m/sec^2
$\theta_{10}(\theta_{20})$	Initial angle between the leg 1(2) and the slope normal	11.5	Degree
γ	Angle of slope	1	Degree

Let $\theta \quad [\theta_1, \theta_2]^T$ is the generalized coordinate vector of the walker system, where $\theta_1(\theta_2)$ represents the real-time angle between the leg 1(2) and the slope normal. The motion of the walker can be classified as the swing stage and the transition stage. During the swing stage, the dynamic equation of the system can be obtained by Lagrange equation:

$$M(\theta)\ddot{\theta} + C(\theta,\dot{\theta})\dot{\theta} + G(\theta) = 0. \quad (1)$$

$M(\theta)$ is a 2×2 inertia matrix, $C(\theta,\dot{\theta})$ is a 2×2 matrix with the coriolis and centrifugal coefficients, $G(\theta)$ is a 2×1 vector of gravitational torques to the generalized coordinates θ in Eq.(1).

During the transition stage, two events happen simultaneously: the former swing leg hits the ground transferring to the latter stance leg; the former stance leg leaves the ground transferring to the latter swing leg. As implied in the assumptions above, the walker's angular momentum is conserved during the collision. So we can get the transition equation as follows:

$$Q(\theta^+)\dot{\theta}^+ = P(\theta^-)\dot{\theta}^- . \quad (2)$$

$Q(\theta^+)$ and $P(\theta^-)$ are both 2×2 matrixes with the angular momentum coefficients of the walker after and before the impact with ground respectively, $\dot{\theta}^+$ and $\dot{\theta}^-$ are the angular velocities just after and before the transition in Eq.(2).

The details about Eq.(1), Eq.(2) and further more detail information can refer to [9].

Passive walker has a limitation that it only walks on a slope under the effect of gravity. The existence of slope is not only a difficulty to modeling in ADAMS, but also an important influencing factor of the walking stability. Taken care of the two aspects, the model shown in Fig.1(a) must be reasonably simplified.

Asano et al. proposed the virtual gravity field theory to explain the principle of quasi-passive dynamic walking on flat ground. They viewed the active driving force and gravity working on the bipedal robot as a certain direction virtual gravity field, then the quasi-passive dynamic robot walking on flat ground has a similar characteristic of the pure passive dynamic robot walking on a slope [5]. Inspired with this study method, we make two changes in the model shown as in Fig.1(a): transferring walking on a slope to walking on flat ground, and transferring the vertical gravity to the virtual gravity with horizontal and vertical components. The state of changed model is shown in Fig.1(b). After changing, the changed model's state is equivalent with the former model's state because it just counterclockwise rotates the coordinate around the origin.

3 Dynamic Simulation and Results Discussion

Fig.2. shows a virtual planar straight leg with arc-feet prototype modeled in ADAMS by using the modeling tool and the simplified model as shown in Fig.1(b). The simplified virtual prototype is composed of two rigid links with arcs as legs with feet, one column as hip, and a box with some rectangular planes as a walking flat ground with some discrete steps. These parts are assembled by various joint restrictions such as revolute joint, fixed joint, parallel joint. The main joint restrictions of the virtual prototype are listed in Table 2. To avoid the kneeless model's problem of "foot clearance" occurring at mid stance, several contact restrictions between the arc-feet and ground steps are added.

As the transformation methods mentioned above, the magnitude of the virtual gravity components of X direction is set to $g*\sin\gamma$ and the magnitude of Y direction is set to $-g*\cos\gamma$, which is shown in Fig.2. And the units' environment of the simulation is set to MMKS as shown in Fig.2.

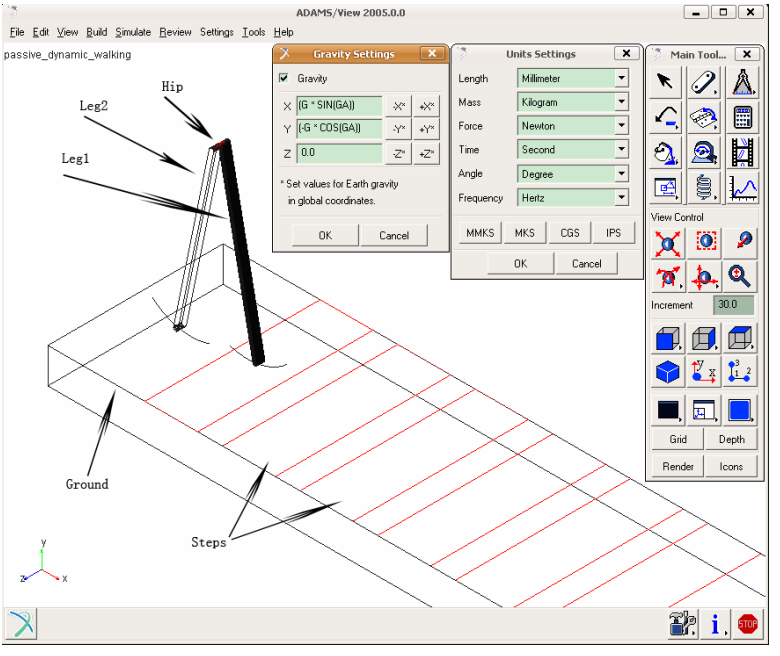


Fig. 2. Virtual planar straight leg with arc-feet prototype in ADAMS

Table 2. Main joint restrictions of the virtual prototype

Joint name	Part1	Part2	Joint restriction
Jprim1	Leg1	Ground	Parallel joint
Joint1	Leg1	Hip	Revolute joint
Joint2	Leg2	Hip	Revolute joint
Joint3	Steps	Ground	Fixed joint

3.1 Dynamic Simulation Analysis

After building the prototype and setting simulation parameters, a dynamic simulation can be started. The time of simulation is set to 1.6s, and the step size of simulation is set to 0.01s. All of these modeling tools and simulation settings are embed in the main toolbox shown as in Fig.2.

Fig.3 shows the graphical results of the dynamic simulation of this virtual prototype. It can be seen that the model can walk stably along the flat ground.

ADAMS provides measurement tools to view the changes of some important data such as the angle velocity, angle acceleration, constraint force and torque in the dynamic simulation. And the tools also can plot arbitrary mixed curves based on the simulation data. The quality of the simulation walking can be tested by analyzing the measurements of the system. Further more, the credibility of the dynamic simulation can be verified by compared with the former recognized numerical simulation data.

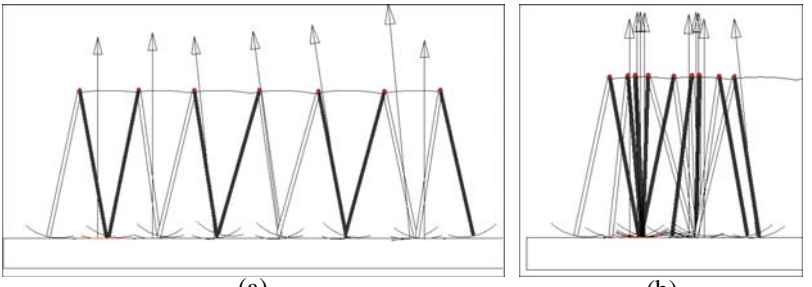


Fig. 3. Dynamic simulation of passive walker. (a) The state of walker in every instantaneous of leg exchange; (b) the continuous change of legs in two successive cycle gaits. The solid line near to the hip is the trace of hip's motion. The arrows are the contact forces between arc-feet and ground, whose length and arrowheads show the magnitude and the direction of the contact forces respectively.

Results of the Leg's Angle. The system's generalized coordinate $\theta \ [\theta_1, \theta_2]^T$ is an important data to weigh the stability of system. To distinguish two curves of θ_1, θ_2 in the coordinate system, we use $\theta' \ [\theta_3, \theta_1]^T$ to replace $\theta \ [\theta_1, \theta_2]^T$, where $\theta_3 \ (\theta_1+\theta_2$ represents the included angle of two legs.

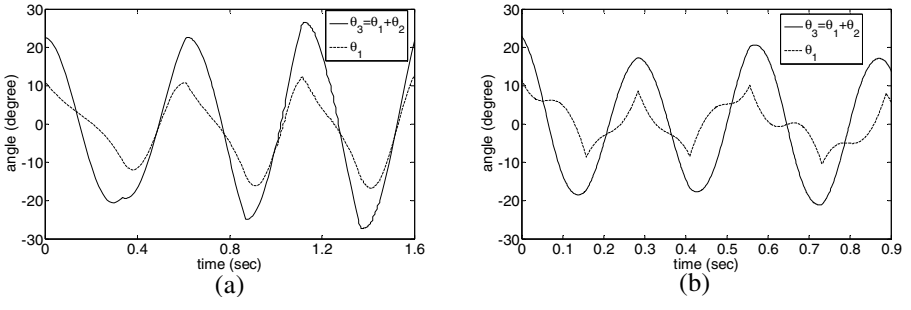


Fig. 4. (a) The changes of $\theta' = [\theta_3$ *(solid curve),* θ_1 *(dotted curve)]*^{T}. (b) the data calculated by numerical simulation at the same parameters.

Fig.4(a) shows the change of $\theta' = [\theta_3, \theta_1]^T$ with time in the dynamic simulation. It is concluded that the gait is basically periodicity and stability. Fig.4(b) shows the same data with the same parameters of passive walker calculated by the numerical simulation method mentioned in [10]. Compared with these curves in Fig.4(a) and Fig.4(b), we conclude that the method of dynamic simulation using ADAMS is available and correct. Fig.5 shows the phase diagram of leg1 based on the dynamics simulation results, which is similar to the phase diagram of numerical simulation.

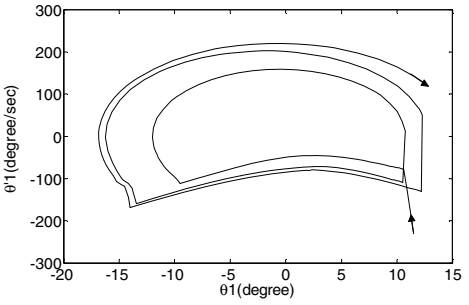


Fig. 5. The phase diagram of leg1 based on the dynamic simulation results

Results of the Hip's Energy. In the passive gaits, there is only gravity applying work to the system, so the mechanical energy of system is conservation during the swing stage. The angular momentum of system is conservation in the collision of swing leg and ground during the transition stage, but it loses part of the system's kinetic energy. Fig.6 shows the changes of the hip's kinetic energy, potential energy and total energy.

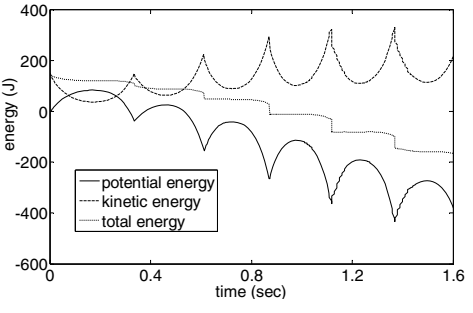


Fig. 6. The changes of hip's kinetic energy (*dotted curve*), potential energy (*solid curve*) and total energy (*point curve*)

Fig.6 shows that the hip's kinetic energy first decreases and then increases while the potential energy first increases and then decreases in a step. The hip's total energy has little changes in a step, but decreases about 0.045 J in the instantaneous of every collision. The lost energy mainly caused by the perfectly inelastic collision. If the system is duly added the lost energy on flat ground, the walker will never stop. That is the principle of quasi-passive dynamic walking and the essence of virtual gravity field theory.

Results of the Contact Force. The contact force is also an important indicator of walking stability and continuity. Fig.7 shows the changes of the vertical component of the contact force between arc-feet and ground. That the force in every step is stable and continuous, except for the pulse force in every collision, means the walking is stable and continuous.

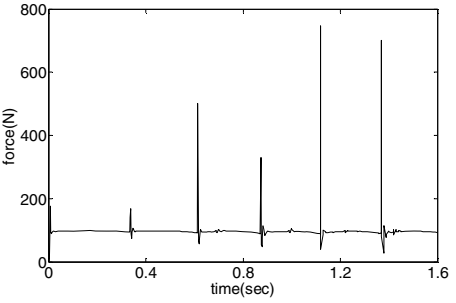


Fig. 7. The changes of the vertical component of the contact force

3.2 Parameterization Simulation Analysis

ADAMS has a useful function that it can analyze the changes of walking simulation when some design parameters change. In the process of analysis, ADAMS takes various design parameter values to simulate a series of analysis and return the simulate results. We take three parameters to find the effect to system: slope angle γ, distance of hip and leg's center of mass *C* and radius of arc-feet *R*.

We choose the changes of leg's included angle θ_3 in simulation analysis to observe the effects of walking, for it is an important data to weigh the stability of the system. To analyze the influence on gait stability of one parameter, we only change a parameter and fix other parameters in a parameterization simulation analysis process. Fig.8 shows the changes of leg's included angle θ_3 under the different of slope angle γ, distance of hip and leg's centre of mass *C* and radius of arc-feet *R* respectively.

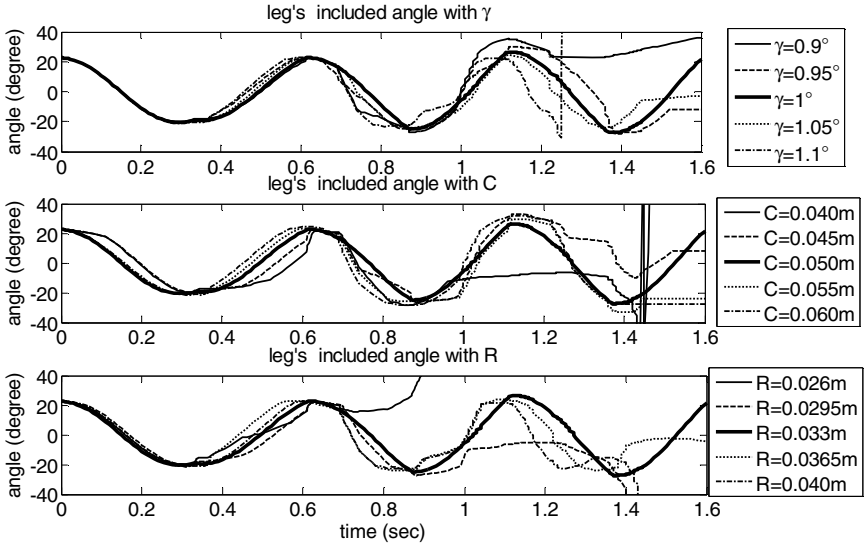


Fig. 8. The changes of leg's included angle θ_3 under the different of slope angle γ, distance of hip and leg's centre of mass *C* and radius of arc-feet *R* respectively

From the results of parameterization simulation analysis as shown in Fig.8, we can obtain three conclusions as follows.

1. Slope angle γ is one of the critical factors on walking stability. Smaller or larger slope angle will reduce the walking stability, and the larger slope angle the faster walking velocity.

2. That the leg's center of mass is nearer to hip will make the walking lose stability earlier. In contrast, that the leg's center of mass is nearer to foot will remain walking, but walking slower than normal.

3. Radius of arc-feet R is another critical factor on walking stability. The smaller radius will let walking access into instability earlier. But it has no effect on walking velocity and step length.

4 Conclusions and Future Work

ADAMS provides a friendly, visualization and parameterized simulation environment, which is much simpler than numerical simulation and experiment trial. We build a simple planar straight leg with arc-feet passive walker model based on virtual gravity theory using ADAMS, and obtain a dynamic walking simulation. The results of dynamic simulation are shown by graphics and curves, and we can easily get the real-time kinematics and dynamics data about each rigid body such as the angle displacement, angle velocity, angle acceleration, constraint force and torque. The results of dynamic simulation are believable by comparing with the recognized research data and the influence on gait stability of three critical physical parameters is obtained by using the parameterization analysis tools. From these analyses, we can conclude that the suitable initial parameter is a critical factor to a simplest passive walker's stability on slope ground.

The dynamic simulation of a planar straight leg with arc-feet passive walker model has been realized, but our ultimate goal is to achieve the three-dimensional humanoid passive walker model's simulation. It has a lot of work to do for this.

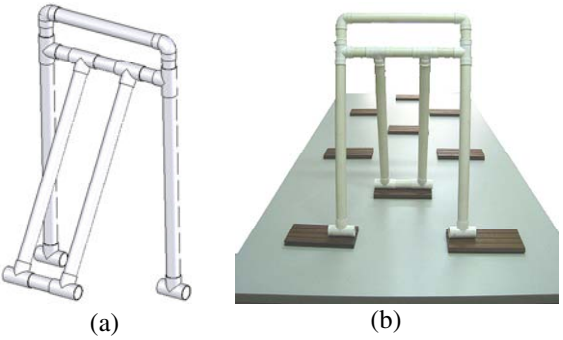


Fig. 9. (a) An ideal design model of passive walker; (b) the real prototype mainly assembled by pipe materials

The dynamic simulation of system is a useful and easier way to study the bipedal robot's gait, but the simulation results also should be verified by some little experimental test. Fig.9 shows the simple four straight legs passive walker prototype inspired by McGeer's passive dynamic walker. The real prototype as shown in Fig.9(b) is composed of different shapes of PVC-U pipe materials, a steel spindle and four rolling bearings. The real prototype can walk down along the particular angle slope if the swing legs have correct initial angle velocity. The prototype has been accomplished, but the experimental data which will compare with the results of simulation still need further testing.

Acknowledgments. This work was supported by National Natural Science Foundation of China under Grant No.50775072.

References

1. McGeer, T.: Passive Dynamic Walking. The International Journal of Robotics Research 9, 62–82 (1990)
2. McGeer, T.: Passive Walking with Knees. In: Proceedings of the 1990 IEEE International Conference on Robotics and Automation, pp. 1640–1645. IEEE Press, New York (1990)
3. Collins, S., Ruina, A., Tedrake, R., Wisse, M.: Efficient Bipedal Robots Based on Passive-Dynamic Walkers. Science 307, 1082–1085 (2005)
4. Goswami, A., Thuilot, B., Espiau, B.: A Study of the Passive Gait of a Compass-Like Biped Robot: Symmetry and Chaos. The International Journal of Robotics Research 17, 1282–1301 (1998)
5. Asano, F., Yamakita, M., Kamamichi, N., Luo, Z.W.: A Novel Gait Generation for Biped Walking Robots Based on Mechanical Energy Constraint. IEEE Transactions on Robotic and Automation 20, 565–573 (2004)
6. Wu, Q., Sabet, N.: An Experimental Study of Passive Dynamic Walking. Robotica 22, 251–262 (2004)
7. Wisse, M.: Essentials of Dynamic Walking: Analysis and Design of Two-Legged Robots. Ph.D. Thesis, TU Delft, Netherlands (2004)
8. MSC.Software Corporation, `http://www.mscsoftware.com/products/mdadams.cfm`
9. Goswami, A., Espiau, B., Keramane, A.: Limit Cycles in a Passive Compass Gait Biped and Passivity-Mimicking Control Laws. Autonomous Robots 4, 273–286 (1997)
10. Wisse, M., Schwab, A.L.: First Steps in Passive Dynamic Walking. Climbing and Walking Robots, 745–756 (2005)

A New Strategy of the Robot Assembly Motion Planning*,**

Yanchun Xia[1], Yuehong Yin[2], Yuewei Bai[1], and Yafei He[1]

[1] School of Mechanical & Electronic Engineering, Shanghai Second Polytechnic University, No. 2360 Jinhai Road Pudong New Area Shanghai, Shanghai, P.R. China
mollyxia@163.com
[2] Research Institute of Robotics, Shanghai Jiaotong University, Shanghai, PRC

Abstract. A motion planning strategy, a dissymmetrical T-shaped peg into a C-shaped slot, is proposed based on medial axis diagram knowledge and force control technology. It is very important to understand the assembly environment in the process of the active assembly operation. The assembly environment can be understood based on the analysis of the geometrical and force conditions. The medial axis diagram is used to analyze the geometric constraint relations of the parts during the robot assembly, and by which, the whole process of the complex assembly operation can be planned. As a result, the robot can understand the almost environment information based on the planning knowledge. Due to the uncertainties in the assembly process, the force sensing and force control are executed for apperceiving the precise physical relations. Combining the medial axis diagram knowledge-based with force sensing and force control strategy, the assembly task of T-shaped peg into C-shaped slot can be completed high efficiently, and it can be used for references to the study of the robot assembly.

Keywords: Coarse and fine motion planning, Environment understanding, Force control, Medial axis diagram, Contact constrain.

1 Introduction

The study of motion planning strategy is the key issue in robot assembly planning. In order to automatically execute robot motion planning, a key problem of transferring a given sequence of assembly operations into feasible motion plans must be solved. Most researches are focused to the following aspects: 1) gross motion planning[1], 2) grasp planning[2], and 3) motion planning[3].

Robot assembly is a constraint manipulation[4], which is a kind of compliance motion under force control essentially[5]. It is a very difficult task, because it needs to

* Supported by Leading Academic Discipline Project of Shanghai Municipal Education Commission, Project Number: J51802.

** Supported by Shanghai Municipal Education Commission Fund (EGD08004) and School Fund of Shanghai Second Polytechnic University (QD2008009).

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 1246–1254, 2009.

study and understand environment fully by surface tracing, based on which the strategy of control can be carried on well. The geometric approach is one of the trends in assembly planning. Krishnan and Sanderson[6] introduced an algebra of polyhedral cones which provides a tool for combining geometric constraints from different part relations. Claudio Mirolo and Pagello[7] defined a spatial reasoning system based on geometric modeling. We are aware that pure geometry is not sufficient to approach the whole robot motion planning, and that major progress will follow a better integration of the sensory feedback in robot control. In order to ensure the compliance of the manipulation and validity of the control, the ideal force/position tracing and control strategy must be implemented[8,9].

The analysis in this paper is part of the assembly planning process. Here our research results are mainly expounded in three directions. First, the geometric states are analyzed according to the mutual configurative relationships of assembly parts. Then, the assembly environment can be understood by the medial axis diagram of assembly parts. Finally, the assembly strategy based on the environment understanding and the force sensing and force control is presented. According to the above analysis, the whole assembly strategy is engendered through combining geometric reasoning, force sensing and force control.

2 Geometric Analysis

There are many configuration states for the assembly process of inserting a dissymmetrical T-shaped peg into C-shaped slot. The mutual relationships of these configuration states are analyzed using geometric analysis.

2.1 The Assembability Analysis of the Geometric and Force

Figure 1 shows the geometrical relations of the T-shaped peg and C-shaped slot. L_1, L_2 are the length of the T-shaped peg, which divided by the axes of the tail part of the T-shaped peg, and $L=L_1+L_2$. H is the width of the clearance of C-shaped slot, t_c is the thickness of the C-shaped slot, t_1 and t_2 are the thickness of the tail part and head part of the T-shaped peg respectively. Supposed $L_2=kL_1$, then $L_1=L/(1+k)$.

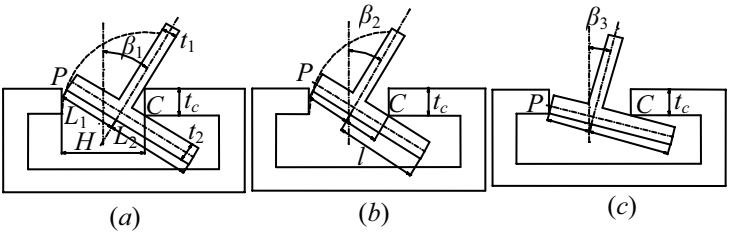


Fig. 1. The geometric representation of T-shaped part

When the contact state is as Figure 1(a), the condition of the assembability is,

$$\overline{CP} = \sqrt{(t_c \sin\beta_1 + \tfrac{1}{2}t_1 + L_1)^2 + t_2^{\,2}} \le h \tag{1}$$

When T-shaped peg rotates from β_1 to β_2 with the rotate center of the point C, as shown i n Figure 1(c) and when $\overline{CP} = h$, the permission maximum length of the T-shaped peg is,

$$L_{\max} = (1+k)\left(\sqrt{h^2 - t_2^{\,2}} - t_c \sin\beta_1 - \frac{1}{2}t_1\right) \tag{2}$$

When the contact state is as Figure 1(b), it has $\overline{CP} = \sqrt{(l^2 + L_1)^2 + t_2^{\,2}} = h$. And when Figure 1($a$), it is $\overline{CP} < h$, which can satisfy the assembly condition.

If T-shaped peg's maximum exterior contour in the scope of the C-shaped slot's clearance with the obliquity angle β, then the condition of the T doesn't contact with the C can be expressed as,

$$(L + t_2 \tan\beta)\cos\beta \le h \tag{3}$$

By which, the critical obliquity angle can be drawn.

2.2 Description of the Representation

The task of T-shaped peg into C-shaped slot is a three degree-of-freedom problem: two for translation and one for rotation. The geometry model and parameters are shown in Figure2.

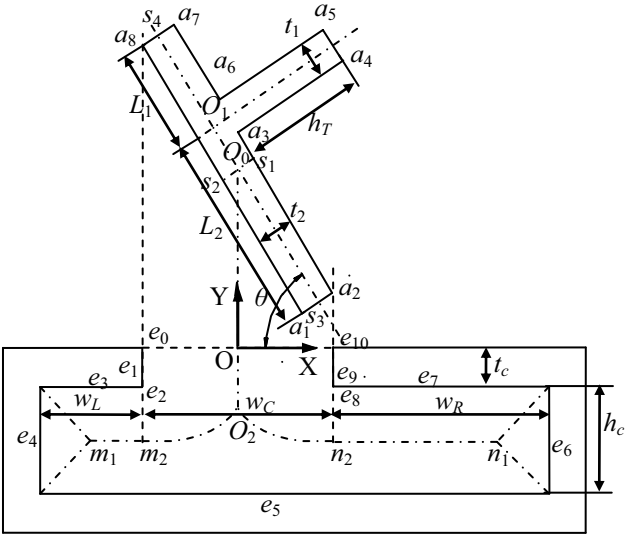


Fig. 2. The mating sketch map of T-shaped peg and C-shaped slot

Figure2 shows that the T-shaped peg and C-shaped slot are dissymmetric parts whose contours are denoted by a_1, a_2, ..., a_8 and e_0, e_1, ..., e_{12} respectively. Because of dissymmetry, the assembly task can be completed in a single direction. Let *O-XY* be a coordinate system attached to the C-shaped slot. The coordinates of the exterior contours points of the T-shaped peg are given as T (x_T, y_T) in the *O-XY* coordinate system, and the pose of the T-shaped peg is denoted by the angle θ, which can be formed from axis line of the head part of the T-shaped peg to the X axis. Thus, the T-shaped peg can be described using the coordinates (x_T, y_T, θ).

3 The Construction of the Medial Axis Diagram

3.1 Definition of the Medial Axis Diagram

The medial axis is computational geometric technology[10], which is evolved from Voronoi diagram. It partitions the space into sub-areas according to distance. Each sub-area belongs to a graphics element and the distance of the points in the sub-area to the graphics element is nearer than to the other graphics elements. If B $\{B_0,\ldots, B_n\}$ is the a obstacle set of the work space W and ∂B_i is the boundary of road-block B_i, then $\partial B = \bigcup_i \partial B_i$. If $d(p)$ is the minimum distance of a point p to a certain point of the boundary, then $d(p) = \|p - c(p)\|$. Where $c(p)$ is the points on the boundary. So, the definition of the medial axis is,

$$MA(W) = p : \exists a, b \in \partial B, \|p - a\| = \|p - b\| = d(p) \tag{4}$$

3.2 The Medial Axis Diagram of the Assembly Parts

According to the definition mention in the above, the medial diagram of the assembly parts is constructed using the algorithm of the reference [10]. For T-shaped peg and C-shaped slot, the medial line is made up by the point sets which have equal distance to the all boundary elements. So, these point sets are the bisectors of the boundaries in certain regions. The interior contour of C-shaped slot is made up by the boundary elements of e_0, e_1,…, e_{10}. The medial axis lines of the interior contour of the C-shaped slot are made up by the bisectors of $b(e_0,e_{10})$, $b(e_1,e_9)$, $b(e_2,e_8)$, $b(e_2,e_5)$, $b(e_3,e_4)$, $b(e_3,e_5)$, $b(e_4,e_5)$, $b(e_5,e_6)$, $b(e_5,e_7)$, $b(e_5, e_8)$ and $b(e_6,e_7)$. Unite all the bisectors, the medial diagram of the C-shaped slot's interior contour can be made. Figure 2 shows the interior medial axis diagram of the C-shaped slot and T-shaped peg. The medial axis line is denoted by dash dot line. The midline s_1s_2 of the T-shaped peg, intersects its axis line s_3s_4 at point O_0, which is on the medial axis line O_0O_2. All the environmental knowledge is shown in the figure 2.

Supposed that the friction coefficient and the state of the surface of the T-shaped peg and C-shaped slot are the same. The contact force of the two parts can be constant when the T-shaped peg moves along the medial axis. So, the moving track along the surface of the C-shaped slot can be called 'equal force line'. If the initial projection of T-shaped peg's exterior contour is within the C-shaped slot's clearance scope, under

the guidance of the medial axis diagram, the characteristic points on medial line of the T-shaped peg can move along the medial axis lines of the C-shaped slot under the constant contact force, such as the ideal zero force, and complete the assembly task.

In order to implement of assembly task successfully, the posture of the T-shaped peg must be adjusted until it resembles the state shown in Figure 2, where the horizontal projection distance of the T-shaped peg is smaller than or equal to the clearance length W_C of the C-shaped slot. Before the point s_3 reaches the medial axis line O_2n_2, the point O_0 moves along the medial axes line O_0O_2 need to be guaranteed. At the same time, the point a_8 must move along the direction of line e_0e_1, and line segment a_3a_4 does not contact contour element e_{10}. Then, the point s_3 moves along the medial axis line O_2n_2 and n_2n_1, until the line segment a_3a_4 contacts e_{10}. Afterwards, the posture of the T-shaped peg is adjusted until its head part is completely in C-shaped slot. After translational and rotational adjustments, the points s_3 and s_4 locate at the medial axis lines n_1n_2 and m_1m_2, respectively, and the O_0 at the extended line of O_0O_2. At last, move T-shaped peg upward until it reaches the final position. The medial axis diagram optimizes the geometric reasoning analysis to a certain extent. Although the medial axis has a good geometric character, it doesn't have the dynamic character. Due to the uncertainties in the assembly process, contact always exists. Therefore, for the success of assembly task, it is important to combine the static character of computational geometry with the dynamic character of active force control.

4 Active Assembly Strategy under Force Control

The medial axis diagram gives the all the knowledge and information of the assembly environment, including the feasible location relations and the operation procedures of the assembly task of the T-shaped peg into C-shaped slot. However, the exact states, the motion directions and the magnitude of the state transitions in the process of the assembly. How to get the exact measurable posture relations of the assembly parts is very important and difficult. Active visual and active force control are the main methods. Because of the low location precision and the complexity of the images dealing, it is not the best method in the robot assembly operation. With the medial axis diagram, robot motion planning can be carried out while controlling and keeping the contact force constant under active force control. The contact force information and the coordinates of the system, the motion direction and the location relations of the assembly parts can be known by the integrating the six-dimension force sensing information and environment knowledge.

4.1 Position and Force Sensing

According to geometric constraints and coordinate transformation, it has

$$F_a = (J^{\mathrm{T}})^{-1} F_s \tag{5}$$

where J is the Jacobian matrix, F_s is F/T value in sensor coordinate , F_a is the value in tool coordinate which is related to the direction of vector and friction coefficient. The friction coefficient can determine the directions of tangential vector and normal vector. The angle α between the normal vector n_c and resultant force F satisfies:

$$\alpha = \arctan(1/\mu) \tag{6}$$

At the beginning of assembly, the T-shaped peg is tilted at a certain angle. The robot holds T-shaped peg moving to C-shaped slot until the contact engendered. Contact information contains the direction of the contact surface normal, and its relative location and orientation. Though the locations of contact points might be estimated geometrically in motion planning, they may not be accurate due to various control errors. There are on-line contact point localization algorithms developed based on F/T sensors. Generally, the location of the contact point can not be determined from the sensed information alone. The geometric constraint $g(r)=0$ has to be imposed to get a finite number of solutions. The relation between contact position vector r and contact force F is

$$r=\Gamma(F, g) \tag{7}$$

The relation between sensor data (f_s, τ_s) and contact position vector r is

$$\tau_s=r\times f_s \tag{8}$$

Eq.(8) can be rewritten as a matrix equation using a skew symmetric matrix $\tilde{f}_s$.

$$\tau_s = \tilde{f}_s r \tag{9}$$

where

$$\tilde{f}_s = \begin{bmatrix} 0 & -f_z & f_y \\ f_z & 0 & -f_x \\ -f_y & f_x & 0 \end{bmatrix}$$

r can be determined by Equations (7) and (9).

Because the robot can gain information on the location of the contact point and the contact F/T in course of tracking the assembled part's surface, it can understand C's geometrical contour. The normal vector tilt angle of the contact points are shown in the Figure 3. In control process, the contact forces are controlled to a proper bound to maintain the force tolerance with force feedback. The changeable process of measured forces F_x, F_y in T insertion are shown in Figure 4.

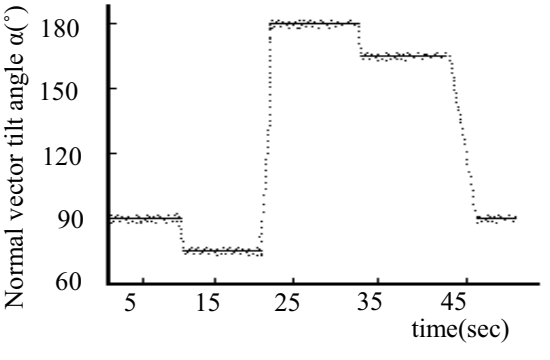


Fig. 3. Normal vector tilt angle of contact point

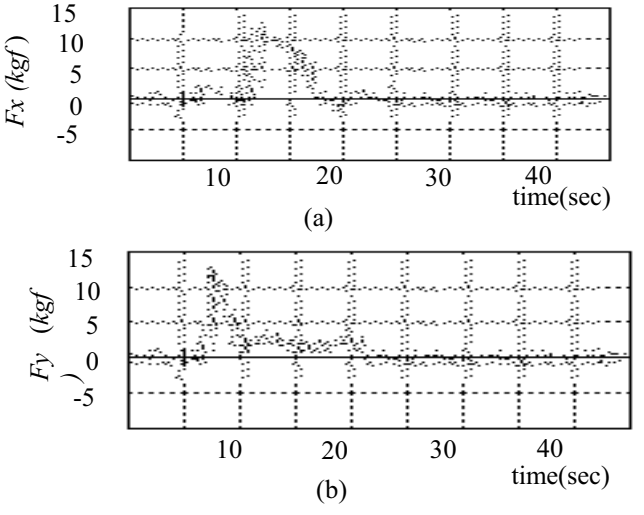


Fig. 4. The curve map of tracking force

The robot assembly manipulation is a continuous process. For computational convenience, the process can be dealt in discrete means. The velocity expression is $v = S/T$, where S is the length of the discrete step, T is the cycle of the robot's position servo. The length setting of the discrete step S is very important, because a big step could jump over several states, which could affect implementation of the assembly strategy. Here S=0.05mm and T=12ms. The core problem of sensing the exterior environment is mastering the position coordinates of several key points, such as points e_0, e_2, e_8 and e_{10}, at which the normal and tangible vector of the contact points will be changed. Based on T's geometric characteristic and kinematic behavior, the functional map relating the correction of motion is generated and stored as a control law, by which the robot can insert T-shaped peg into C-shaped slot.

4.2 The Analysis of Assembly Strategy

Figure 4 shows the motion process under contact and the sketch map of force sensing in this assembly process. The specific assembly strategy is described as follows.

As shown in Figure 5(a) and (b), the posture of the dissymmetrical T-shaped peg is adjusted. If point a_2 reaches line e_9 (see Figure 5(c)), it indicates that the adjusting position is correct. There is a critical state between Figure 5(a) and (b). Then, move upward the T-shaped peg and keep point a_2 in contact with element e_9. The contact state of point a_2 contacts reflex vertex e_{10} is the critical state of this contact state. After that, adjust the position of the T-shaped peg, until its geometrical center O_0 is located at the medial axis line OO_2. At the same time, the T-shaped peg rotates around the O_0 point, which allows the horizontal projection distance of L to be smaller than or equal to the clearance length W_C of the C-shaped slot. (see Figure 2 and 5(e)) . Before the point a_2 reaches the reflex vertex e_8, it needs to guarantee that point O_0 moves along the medial axis line OO_2. At the same time, a_2 remains in contact with line e_9,

and point a_8 moves along the direction of line e_0e_1 but does not contact line e_1 and line segment a_3a_4 doesn't make contact with contour element e_{10} during the motion process. When point a_2 has reached the reflex vertex e_8, the T-shaped peg moves to the interior of the C-shaped slot under conditions that the exterior contour line a_2a_3 of the T-shaped peg remains in contact with the reflex vertex e_8. After point s_3 has reached the medial axis line n_2n_1, point s_3 moves along it and contour line a_2a_3 keeps in contact with reflex vertex e_8. As shown in Figure 5(g) the broken dot arc is the arc whose center is e_8 and radius is W_C. When it satisfies the geometric condition $a_8e_8 \leq W_C$, T can rotate around the contact point e_8 until the element a_7 falls into the clearance of the C-shaped slot with no-jamming.

At last, T-shaped peg is moved horizontally to the left until it reaches the final assembly position.

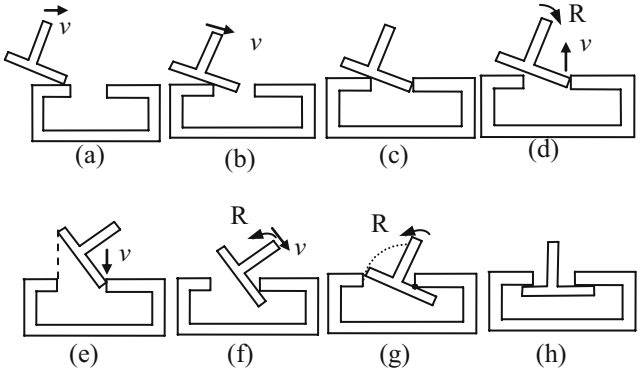


Fig. 5. The sketch map of force sensing

5 Conclusions

Insertion of complex shapes requires multi-step motions, and may require a long computation time. Motion planning strategy for a dissymmetrical T-shaped peg into C-shaped slot has been presented. For big clearances, it is easy to execute the assembly motion planning. A strategy for small clearance assembly task was presented here. The new assembly strategy is proposed based on the environmental knowledge obtained from the medial axis diagram of the assembly parts and force sensing and force control information in the process of the assembly operation. The assembling task of the T-shaped peg into C-shaped slot was completed successfully using the PUMA robot.

References

[1] Ilari, J., Torras, C.: A configuration space heuristic approach. The Int. J. of Robotics Research 9(1), 75 (1990)

[2] Grupen, R.A.: Planning grasp strategies for multi-fingered robot hands. In: Proc. of the 1991 IEEE Int. Conference on Robotics and Automation Sacramento, California, April 1991, p. 646 (1991)

[3] Sturges, R.H., Laowattana, S.: Fine motion planning through constraint network analysis. In: IEEE Int. Sym. on Assembly and Task Planning Pittsburgh, Pennsylvania, August 1995, p. 160 (1995)
[4] Schimmels, J.M., Peshkin, M.A.: Admittance matrix design for force-guided assembly. Transactions on Robotics and Automation 8(2), 213–227 (1992)
[5] Yin, Y.H., Wei, Z.X., Zhu, J.Y.: New intelligent strategies for tracing unknown environment of robot. Chinese Journal of Mechanical Engineering (Chinese) 35(1), 51–56 (1999)
[6] Krishnan, S.S., Sanderson, A.C.: Reasoning about geometric constraints for assembly sequence planning. In: Proc. of the 1991 IEEE Int. Conf. on Robotics and Automation Sacramento, California., April 1991, p. 776 (1991)
[7] Mirolo, C., Pagello, E.: Local geometric issues for spatial reasoning in robot motion planning. In: IEEE/RSJ Int. Workshop on Intelligent Robots and Sys., Osaka, November 1991, p. 569 (1991)
[8] Yin, Y.H., Fu, Z.: Robotic complex assembly strategies based on Voronoi diagram. Journal of Shanghai Jiaotong University (Chinese) 10, 1408–1412 (2002)
[9] Yanchun, X., Yuehong, Y., Zhaoneng, C.: A Novel Methodology for Complex Part Motion Planning. High Technology Letters 11(3), 245–249 (2005)
[10] Held, M.: Voronoi diagrams and offset curves of curvilinear polygons. Computer Aided Design 30(4), 287 (1998)

From Morphologies of Six-, Four- and Two-Legged Animals to the HexaQuaBip Robot's Reconfigurable Kinematics

Alexandre Veinguertener, Thierry Hoinville,
Olivier Bruneau, and Jean-Guy Fontaine

Italian Institute of Technology – Via Morego, 30 – 16163 Genova, Italy
Université de Versailles – 10/12, avenue de l'Europe – 78140 Vélizy, France

Abstract. How can we go beyond the locomotor versatility of current legged robots? We propose an approach, called HexaQuaBip, based on merging the most prevalent legged animal morphologies in a bioinspired polymorphic yet non-modular robot, intended to be able to reconfigure in either hexapodal, quadrupedal or bipedal modes. This paper focuses on reviewing main types of 6-, 4- and 2-legged animal kinematics and results in integrating all of them into a reconfigurable kinematic structure.

1 Introduction

Recently, Bullock [1] has pointed out that "biological devices are 'niche-facing' or *nichiversal*", rather than universal. In other words, because biological mechanisms are above all shaped according to their lineage's environmental conditions (or roughly speaking their niches), they are not expected to properly work apart from these "normal" conditions *(for more precisions see [1])*. This observation was originally raised against the common claim that natural and bio-inspired computing would be general-purpose. However, it likely applies to any product of natural evolution or biomimetic device and, especially for what we are interested in, to natural locomotor apparatus and their artificial counterparts.

The approach we propose, so-called HexaQuaBip, consists in studying how to design and control what would be a polymorphic *hexa*podal-*qua*durpedal-*bi*pedal robot. Our hypothesis is that such a reconfigurable artefact (yet non-modular), by merging the most prevalent legged animal morphologies, would afford unseen terrestrial locomotor versatility, that could even go beyond those of living beings. What we mean by locomotor versatility here is related to how many diverse environments a given locomotor apparatus can adapt to. Conceptually, this notion can be thought, given a locomotor multi-criterion space (eg. a combination of speed, energy, stability, agility, etc. criterions), as the volume of all tradeoffs satisfied by a locomotor apparatus and the locomotion modes it supports.

As a main constraint on shaping this volume, this article proposes a first bio-inspired kinematic structure for a reconfigurable HQB robot.

We will first review terrestrial locomotor versatility in robotics (sec. 2). Then, we will describe the core HQB approach, as well as what motivates it and how

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 1255–1265, 2009.

it relates to previous works (sec. 3). In section 4, we will describe kinematic structures of hexapodal, quadrupedal and bipedal animals, which will then be merged to define kinematics of HQB robot (sec. 5). Finally, we will conclude by giving insight about hardware implementation of a first HQB prototype (sec. 6).

2 State of the Art: Legged Robots Exhibiting the Most Versatile Land Locomotions

RHex. (Fig. 1) [2] is a remote-controlled man-portable hexapod robot inspired from cockroaches. It weighs about 8kg, while being half a meter long and 0.13m high. RHex has been designed with very simple kinematics. All its semi-wheel-like legs are actuated around only one degree of freedom.

RHex is surely the most multi-terrain hexapod ever built. RHex can move in various outdoor fields like rock, mud, sand or high vegetation. It can climb over railroad tracks, telephone poles, up steep slopes and stairways. Some bipedal walking gait has been also implemented on it with success. RHex's remarkable locomotor versatility undoubtedly comes from its particular "legs" which act as semi-wheels and involve significant compliance.

However, beyond its incredible robustness, RHex's low-articulated morphology inevitably limits locomotor adaptivity. For example, stride length is fixed by the diameter of its wheel-like legs. There is no one contact point but a succession of it and they cannot be chosen. Also, certain major environmental adaptations need operator intervention, like changing legs to paddles for swimming.

BigDog. (Fig. 2) [3] is a quadruped donkey-like robot measuring 1m in length, 70cm in height and weighing 75kg. Its morphology is composed with a rigid body, each leg is made by three segments linked by hinge joints. Legs are connected to the body by two Degrees of Freedom (DoFs) joints: one about the pitch axis and the other one about the roll axis. All joints are actuated by hydraulic pistons, which pressure is maintained by a combustion engine.

BigDog robot certainly displays the most adaptive and robust locomotion of legged robots. It has been shown running at 6.4km/h, climbing up a slope up to 35 degrees, walking across rubble and carrying 150kg load. BigDog is also able to do highly reactive balance recovery from kicking and ice slipping.

Fig. 1. RHex's versatile locomotion. Left: Walking on mud. Middle: Climbing stairs. Right: Bipedal walk.

Fig. 2. BigDog highly versatile locomotion. Left: Iced road. Right: Rocky terrain.

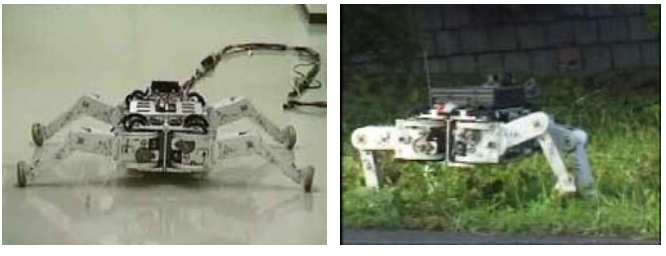


Fig. 3. Roller-Walker hybrid locomotion. Left: Skating on flat ground. Right: Walking in the grass.

The great adaptivity of BigDog obviously comes from its bio-inspired kinematics, its powerful actuation combined to its highly reactive and dynamic control strategy. Since this project is supported by military program, little is known about BigDog control implementation. But, we can infer from available demonstration, that its highly dynamic gaits are not compatible with static stability and energy saving that are desirable at low speeds. This assumption is confirmed by the designers' choice of using strong yet noisier internal combustion engine. Also, even at very low speed, BigDog displays highly dynamic gaits as trot, which could be replaced by walking gait for better energy consumption, as horses do in nature.

Roller-Walker. (Fig. 3) [4] is a hybrid locomotion quadruped robot, equipped with a special foot mechanism. Each of its four legs has three actuated DoFs plus a special ankle joint, that allow switching between passive wheel (rolling mode: horizontal wheel's axis) and round foot sole (walking mode: vertical wheel's axis).

Thus, on flat surfaces, Roller-Walker can move quickly using few energy by using skating motion. This mode can be also used to climb up or down little slopes. In contrast, walking mode is slower, but allows Roller-Walker to cross obstacles or moves on uneven surfaces.

However, its control is mostly based on preprogrammed gaits and feet reconfiguration, which limits greatly its locomotor adaptivity and robustness.

Table 1. The HQB approach related to other works. From conventional non-reconfigurable robots, through reconfigurable ones, up to multi-robot systems (aka. robotic swarms), we can draw a continuous trade-off between potential versatility and simplicity of control. At one extreme, non-modular robots like BigDog seem easier to control, but uniqueness of their morphologies inspired from given animal species constrains their applicability to restricted niches. Self-reconfigurable modular robots offer more versatility through their adaptable metamorphic morphologies, but complex algorithms are needed to plan efficient docking sequences of their cells. At the other extreme, swarm-based robots, especially heterogeneous ones, can spread through virtually any environments, but their emerging behaviors are very difficult to handle. Within HQB project, we propose to extend multimodal locomotion approach, explored in projects such as Roller-Walker or Gorilla Robot, by designing a polymorphic non-modular robot that can reconfigure as six-, four- and two-legged animal main taxa.

Versatility	−−	−	+	++
Simplicity	++	+	−	−−
Approaches	Conventional	Self-Reconfiguration		Swarm-based
Morphology	Monomorphic	Polymorphic	Metamorphic	Amorphic
Modularity	Standalone		Dockable Cells	Autonomous Agents
Bio-Inspiration	Species	Taxon	Tissue	Natural Swarming
Sub-approaches	*Unimodal locomotion*		*Chain-type*	*Homogeneous*
Project examples	• BigDog [3]	*Multimodal locomotion*	• Polybot [7]	• Swarm-bots [10]
		• Roller-Walker [4]	• M-TRAN [8]	
		• Gorilla Robot [5], [6]	*Lattice-type*	*Heterogeneous*
		• **HexaQuaBip**	• ATRON [9]	• Swarmanoid [11]

3 The HexaQuaBip Approach

Locomotor versatility requires from morphology a great flexibility in term of performance achievement. For instance, we expect a versatile locomotor apparatus to be able to exhibit, *at least successively*, from robust to rapid locomotion, through agile or energy-conservative one, in order to adapt to various environments. Given the diversity of existing land environments, we believe that the sole realistic alternative is to resort to multimodal locomotion. Obviously, conventional "fixed-morphologies" are somewhat limited to afford it. Conversely, metamorphic morphologies, as devised in the currently developing field of self-reconfigurable robotics, promise to exceed conventional robot flexibility [8]. Indeed, thanks to their high modularity, which is based on using "mechanical cells" that can connect each other, metamorphic robots are able to change their own shapes according to new circumstances. Thus, they are thought, far beyond locomotor versatility, to be potentially multifunctional.

However, such a cellular approach implies an extremely challenging complexity [8]. First, hardware implementations usually rely on small and fragile mechatronics, especially as regards the docking system of modules. Second, powerful algorithms are often needed to tackle the combinatorial explosion involved in shape-shifting. Third, distributed control assumes to solve hard issues like information synchronization or shared decision.

Therefore, it seems painful to follow this approach in order to achieve locomotor versatility "only". Instead, we advocate a better compromise between

versatility and complexity, namely the design of polymorphic standalone robots, rather than metamorphic modular ones (see Tab. 1). Such less reconfigurable robots already started to be investigated in at least two projects: Roller-Walker [4], that can switch between either walking or skating locomotion; and Gorilla robot [6], that is intended to mimic natural morphological flexibility of some apes, who are able to do brachiation, bipedal and quadrupedal locomotion.

With the HQB approach, we aim to go further in robotic polymorphism and multimodal locomotion, by designing kinematics that allows a non-modular robot to reconfigure in most species included in hexapods, quadrupeds and bipeds taxa.

Why drawing inspiration from nature, if we pointed out in introduction the nichiversality of biological devices? Because, while it is true that individual species display a limited locomotor versatility, this is not the same for upper animal taxa, which, being considered as a whole, spread over large and heterogeneous environments. Thus, the taxon gathering hexapods, quadrupeds and bipeds turns out to inhabit virtually all land surfaces. Hence, by merging as much as possible legged morphologies in one polymorphic robot, we hope this one embodies the locomotor versatility characterizing the whole taxon of legged animals. Moreover, through their phylogeny, the imitated legged animals necessarily share some similarities that we expect to exploit.

4 Two-, Four- and Six-Legged Animals' Morphologies

In a previous work [12], we have reviewed the main DoFs and segment length ratios of animals' limbs and merged them into a versatile reconfigurable four-segmented leg. The resultant kinematics appears as following: three DoFs between the body and the first segment, one DoF between the 1st and 2nd segment, three DoF linking the 2nd and 3rd segment and a last hinge joint between the 3rd and 4th segment. To those primary results, here we add the various DoFs that the body minimally requires for each locomotion mode and reconfigurability.

Hexapods: Insects' body is considered owing two DoFs, one between each thorax part, both hinge-like about pitch axis. Those DoFs allow the cockroach to flex its body in order to help changing from one surface to another with different inclinations. For instance, when the insect is moving from a vertical surface to a horizontal one or when it crosses obstacles [13].

Quadrupeds: While walking, legged reptiles adopt a sprawled posture relatively similar to insects [13]. Their body mostly lies on the ground. Hind limbs, which could be related to cockroach rear legs [13], being mostly rear oriented and moving in a plane almost parallel to the ground. Fore limbs, assimilated to the front legs of cockroaches [13], move along a plane mostly perpendicular to the spine axis. Reptile's spine, on contrary to the quite rigid exoskeleton of insects, bends greatly about yaw axis during locomotion (see Fig. 4) and from this comes a significant increase of stride length. Blob [14], Goslow [15] and Biewener [16]

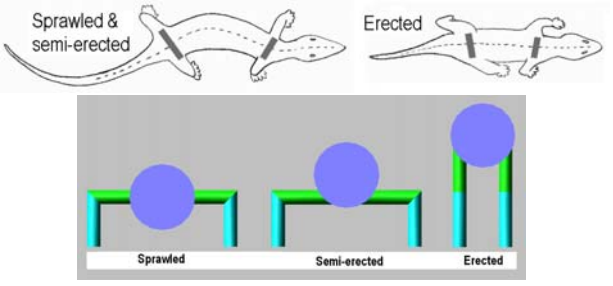


Fig. 4. Top: Effect on spine. Bottom: Reptile postures.

investigated the evolution of postures, from sprawled or semi-erected to erected ones (see Fig. 4). They have showed that heavy reptiles, such as alligators or iguana, decrease the stress in their bones created from weight and torsion, by taking more erected postures. Reptiles are thus adopting postures closer to mammals. In this case, front and hind limbs operated mostly under the body. The spine does not bend about the yaw axis anymore but stays relatively straight, especially while running [14].

Concerning now mammals, the flex-ext of their spine about pitch axis allows an augmentation of stride length especially in the case of fast gaits such as gallop and gives a good impulse in the case of bouncing or jumping. This flex occurs mainly at the second third of the spine starting from head, because of the ribcage rigidity.

Bipeds: Human morphology has been investigated extensively [17,18,19]. Based on those anthropomorphic models, Bruneau [20] established one taking into account the minimal requirements for locomotion such as geometry, mass, joint motion. As result, six DoFs have been allocated to each leg: three DoFs in the hip, one in the knee and two for ankle (pitch and roll axis). The rest of the body helping in stability control have been model as following: three DoFs for each arm (two for the shoulder, one for the elbow), four DoFs for torso (three linking the pelvis to the torso and one folding fore-aft the torso itself into two parts, e.g. Fig. 7).

Regarding birds, Hugel et al. [21] consider the body like one none articulated part.

5 HexaQuaBip's Kinematic Design

Here, we propose to merge them to figure out the DoFs that the HQB must minimally display in order to fit all locomotion modes.

Our approach is to draw analogies between above described kinematic models.

In the hexapod section (4), we have seen that in terms of structure HQB needs three segments (prothorax, mesothorax, metathorax: dark blue and silver in Fig. 5) making up the body and six legs composed by four segments. Regarding

Table 2. Instantiation of natural trunk/thorax segments into HQB's morphology. (*) Here, we take inspiration from mammalian quadrupeds' hindlimbs only, because their forelimbs display a huge variety related to not only locomotory function.

Locomotion modes		**HQB's body segments**				
		b1	**b2**	**b3**	**b4**	**b5**
Hexapod		prothorax		mesothorax	metathorax	
Quadruped		pelvis(*)		trunk		pelvis
Biped	Theropod+Bird	upper body (ie. without the legs)				
	Human	torso+shoulders			lumbars	pelvis

Table 3. DoFs needed for HQB's body according to locomotion modes. Orientations (according to hexapodal standard posture, Fig. 5): Y, yaw; P, pitch; R, roll.

Locomotion modes		**HQB's body joints**			
		b1/b2	**b2/b3**	**b3/b4**	**b4/b5**
Hexapod			P	P	
Quadruped	Mammals		P	P	
	Reptiles	Y			Y
Biped	Theropod+Bird				
	Human			P	Y,R
HexaQuaBip		**Y**	**P**	**P**	**Y,R**

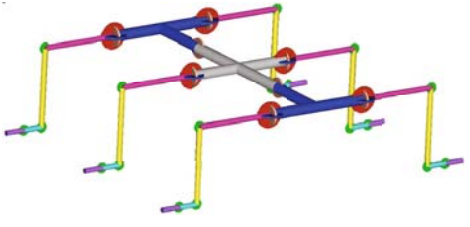


Fig. 5. Minimal HQB's kinematics for hexapods

DoFs, we showed that the body must be endowed with one articulation between each body segment, about pitch axis, allowing the robot to adapt to the ground properties and to cross over high obstacles. Thus Hexa- kinematics will be like shown in Fig. 5 and own the DoFs summarized in Table 3 and 2.

From the point of view of mammalian quadrupeds, an important point is the spine flexibility, ventrally-dorsally, at two third of spine length starting from head. Then, a good way to stand HexaQua- frame is on the front and rear legs taking advantage of the body joint already used in Hexapod mode. As second point is necessity to have pelvis mobility around yaw axis for reptiles. Consequently, we choose to cut each pelvis (e.g. prothorax and metathorax: dark

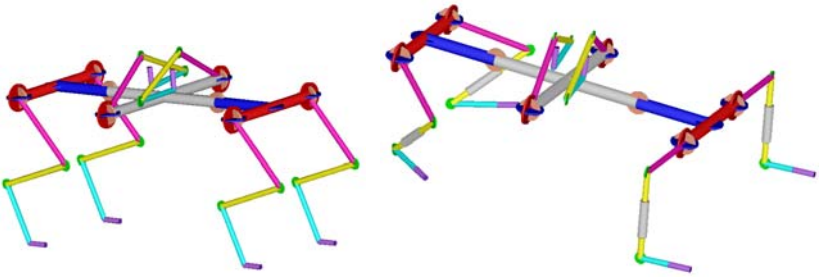


Fig. 6. Minimal HQB's kinematics for hexapods plus quadrupeds. Left: Mammal posture. Right: Reptile posture.

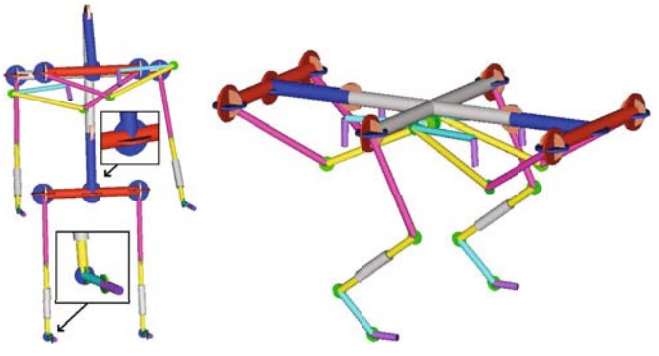


Fig. 7. Complete HQB's kinematics for hexapods plus quadrupeds plus bipeds. Left: Human posture. Right: Bird posture.

blue in Fig. 5 becoming red and dark blue in Fig. 6) into two parts in order to endow them with one rotating DoF about yaw axis (e.g. Fig. 4).

For biped case, on one hand we have birds which need a structure made out of opposed double pendulum and on the other hand, human being which is made out of one or more inverted pendulum.

To cope with human frame, it is first necessary to erect the robot. Since we have a pelvis and a long torso, we have chosen to stand the robot on the rear legs. In this position, we can easily make parallel between HexaQuaBip frame and human model. In fact, according to our review (see section 4) only one DoF has to be added to the torso (about yaw axis while structure erected). That new kinematics of HexaQuaBip frame is shown in Fig. 7 and summarized in Table 3 and 2.

On the other hand, for birds case it appears judicious to use the middle legs in order to balance HexaQuaBip's body such as birds do. In this configuration, body's DoFs themselves are not really useful since bird's body does not flex greatly. However, having those DoFs gives a good opportunity to manage robot weight in order to optimize its center of mass control and recreate the opposed pendulum configuration of birds (see Fig. 7).

6 Hardware Implementation

HexaQuaBip hardware implementation is far from being trivial. In fact, its great number of degrees of freedom allows the robot to move in whatever terrain but the same complexity make tough the mecanical design. However in HQB concept, it is essential to have a real robot in order to put the robot is true situation that are often hard to model such as mud, snow, etc.

For the hardware implementation we plan, as preliminary study, to use actuators with variable stiffness of the main body and hip joint that swing the legs will be obtained by serial elastic actuators with active/passive compliance [22] for multi degrees of freedom joint. In fact, this actuators offers an useful damping foot strikes. Another advantage of those actuators is their small size that can directly be included into the core of a part itself such as pelvis. For HexaQuaBip, it is also needed to have a good range of rotation to change from quadruped (torso horizontal) to human configuration (torso vertical). On the joint between the 1st and 2nd segment will be implemented, by the mean of electric motor a cabled solution. This solution allows us to put the actuator along the segment and so to have a mass spread along the whole 1st segment (and not down in joint). This solution can also endow a passive compliance by putting a spring with a stiffness [23] adaptive to reconfiguration of the robot.

7 Conclusion

In this paper, we reviewed various most important kinematics of hexapodal, quadrupedal and bipedal animals. Then, we merged them to obtain a polymorphic reconfigurable robot expected to display a greater locomotor versatility.

As regards perspective of this work, a next step of our project will be to study deeper technological implementation of the obtained reconfigurable kinematics using currently available hardware such as serial elastic actuators. A complementary task would be to design reconfigurable feet, from unguligrade hooves to plantigrade soles, due to their huge importance in locomotion. At mid- and long-term, bio-inspired control strategies will be proposed to achieve gait adaptations to morphology reconfigurations and morphology optimization for each fixed gait.

References

1. Bullock, S.: The fallacy of general purpose bio-inspired computing. In: Rocha, L.M., Yaeger, L.S., Bedau, M.A., Floreano, D., Goldstone, R.L., Vespignani, A. (eds.) Artificial Life X: Proceedings of the Tenth International Conference on the Simulation and Synthesis of Living Systems, pp. 540–545. The MIT Press, Bradford Books (2006)
2. Neville, N., Buehler, M.: Towards bipedal running of a six legged robot. In: 12th Yale Workshop on Adaptive and Learning Systems (2003)

3. Raibert, M., Blankespoor, K., Nelson, G., Playter, R.: The BigDog Team: Bigdog, the rough-terrain quadruped robot. In: The International Federation of Automatic Control (2008)
4. Endo, G., Hirose, S.: Study on roller-walker (multi-mode steering control and self-contained locomotion). In: International Conference on Robotics & Automation (2000)
5. Aoi, S., Egi, Y., Ichikawa, A., Tsuchiya, K.: Experimental verification of gait transition from quadrupedal to bipedal locomotion of an oscillator-driven biped robot m. In: International Conference on Intelligent Robots and Systems (2008)
6. Aoyama, T., Sekiyama, K., Hasegawa, Y., Fukuda, T.: Analysis of relationship between limb length and joint load in quadruped walking on the slope. In: International Conference on Intelligent Robots and Systems IROS 2008, pp. 3908–3913 (2008)
7. Yim, M., Zhang, Y., Roufas, K., Duff, D., Eldershaw, C.: Connecting and disconnecting for chain self-reconfiguration with polybot. Proc. IEEE/ASME Trans. Mechatron 7(4), 442–451 (2002)
8. Murata, S., Kurokawa, H.: Self-reconfigurable robots: Shape-changing cellular robots can exceed conventional robot flexibility. IEEE Robotics & Automation Magazine 14, 71–78 (2007)
9. Jorgensen, M., Ostergaard, E., Lund, H.: Modular atron: modules for a self-reconfigurable robot. In: Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2004), vol. 2, pp. 2068–2073 (2004)
10. Groß, R., Bonani, M., Mondada, F., Dorigo, M.: Autonomous self-assembly in swarm-bots. IEEE Transactions on Robotics 22(6), 1115–1130 (2006)
11. Christensen, A., O'Grady, R., Dorigo, M.: Morphology control in a multirobot system. IEEE Robotics & Automation Magazine 11(6), 732–742 (2007)
12. Veinguertener, A., Hoinville, T., Bruneau, O., Fontaine, J.G.: Morphological design of the bio-inspired reconfigurable hexaquabip robot. In: Climbing and Walking Robots and the Support Technologies for Mobile Machines (2009)
13. Ritzmann, R.E., Quinn, R.D., Fischer, M.S.: Convergent evolution and locomotion through complex terrain by insects, vertebrates and robots. Arthropod Structure & Development 33, 361–379 (2004)
14. Blob, R.W., Biewener, A.A.: In vivo locomotor strain in the hindlimb bones of alligator mississippiensis and iguana iguana: implications for the evolution of limb bone safety factor and non-sprawling limb posture. Journal of Experiment Biology 202, 1023–1046 (1999)
15. Goslow, G.E., Reinking, R.M., Stuart, D.G.: The cat step cycle: hind limb joint angles and muscle lengths during unrestrained locomotion. Journal of Morphology 141(1), 1–41 (1973)
16. Biewener, A.A.: Biomechanical consequences of scaling. Journal of Experiment Biology 208(Pt 9), 1665–1676 (2005)
17. Hanavan, E.P.: A mathematical model of the human body. AMRL TR 102, 1–149 (1964)
18. Hodgins, J.: Three-dimensional human running. In: Proc. IEEE International Conference on Robotics and Automation, vol. 4, pp. 3271–3276 (1996)
19. McKenna, M., Zeltzler, D.: Dynamic simulation of a complex human figure model with low level behavior control, vol. 5, pp. 431–456. MIT Press, Cambridge (1996)
20. Gravez, F., Bruneau, O., Ouezdou, F.: Analytical and automatic modeling of digital humanoids. International Journal of Humanoid Robotics 2, 337–359 (2005)

21. Hugel, V., Hackert, R., Abourachid, A.: Exploiting bird locomotion kinematics data for robotics modeling. CoRR abs/0807.3225 (2008)
22. Tsagarakis, N.G., Laffranchi, M., Vamderborght, B., Caldwell, D.: A compact soft actuator unit for small scale human friendly robots. In: International Conference on Robotics and Automation (2009)
23. Blickhan, R., Full, R.J.: Similarity in multilegged locomotion: Bouncing like a monopode. Journal of Comparative Physiology 73, 509–517 (1993)

Symbricator3D – A Distributed Simulation Environment for Modular Robots

Lutz Winkler and Heinz Wörn

Institute for Process Control and Robotics (IPR), Karlsruhe Institute of Technology (KIT), D-76131 Karlsruhe, Germany

Abstract. A distributed simulation environment is introduced in this paper that allows to simulate robot swarms as well as modular robots. This simulator is in use within the European projects 'SYMBRION' and 'REPLICATOR'. The project's main goal is to build robots that can aggregate to multi-robot organisms. These robots will be able to act individually, as a swarm or within an organism. The simulator needs to be able to simulate all of these behaviors. As the number of robots can be large, both in the swarm as well as in the organism, the simulator needs to be distributed on several computers. As the demand varies between the different behaviors –swarm and organism behavior– the behavior of the simulator needs to vary as well, e.g. for a swarm, a dynamic simulation is not necessary, whereas for an organism, a fast dynamic simulation is obligatory. We are therefore developing the Symbricator3D simulation environment, that will fulfill all the described requirements.

1 Introduction

The SYMBRION[1] and REPLICATOR[2] project's main goal is to develop robots that are cabable of building a swarm as well as building large organisms for investigating and developing novel principles of evolution and adaptation for symbiotic organisms [1,2,3]. Each robot is fully autonomous and can act in the environment either alone or can be physically connected to other robots to build a larger organism to solve tasks that it cannot solve alone. Such an organism will be provided with an internal bus system for sharing computing power. Additionally, robots can share electrical power within the organism. Combining evolutionary principles with swarm behaviors, the organism will autonomously manage their own hardware and software organization to become self-configuring, self-healing, self-optimizing and self-protecting. Combining the advantages of swarm robotics and the advantages from reconfigurable systems, adaptive, evolvable and scalable robotic systems can be built, which are able to co-evolve and cooperate with each other without human supervision in new and unpredictable environmental situations.

As the Symbrion and Replicator robots are currently not available and experiments with them will cost an extensive amount of time, an accurate and fast simulation is required to evaluate novel bio-inspired algorithms. The simulator does not only need to

[1] www.symbrion.eu
[2] www.replicators.eu

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 1266–1277, 2009.

provide an accurate dynamic simulation, but also a detailed sensor and actuator model. The algorithms and programs that are later meant to run on the robot platform also need to run in the simulator without major changes in the source code. Last but not least, as the high demands to the simulator cannot be achieved by a single computer, a distributed simulation is mandatory. We are developing therefore the Symbricator3D simulation environment to comply with all the high demands given in these projects.

The paper is organized in the following way. We will first give a short background overview where we will describe the underlying framework and why we choose that particular framework. Then we will introduce the design implementation of Symbricator3D and describe the different modules of the simulator. At the end we give a short evaluation followed by conclusion.

2 Background and Motivation

A simulation environment for modular robots, such as those we are building in the SYMBRION and REPLICATOR projects need to cover several tasks. Those are in particular:

Single Robot Control. Each single robot needs to be seen as an entity and needs to be simulated separately to help developing swarm algorithms. The single robot can be simulated dynamically, but does not need to be simulated that way. A 2D-simulator is often sufficient enough.

Organism Motion. When the robots are connected together, they build an organism. This organism also needs to be simulated. As the organism often accomplishes movements in the 3D-space, a 3D-simulator, that either simulates the movements kinematically or dynamically is essential.

Evolving Organisms Structures. Like natural evolution, different organism configurations will compete each other in every generation. The best configurations will be used for the next generation to form slightly different organisms. Over time the population improves, and eventually a satisfactory organism will be evolved that can be used on the real robot. The simulator needs to provide functions to support organism evolution.

Developing and Evaluating Control Algorithms. Like organism structures, organism control algorithms will also be evolved and the simulator needs therefore provide functionality for evolutionary control algorithms.

According to these different tasks the simulator needs to fulfill, a lot of requirements are arising. The simulator needs to be able to simulate large robot swarms as well as complex organisms containing lots of degrees of freedom. Different dynamic and non-dynamic simulators within the simulation engine are therefore necessary and switching between them during runtime is mandatory. Additionally, sensors and actuators of each robot need to be simulated as accurate as possible. Detailed sensor and actuator models are therefore important. The algorithms and programs that are later meant to run on the robot platform also need to run in the simulator without major changes in the source code. Different environmental features, either artificial like electrical sockets or switches or natural like temperature or humidity need to be implemented in the

simulator to support different scenarios. As sixteen partners with more than 50 people are involved in both projects who are using and developing the simulation, a modular design of the simulator is mandatory, so that for each experiment a different configuration can be used. This modular design will also help to distribute the simulation over different computers in a network. Such a distributed simulation is also strongly required, as the requirements to the whole simulator are too high to be executed on a single computer.

These high demands are only covered to some extend by current open source robot platform simulators, such as the very popular Player/Stage project [4], which together with the Gazebo project offers a platform for dynamics simulation and also supports controller distribution over a network. A fully distributed simulation as it is described in [5] however is not supported. The ability to use and switch between different dynamic and non-dynamic simulators at runtime need also be implemented as the user currently need to decide before starting the simulation whether he wants to run a fast 2D simulation (Stage) or a 3D dynamic simulation (Gazebo).

We decided therefore to build a new simulator that builds on the Delta-3D game engine[3] which is also appropriate for a wide variety of modeling and simulation applications. It offers among others support to different dynamics engines such as Open Dynamics Engine(ODE) and Nvidia PhysX as well as an physics abstraction layer (PAL) to include other dynamics engines. Additionally a variety of network libraries are supported such as the High Level Architecture [6], a state-of-the-art distributed simulation standard. With its modular design Delta-3D also eases the development and the maintenance of the simulator.

The Symbricator3D simulation environment is built upon the ”Delta-3D” game engine, which is a well-supported and a fully-funded open source project [7]. It is also appropriate for a wide variety of modeling and simulation applications. It supports a wide range of driver libraries such as the Open Dynamics Engine (ODE) and Nvidia PhysX for dynamic simulation, Open Scene Graph (OSG) for the graphic display and the High Level Architecture (HLA) for distributed simulation [5]. Its modular architecture allows full access to these underlying libraries so that we have full flexibility in developing software without being restricted to the API of Delta-3D, which also offers a tool which facilitates in easy creation of three dimensional maps such as a ”maze”, which can be loaded into the simulation environment during the runtime. With this tool we have the possibility of loading different scenarios during runtime, test different controllers and evaluate their behavior very easily.

One of the major concerns in the simulation is the dynamics engine. As organisms consisting of lots of robots need to be simulated, stability is a crucial criteria. On the other hand – as the projects are declared as open source, open science and open hardware – a commercial solution such as the ”Havok” engine[4] is not desirable. Additionally, a platform independent dynamic engine is preferable, however not obligatory. Delta-3D supports the Open Dynamics Engine(ODE) by default. ODE does however not offer the stability such as the Nvidia PhysX engine, which is also supported by Delta-3D, but currently runs only under Windows platforms. A third option will be the

[3] www.delta3d.org

[4] www.havok.com/content/view/17/30

Impulse-Based-Dynamic-Simulation (IBDS) [8], which is a very promising and a stable dynamics engine that can also be included in the simulation environment. Another mandatory requirement for the dynamic simulation – to ensure stability and simulation rate – is sourcing the computation of the dynamics out to the graphics processing unit (GPU). Currently, we are using ODE for the dynamics simulation. It has been however shown that the ODE is not suitable for our needs and therefore we will switch to another dynamics engine. For that reason, we have preselected the two previously mentioned dynamics engines (IBDS and PhysX). Whereas Nvidia PhysX engine is also supported by Delta-3D and a GPU implementation for Nvidia Graphic cards already exists, the IBDS uses a new, impulse-based approach to simulate dynamics and promises to be more stable and faster. A shader-based implementation for running the IBDS on the GPU is currently under development.

Another very important point is the distributed simulation of the organisms. Besides the dynamic simulation, the simulation of the controller by their own will require a lot of computational power, as each robot will have high computational capability – each robot will be equipped with at least one Cortex[5] microcontroller and one Blackfin[6] microprocessor. Additionally, for each robot all the sensors and actuators need to be simulated accurately. A distributed simulation on different computer systems is therefore essential. The High Level Architecture(HLA) standard is not only supported by Delta-3D, but also offers the possibility for different kinds of simulations to interact with eachother. For example, the dynamic simulation can run separately on a highly GPU-powered computer, whereas the controller simulations can run on another computer, while the arena simulation for evaluating bio-inspired algorithms runs separately. Each of these simulations will use the HLA standard to communicate with each other via a runtime interface (RTI). The RTI is the interface between all those simulations. We use the common, open-source and platform-independent runtime interface CERTI. With this distributed simulation standard we are not only able to run the simulation distributed, but also the parallel development of the Symbricator3D simulation will be assuaged.

3 Implementation

The Symbricator3D simulation environment is programmed in C++ and consists of three main classes: robot actors, robot controllers and simulation components. All these classes are managed by the Delta-3D GameManager, which is responsible for the management and the interaction between all the elements in the scene. The robot actor class represents the geometrical and physical properties of the robot, whereas the robot controller class represents the software that is running on the robot. We are about to create the interface in a way that without or at least with minimal changes in the source code the controller written for the simulation can be transferred to the robot. Controller can run integrated in the simulation, remotely via a TCP/IP connection or on a microcontroller, e.g. on the Cortex controller connected via a RS-232 interface to the simulator.

[5] www.arm.com/products/CPUs/ARMCortex-M3.html

[6] www.analog.com/en/embedded-processing-dsp/blackn/products/index.html

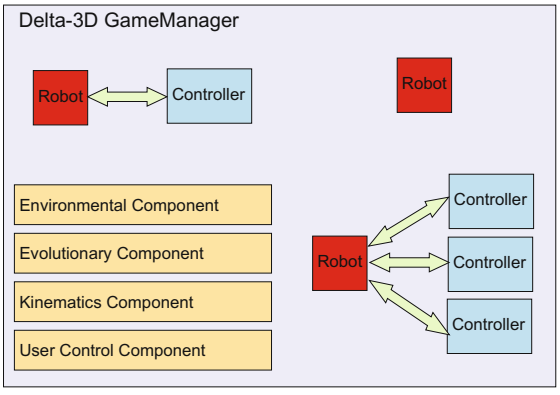


Fig. 1. Structure of the Symbricator3D. The GameManager controls every element in the simulation. Robot Actors (Robot) represent the physical and geometrical properties of the robot and are controlled by controllers. Simulation components are global components for diverse tasks, such as for simulating environmental variables, for evaluating a population or for controlling the simulation.

One robot actor can have several robot controller, but a robot controller can only control one robot actor. The robot actor together with its controllers build the complete robot model in the simulation. The third class, the simulation components are for the control of the simulation, e.g. for evaluating the fitness of all robots or for starting a new population or for giving the user the possibility to interact with the simulation, for example via a graphical user interface. Figure 1 shows the setup of the simulator.

3.1 The Robot Class

The robot actors in Symbricator3D consists of four different elements: bodies, actuators, connectors and sensors. Figure 2 shows the configuration of the robot model in the simulation based on the first prototype.

Bodies describe the geometrical and physical properties of parts of the robot such as the collision shape and the mass. A body is a single unit, there are no parts inside a body that are movable. Actuators are those components that connect two bodies together and that can put a force between these two bodies that they move relatively to each other. Actuators are realized by using joints provided by the dynamics engine. Connector modules for connecting two robots together and sensors modules for perceiving the environment are attached to a body. Elements are stored in so called container classes, which collect all robot elements of the same type. For each type of element there exists a different container class, and there exists therefore a body container, a sensor container, etc. As the simulator can access the robot elements during runtime without prior knowledge of the robot's assembly, creating a robot this way is very flexible, and not only Symbricator robots can be created and simulated, but also other robots in the field of modular and swarm robotics.

During the initialization phase the elements are created and put together according to the given robot design. This way, not only different robots can be created in the

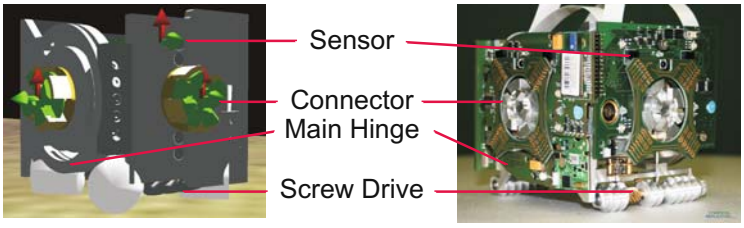


Fig. 2. Current robot design in Simulation compared with the first prototype

simulation, but also tools that robots can attach too. Tools are objects which are not fully autonomous robots, but give the organism additional sensors such as a high resolution camera or additional actuators, such as wheels to make the organism move faster in flat terrain, or other abilities such as a charging unit which enables the organism to recharge its robots at a common power socket.

Sensor Simulation. Currently there are several sensors implemented in the Symbricator3D simulation environment. These sensors are divided in two different classes: internal and external sensors. Internal sensors describe properties of the robot, such as the position of the actuator or the position and orientation of the robot in a global coordinate system. These values can be retrieved from the dynamics simulation or from Delta-3D.

External sensors are those sensors that measure the environment. Currently we have two different classes of external sensors. The first type are the camera-based sensors. These sensors use the camera images from OpenSceneGraph(OSG) to retrieve the sensor data. OSG images do not only provide RGB values in the images, but also the depth component or the luminance for example. Using the OSG images different sensors can be implemented, such as cameras, light sensors, distance sensors, or infrared sensors. The second type of sensor would first need a model which describes the behavior of the environmental properties it describes. This can be the temperature, humidity or gases. These environmental properties are simulated in a separate component in the simulation and the sensors retrieve the data accordingly to their position and orientation.

A very important point in sensor simulation is the development of an accurate sensor model, which should include simulation of several types of deviations, such as noise, drift and sampling rate. We currently focus on improving the sensor model by taking into account these deviations so that the sensor models in simulation represent the sensors on the robot as realistic as possible.

Actuator Simulation. Currently two different robot designs [3] are implemented. One design uses a chain drive and enables the robot to move over small obstacles, while the other one uses a nearly holonomic screw drive that enables the robot to be more flexible in building organisms. Each of these designs includes a main hinge, which is responsible for organism movements.

The main hinge is a joint that can be simulated by using the appropriate joint model of the dynamics engine; the chain and the screw drive are actuators, that can only be

simulated by introducing several helper wheels, that are connected to the main body by a joint and controlled synchronously. The chain drive consists of two chains that move around the robot. As a detailed model of the chains would consume too much computing power the chain drive will be simulated by eight helper wheels. They are connected by a hinge joint to the main body. Figure 3 shows the modeling of the chain drive and the screw drive. The four right wheels (”R”) represent the right chain and the four left wheels (”L”) the left chain. They are controlled exactly like the chain which they represent.

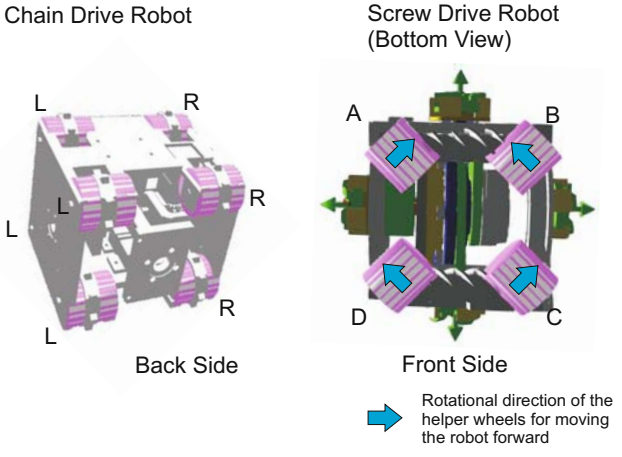


Fig. 3. Simulation model of chain and screw drive. The chain drive is simulated by four right wheels(”R”) and four left wheels (”L”). The screw drive is simulated by four wheels (”A”,”B”,”C”,”D”).

Simulating the screw drive in detail would be even more complicated as the collision model for the screw itself consists of a lot of polygons and collision detection would cost enormous computational power. The screw drive will be approximated therefore by four helper wheels, which are depicted in Figure 3. The screw drive allows four translatory movements (right, left, forward and backward) and four rotatory movements (right/left rotation) around the left axis or right axis. The different movements and how they can be accomplished on real robots and in simulation are listed in List 1. A combined movement of rotation and translation will be accomplished by superposition. The proportion of translatory movement and rotatory movement are therefor calculated and the four wheels will be controlled according to theses mixture of movements.

Like sensors, actuators are also required to be simulated according to noise and drift. Additionally, actuators have constrains such as maximum velocity and force, minimum and maximum joint angles. In the course of the sensor improvement, we are also currently improving the actuator models according to these deviations.

Connectors. Connectors are the physical and graphical representation of the coupling device of the Symbrion and Replicator robots in the simulation, with which the robots

Table 1. Movement possibilities of the screw drive, how they can be accomplished and how they will be simulated

Type of Movement	Front Screw	Back Screw	A	B	C	D
Translation:						
Forward	+1	+1	+1	+1	+1	+1
Backward	-1	-1	-1	-1	-1	-1
Right	+1	-1	+1	-1	+1	-1
Left	-1	+1	-1	+1	-1	+1
Rotation:						
Right around Front Screw	0	-1	+1	-1	0	0
Left around Front Screw	0	+1	-1	+1	0	0
Right around Back Screw	-1	0	0	0	-1	+1
Left around Back Screw	+1	0	0	0	+1	-1

can connect to eachother and once connected share energy and computational power. If a robot wants to connect with another robot, Symbricator3D checks if there is a robot close enough, has the right orientation and is ready for connection. If all three conditions are fulfilled the two robots will be connected automatically.

Besides the physical connection of two robots, the connectors, like the coupling devices on the real robot, also provide communication inside the organism. Two kinds of communication are available within the organism, local communication between two connected robots and organism communication.

Communication. The Symbrion and Replicator robots will be equipped with a variety of communication devices. They will have infrared sensors with which they can not only detect obstacles, but also can communicate with robots in their neighborhood. Additionally, they will be equipped with a ZigBee module for global communication. Within the organism they will have the ability to exchange information with all the other robots in the organism via an internal bus system, or they can send messages to those robots that are directly connected to them. Besides the infrared communication, which we are currently implementing together with the sensor improvement, all other communication types are already provided by Symbricator3D.

A communication message in the simulation consists of a string. The string can represent anything, from a single integer number to a whole program segment that needs to be interchanged with another controller. Symbricator3D checks which robot controller can get the message according to communication type and accessibility, and then sends the message automatically to the right controllers. For example, if a robot sends an infrared message, only the robots (and those controllers) within the range of the IR emitter will receive the message.

3.2 The Controller

Whereas the robot class is the physical and geometrical representation of the robot, the controller is the software representation. It is designed to be run in the simulation

as well as on the micro controller without changes in the source code. This will be achieved through FreeRTOS[7], which is a portable open source mini Real Time Kernel. The SymbricatorOS [1], the operating system running on the Symbrion and Replicator robots which is built on FreeRTOS will also run in the simulation so that every program designed for the SymbricatorOS will run on every microcontroller as well as in the simulation. The controller are not required to run only within the simulation, but can also run on the microcontroller which can be connected to the simulation via a serial interface or a TCP/IP interface. This way we can test controllers throughfully if they fit into the micro controller constraints. A third option is not to exchange the whole program code, between simulation and real robot, but a behavior plan, in which actions accordingly to the sensor input are described. Therefor we use the MDL2ϵ, a motion description language[9], that has already been implemented on different platforms and tested successfully for different robot architectures. The MDL2ϵ parser has been integrated into the simulation environment and can now be used further, e.g. for bio-inspired experiments such as offline learning or evolution algorithms.

3.3 Simulation Components and Distributed Simulation

Simulation components are responsible for contolling and observing the simulation and its elements. They can be used for logging robot behaviors, evaluating controllers, resetting the simulation or even simulating environmental variables or other phenomena. Controllers and simulation components inherit the Delta-3D GMComponent class which has been implemented thread-safe and which can also be run distributed on other computer platforms. For distributed simulation we use the High Level Architecture standard which allows to combine several simulations into one simulation. Each simulation, in HLA terms called federate interacts with other federates through a Runtime Infrastructure (RTI). The RTI is the connection between each simulation. We propose the following structure for a distributed simulation of modular robots. A highly GPU powered computer runs the dynamic simulation of the scene including all the physical robot models. If robots or other objects do not need to be simulated dynamically, a kinematic or a fast 2D agent simulator takes the task. Several processors simulate sensor and controller behavior and additionally some environmental simulations can be used to simulate phenomena such as temperature or humidity dispersion and diffusion. Each federate can be run distributed. The RTI makes sure that the information which is needed by other federates will be provided in time and synchronizes the federates regularly.

4 Results and Future Prospects

The recent version of Symbricator3D is a fully functional application that models the Replicator and Symbrion robots accurately using Open Dynamics Engine for dynamics simulation, and Open Scene Graph for the graphical display. The underlying framework is the game and simulation engine ”Delta-3D”, which also supports the HLA standard

[7] www.freertos.org

for providing distributed simulation. The dynamics engine works properly and is stable for small to medium sized organisms. The models representing the Symbricator and Replicator robots are the CAD designs of these robots, and therefor are identical to the appearance of the actual robots. Symbricator3D provides, beside the keyboard control, a graphical user interface to control the simulation as well as the single robots. The robots can be controlled manually or by controllers which can be run within the simulation or externally.

As the simulator is accurate and stable enough, genetic algorithms can already evolve miscellaneous behaviors for different organism configurations, which is an important step for the development of new bio-inspired algorithms, a major goal of the SYMBRION and REPLICATOR projects. First behaviors for single robots have been evolved [1] as well as for organism with few robot modules [10]. Furthermore, different larger organisms consisting of up to 30 joints have been implemented and simulated in Symbricator3D. In Figure 4 the gait of two organisms, a caterpillar and a wheel are depicted. The gait of the caterpillar is a wave function based on the movements of its biological counterpart, while the movement of the wheel is based on moving the center point of gravity of the organism by a coordinate movement of the hinges of each robot. Simulating larger organisms however becomes instable and inaccurate as the Open Dynamics Engine, our current dynamics simulation, has been turned out not to be suitable for dynamic simulation of closed-loop kinematic chains with many degrees of freedom. Therefore, we plan to change the dynamic simulation to either the Nvidia PhysX engine, which supports hardware acceleration on the GPU, or the impulse based dynamics engine IBDS, whose algorithm for dynamics simulation is faster and much more stable than standard methods [11].

If robots are decoupled from eachother, i.e. if they are in swarm mode, Symbricator3D can simulate the dynamic behaviors of up to 100 robots in real time. As for single robot simulation however the dynamics simulation is often not necessary we plan to implement a 2D simulator which will simulate their behavior using less computational algorithms which do not simulate the dynamics of the robots. This way much more robots can be simulated. A first version of a 2D simulator is currently under development which can

(a) Movement of a Wheel-like Organism

(b) Movement of a Caterpillar Organism

Fig. 4. Organism Movements: (a) Wheel-like Movement: By bending the hinge of the bottom-left robot in the organism, the upper robots move to the right and therefore the organism also moves to the right, (b) Caterpillar Movement: Several robots are docked together in a row and move only their hinges to generate a forward caterpillar-like motion.

simulate up to 500 robots. Nethertheless, sometimes even swarm robots need to be simulated dynamically, e.g. in case of rough terrain or during the coupling sequence. As in swarm mode the dynamic behaviors of the robots are independent of eachother, the dynamics can be calculated in parallel. Hence, we expect a performance increase when using dynamics engines with hardware acceleration support such as the Nvidia PhysX engine which we will examine in the future.

5 Conclusion and Future Work

In this paper we introduced the Symbricator3D simulation environment, a new simulator for modular robots which covers all important aspects for simulating modular robots. The mechanisms as well as the implementations of the particular elements of the Symbrion and Replicator robots have been described. Additionally, to get a fast and stable simulation, we have analyzed how to obtain a distributed simulation and decided to use the High Level Architecture standard to distribute the simulation throughout a network.

In the future we intend to evaluate several other dynamic simulations to get a faster and more stable simulation and we plan to implement further the distribute simulation concept. Our aim is a simulation capable of simulating an organism with up to 100 robots and swarms with up to 500 robots. To reach that goal, we intend to simulate robots only dynamically when necessary and kinematically or with a 2D-simulator otherwise, so that we are able to simulate much more complex organisms.

Acknowledgment

The 'SYMBRION' project is funded by the European Commission within the work program 'Future and Emergent Technologies, Proactive' under the grant agreement no. 216342. The 'REPLICATOR' project is funded within the work program 'Cognitive Systems, Interaction, Robotics' under the grant agreement no. 216240. We want to thank all members of the project for their cooperation and their fruitful discussion.

References

1. Kernbach, S., Meister, E., Schlachter, F., Jebens, K., Szymanski, M., Liedke, J., Laneri, D., Winkler, L., Schmickl, T., Thenius, R., Corradi, P., Ricotti, L.: Symbiotic robot organisms: Replicator and symbrion projects. In: PerMIS 2008, Gaithersburg, MD, USA (2008)
2. Schlachter, F., Meister, E., Kernbach, S., Levi, P.: Evolve-ability of the robot platform in the symbrion project. In: Workshop on Pervasive Adaptive Systems, Second IEEE International Conference on Self-Adaptive and Self-Organizing Systems, Venice (2008)
3. Kernbach, S., Ricotti, L., Liedke, J., Corradi, P., Rothermel, M.: Study of Macroscopic Morphological Features of Symbiotic Robotic Organisms. In: Proceedings of the workshop on self-reconfigurable robots, IROS 2008, Nice, pp. 18–25 (2008)
4. Gerkey, B., Vaughan, R., Howard, A.: The player/stage project: Tools for multi-robot and distributed sensor systems. In: Proceedings of the 11th International Conf. on Advanced Robotics, pp. 317–323. IEEE Press, Piscataway (2003)

5. Dahmann, J.S.: Standards for simulation: As simple as possible but not simpler the high level architecture for simulation. Simulation, 6 71(6), 378–387 (1998)
6. Fullford, D.: Distributed interactive simulation: it's past, present, and future. In: Proceedings of Simulation Conference, 1996, pp. 179–185 (Winter 1996)
7. Darken, R., McDowell, P., Johnson, E.: The delta3d open source game engine. IEEE Computer Graphics and Applications 25(3), 10–12 (2005)
8. Bender, J.: Impulse-based dynamic simulation in linear time. Computer Animation and Virtual Worlds 18(4-5), 225–233 (2007)
9. Szymanski, M., Wörn, H.: Jamos - a mdl2e based operating system for swarm micro robotics. In: IEEE Swarm Intelligence Symposium, pp. 324–331 (2007)
10. Schmickl, T., Stradner, J., Hamann, H., Crailsheim, K.: Major feedbacks that support artificial evolution in multi-modular robotics. In: Exploring New Horizons in Evolutionary Design of Robots (EvoDeRob) – IROS 2009 workshop. LNCS. Springer, Heidelberg (2009)
11. Schmitt, A., Bender, J.: Impulse-based dynamic simulation of multibody systems: Numerical comparison with standard methods. In: Proc. Automation of Discrete Production Engineering, pp. 324–329 (2005)

Mono Landmark Localization for an Autonomous Navigation of a Cooperative Mobile Robot Formation

Hugues Sert[1,2], Annemarie Kökösy[1,2], Jorge Palos[1], and Wilfrid Perruquetti[2,3]

[1] ISEN, 41 bvd Vauban, 59 046 Lille Cedex, France
[2] LAGIS UMR CNRS 8146, BP 48, Cité Scientifique, 59651 Villeneuve-d'Ascq, Cedex, France
hugues.sert@isen.fr, annemarie.kokosy@isen.fr, jorge.palos@isen.fr
[3] ALIEN, INRIA Lille - Nord Europe Parc Scientifique de la Haute Borne 40, avenue Halley Bt.A, Park Plaza 59650 Villeneuve d'Ascq, France*
wilfrid.perruquetti@inria.fr

Abstract. In this paper the problem of the autonomous navigation of a cooperative robot formation is studied. Robots navigate in an unknown environment, encumbered by obstacles which have a circular shape or which can be modeled by polygons. Robots must move between an initial known position and a final known position. While going ahead robots discover the obstacles which block their passage and must avoid the collision and go to their final destination. This paper tackles two problems:

1. path planning, which is described as an optimal control problem under constraints (robot physical limits, collision avoidance).
2. localization, in order to plan its trajectory, the robot has to know its localization in the environment : a mono-landmark algorithm of localization in 2D is proposed. This mono-landmark algorithm tracks a landmark during the navigation and uses it to localize the robot in an algebraic framework.

Keywords: Cooperative robot formation, decentralized intelligence, path planning, localization mono landmark algorithm, obstacles avoidance, nonholonomic mobile robots, deterministic global localization, mono-camera localization.

1 Introduction

The applications of the multi-robot cooperation can be found in the military, civil and spatial domains. It concerns missions which were automated historically, either because they represent a danger or a certain difficulty for a human, or because a machine is more adapted and more effective than humans. Cooperative

* This work was partially supported by the Catholic University of Lille, France, under Entités Mobiles project and by INRIA–ALIEN.

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 1278–1292, 2009.

robotics, strongly imbricated with service robotics, concerns a lot of applications in which the researchers are particularly interested:

1. The carriage of loads and the collection of objects in dangerous zones for humans.
2. The use of cooperative machines on construction sites could accelerate the speed of the works and allow impracticable operations for the human.
3. The detection and the coordinated neutralization of land mines. The cooperation would allow to cover the widest zone without many more operators, while reducing the duration of the operation.
4. The security/surveillance of properties: the surveillance and the detection of intervention are tasks that robots can typically carry out. The cooperation allows the surveillance of a larger perimeter.

In order to be able to navigate autonomously, the robot must know where it is in the environment, what kind of obstacles are in this environment and how it can avoid the collision with them to go to its final goal. Figure 1 presents the general architecture of a mobile robot. The **sensors** of the robot are proprioceptive (wheel encoders, gyroscopes, inertial navigation) or exteroceptive (cameras, compass, GPS, laser range-finders). The measurements obtained from its **sensors**, are processed in blocks **"perception"** and **"localization"** to obtain information on the robot localization (position and orientation) and on the environment in which it navigates. The information of blocks **"perception"** and **"localization"** is sent to the block **"strategy"** which has to estimate the situation in which the robot is, and, if necessary, to ask the information from its neighbors to make a decision on the direction of the robot movement. This decision can be, for example, the coordinates of an intermediate objective to achieve (in the case where the robot is in a local minima position), or new constraints which the planned trajectory has to satisfy (for example switching to an

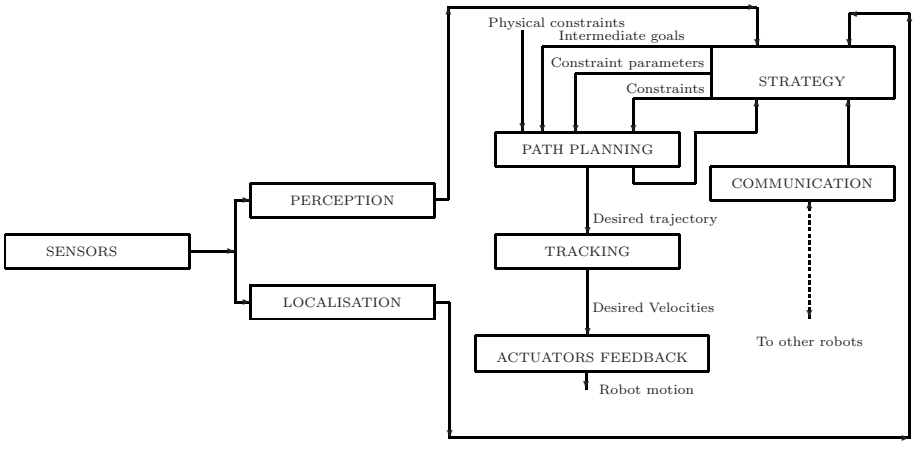


Fig. 1. General architecture for a mobile robot autonomous navigation

energy saving mode if the block **"strategy"** detects that the robot energy level is low and that it is really necessary that the robot achieves its mission). The block **"strategy"** can also only decide to change the value of the parameters of certain constraints.

The block **"path planning"** calculates the optimal trajectory of the robot which satisfies the cost function and the constraints, whereas the block **"tracking"** calculates the desired velocities of the robot (speeds to be applied to wheels) to guarantee that the robot will track the optimal trajectory. The block **"actuators feedback"** guarantees that the desired calculated speeds will be equal to the speeds of the robot wheels. The block **"communication"** ensures information exchanges between the blocks of the robot and with the other robots of the formation.

This paper focuses on blocks **"path planning"** and **"localization"**. Section 2 presents the mobile robot modelisation used in this paper. Section 3 deals with the main results of this research work. In section 3.1 the path planning algorithm will be explained. This part is a synthesis of previous work. In section 3.2 the contribution of this paper will be explained. It concerns the robot localization by using a unique landmark. Originalities of this localization algorithm are:

- The use of only one landmark tracking during the navigation,
- No hypothesis on the nature of the signal is necessary,
- Implementation in real time is possible,
- No step of initialization is necessary.

A conclusion and some discussions on further works are done in section 4.

2 Mobile Robot Modelisation

Consider a multi-robot system composed of N_a wheeled mobile robots. Each mobile robot is of unicycle type, with two driving wheels mounted on the same axis and independently controlled by two actuators (DC motors). The geometric shape of each robot R_i belonging to a mobile robot formation is represented by a 2-D circle of center $C_i = (x_i, y_i)$ and of radius ρ_i. The $i-th$ robot is fully described by a three dimensional vector of generalized coordinates q_i constituted by the coordinates (x_i, y_i) of the midpoint between the two driving wheels and by the orientation angle θ_i with respect to a fixed frame:

$$q_i = [x_i, y_i, \theta_i]^T. \tag{1}$$

Under the hypothesis of *pure rolling and non slipping condition*, the vehicle satisfies the nonholonomic constraint:

$$\left[-\sin\theta_i \ \cos\theta_i \ 0 \right] \dot{\mathbf{q}}_i = 0$$

and the ideal kinematic equations are:

$$\dot{\mathbf{q}}_i = \begin{bmatrix} \cos\theta_i & 0 \\ \sin\theta_i & 0 \\ 0 & 1 \end{bmatrix} u_i \tag{2}$$

where $u_i = [v_i \ w_i]^T$ and v_i and w_i are the linear and angular velocities.

Robots have on-board sensors which enable them to detect surrounding obstacles and vehicles within a range. This range is described by a circle centered at C_i. The broadcasting range, $d_{i,com}(> 0)$, is limited.

3 Main Results

3.1 Path Planning

In this paper, the problem of interest is the navigation of cooperative nonholonomic mobile robots evolving in environments with obstacles. Cooperative approaches for path planning can be broken down into centralized and decentralized (distributed) methods. The centralized approach is usually based on the resolution of an optimal control problem via game theoretical framework [1], semi-definite programming [2] or non-linear programming [3]. However, its computation time which scales exponentially with the number of vehicles, its communication requirement and its lack of security make it prohibitive. This is true even when the most advanced optimization solvers and much simplified linear vehicle dynamics are used [4]. The main challenge is to formulate a simple decentralized strategy which results in a team behavior similar to what is obtained with a centralized approach. Some decentralized algorithms are based on a receding horizon approach [5], [6], [7], [8,9]. Other decentralized strategies, which can be computationally fast, include rule-based approaches [10], probabilistic approaches [11,12] and artificial potential fields [13,14,15]. However, most of these techniques are successfully developed for oversimplified vehicle dynamics without explicitly taking the constraints into account. In this paper a decentralized path planning will be proposed which takes the robot kinematic model and its physical limitations into account.

The goal of the robots is to navigate without collision in an unknown environment with obstacles. Only the initial and final positions of each robot are known. The robots must maintain a communication between them in order to be able to exchange information about their positions and intentions. It is possible to express this goal as an optimization problem under constraints.

3.1.1 Planification Problem Formulation

Find the optimal control $u_i(t)$ and the optimal trajectory $q_i(t)$ of each robot which minimize the cost function:

$$J = \int_{t_0}^{t_f} L_i(q_i, u_i, t)dt \tag{3}$$

where t_0 and t_f are the initial and the final times respectively, L_i is the cost function. For example it is possible to minimize the energy level so in this case $L_i = u^2$. This control musst also satisfy the following constraints $\forall t \in [t_0, t_f]$:

C_1 *the optimal trajectory and the optimal control are solutions of the Kinematic model of robot R_i (2).*

C_2 *the control bounds:*

$$|v_i| \leq v_{i_{max}}, \quad |w_i| \leq w_{i_{max}},$$

where $v_{i_{max}}$ *and* $w_{i_{max}}$ *are the maximal values of the linear and angular velocities of robot* R_i*.*

C_3 *the terminal conditions are respected:*

$$q_i(t_0) = q_i(0),\ q_i(t_f) = q_{i_{final}},\ u_i(t_0) = u_i(0),\ u_i(t_f) = u_{i_{final}}. \tag{4}$$

C_4 *the collision avoidance with all detected obstacles* O_{mi}*, which have a circular shape and belong to the set* $\mathcal{O}_i(\tau_k)$ *of all detected circular obstacles at time* τ_k*:*

$$\forall O_{mi} \in \mathcal{O}_i(\tau_k),\ \ d(R_i, O_{mi}) = \sqrt{(x_i - x^o_{mi})^2 + (y_i - y^o_{mi})^2} \geq \rho_i + r^o_{mi},$$

where $d(R_i, O_{mi})$ *is the euclidian distance between robot* R_i *and obstacle* O_{mi}*, described by its center position* (x^o_{mi}, y^o_{mi}) *and its radius* r^o_{mi}*.*

C_5 *the distance between the robot and the* N_p *polygonal obstacles, where polygonal obstacle* PO_i *is approximated by a succession of* N_i *segments* $\mathbf{S}_{ij}$*. This distancemust obey the following condition:*

$$d(R, PO_i) > \zeta > 0,\ \forall i \in 1, ..., N_p, \tag{5}$$

where $d(R, PO_i)$ *is the distance between the robot and the visible part by the robot sensors of obstacle* PO_i*,*

$$d(R, PO_i) = \min_j d(R, \mathbf{S}_{ij}),\ j \in 1, ..., N_i. \tag{6}$$

and $d(R, \mathbf{S}_{ij})$ *is the distance between the robot and segment* $\mathbf{S}_{ij}$*:*

$$d(R, \mathbf{S}_{ij}) = \min_{M \in S_{ij}} d(R, M) \tag{7}$$

C_6 *the collision avoidance with each other robot* R_j *belonging to the robot swarm:*

$$\forall j \in \{1, \ldots, N_a\},\ i \neq j,\ \ d(R_i, R_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \geq \rho_i + \rho_j.$$

C_7 *the communication between robots belonging to the robot swarm:*

$$\forall j \in \{1, \ldots, N_a\}, i \neq j,\ d(R_i, R_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \geq min(d_{i,com}, d_{j,com}),$$

where $d_{i,com}$ *and* $d_{j,com}$ *are the maximal broadcasting range of robots* R_i *and* R_j *respectively.*

3.1.2 Calculation of an Optimal Trajectory

In order to solve the path planning problem, the flatness property of the system [16] is used, which reduces the computation time. The optimal control problem is transformed into a parameter optimization problem by using the B-spline functions to describe the flat outputs of the robot. As robots do not know in advance the environment in which they evolve, it is not possible to calculate the trajectory for each robot. That is why the receding horizon planning was used. The trajectory of each robot of the formation is calculated at each time interval $[\tau_k, \tau_k + T_c]$, where T_c is the update period. A real-time planning algorithm for a single robot which moves in an environment with obstacles which have a circular shape has been proposed in [17]. A generalization of this algorithm for obstacles which can be included in a polygon has been developed in [18].
For the robot formation, a decentralized algorithm is proposed. It is updated each τ_k, and it is divided into two steps:

1. **Step 1.** Each robot R_i computes its trajectory by only taking the obstacles into account. Constraints C_6 and C_7 are not included in this first step. This trajectory is called **intuitive trajectory**. It is evaluated over the obstacles detection horizon T_d. At the same time, robot R_i analyzes, by using its sensors, the potential inter-robots collision as well as problem of los of communication, in order to know which are the robots with which it has to exchange its trajectory.
2. **Step 2.** The robots which can be in collision or lose the communication with the other robots of the formation adjust their intuitive trajectories by taking constraints C_6 and C_7 into account. This new trajectory is called **planned trajectory** and it is the trajectory that the robot must track at time interval $[\tau_k + T_c, \tau_k + 2T_c[$. This trajectory is evaluated over planning horizon T_p.

The details about the algorithm calculation and implementation can be found in [19].

3.1.3 Experimental Results for the Autonomous Navigation

Robot description. The proposed decentralized motion planning algorithm has been implemented on three mobile Pekee robots manufactured by Wany Robotics. The robot radius is $\rho = 0.25m$. An Intel 486 micro-processor running at $75MHz$ operating under Linux real time hosts the integral sliding mode tracking controller written in C. Pekee is equipped with 15 infra-red telemeters sensors, two encoders, a WiFi wireless cartridge and a miniature color vision camera C-Cam8. The maximal broadcasting range is $d_{com} = 2m$. The vision camera is fixed in the robot coordinate system $(x_c, y_c, z_c) = (0, 0, 0.25)$. The maximum linear and angular speeds are respectively equal to $v_{max} = 0.35$m/s and $w_{max} = 0.8$rad/s.

Algorithm implementation. The initial and final states and controls for the first robot are: $q_1(t_0) = (0.5m, 0m, 0deg)$, $v_1(t_0) = 0m/s$, $w_1(t_0) = 0rad/s$ $q_1(t_f) = (9m, 0m, 90deg)$, $v_1(t_f) = 0m/s$, $w_1(t_f) = 0rad/s$. The second robot

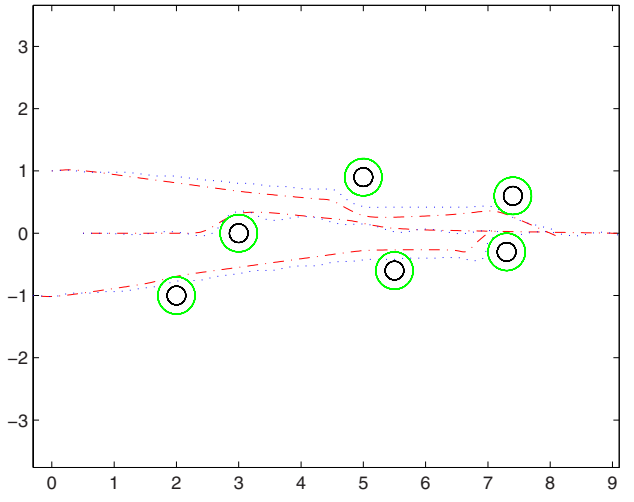


Fig. 2. Resulting trajectories of the three robots

has the following initial and final states and controls: $q_2(t_0) = (0m, 1m, 0deg)$, $v_2(t_0) = 0m/s$, $w_2(t_0) = 0rad/s$ $q_2(t_f) = (8m, 0m, 90deg)$, $v_2(t_f) = 0m/s$, $w_2(t_f) = 0rad/s$. The third robot has the following initial and final states and controls: $q_3(t_0) = (-0.3m, -1m, 0deg)$, $v_3(t_0) = 0m/s$, $w_3(t_0) = 0rad/s$ $q_3(t_f) = (7m, 0m, 0deg)$, $v_3(t_f) = 0m/s$, $w_3(t_f) = 0rad/s$. In this application, only the infra-red sensors are used for the perception of the environment. The values for the update period, the planning horizon and the obstacles detection horizon are chosen as follow: $T_c = 0.5s$, $T_p = 2s$ and $T_d = 2.5s$. The task is to drive these robots from the initial position to the final known position while avoiding collisions and maintaining the communication constraints. Due to the existence of obstacles, the robots must pass through narrow ways and constraint each other in the team. The robot trajectories are depicted in Figure 2. The videos of the experiment can be visualized on the web site `http://oban.isen.fr/~sst_lille/` under the heading *Recherche*.

3.2 Localization

3.2.1 Introduction

Localization is one of the fundamental axes of research in mobile robotics because it is not possible to speak about autonomy for a robot without speaking about localization. So the question "where am I?" [20] must be resolved. A localization via proprioceptive sensors (wheel encoders, gyroscopes, inertial navigation) is possible (call relative localization) but presents problems of long-term drift [21]. An absolute localization via the exteroceptive (cameras, compass, GPS, laser range-finders) allows to avoid this drift. In this part The new algorithm is proposed within the framework of the absolute landmark localization.

There are two types of landmarks, active and passive. Passive landmarks could be natural (their existence is independent of the robot localization) or artificial. Active landmarks need great emission power and may be not discrete enough in an environment where electromagnetic noise is annoying.

Two types of methods are possible for the localization by landmarks, the first one is deterministic and the second one is probabilistic. Deterministic methods are generally based on the triangulation method, which needs, at least, three landmarks at each localization time. Some probabilistic methods could be found in [22]. However, probabilistic methods use an hypothesis on the nature of the noise which affects the signal.

In the deterministic framework it is necessary to have at least three landmarks in order to localize the robot, which could be a problem, for example, in some very structural environments were landmarks are not distinguishable. This paper presents a new method under the algebraic framework rooted from ALIEN project with only one landmark, which is tracking during the navigation. In opposition with the statistical method, this method proposed herein does not use information on the signal nature, in particular on the noise which affects it. Furthermore no initialization step is necessary. So, with the use of the observability property of the angular measurements, this paper presents a new method of localization.

In the following, the principle of the localization method is presented, then the paper deals with the chosen method of numerical differentiation in a noisy environment. Then a resolution in the two dimensional case is done and a synthesis under a theorem formis given. The novelty of this part is in the principle and in the resolution of the two dimensional case. The chosen method of integration comes from [23] and is adapted to the localization problem.

3.2.2 Principle

Let be a state model system described by the following equation:

$$\begin{cases} \dot{q_i} = F(q_i, u_i) \\ z = H(q_i) \end{cases} \tag{8}$$

with q_i the state vector, u_i the command vector and z the measurement vector. The proposed method of localisation consists in finding the vector function $K(z, \dot{z}, \ddot{z}, ..., z^{(n)})$ such as:

$$q_i = K(z, \dot{z}, \ddot{z}, ..., z^{(n)}) \tag{9}$$

This is possible if the system is observable for the vector of measurement z (see [24] and [25] for complements on the observability of non linear systems). Finding the formal expression of K allows to prove the observability of the system, indeed the criterion of Fliess ([26]) expresses that any variable of an observable system satisfies an algebraic equation linking the output, the input and a finite number of their derivatives. With the knowledge of a formal expression of K and an estimation of the successive derivative of z, it is then possible to calculate an estimation of state q_i for any time t.

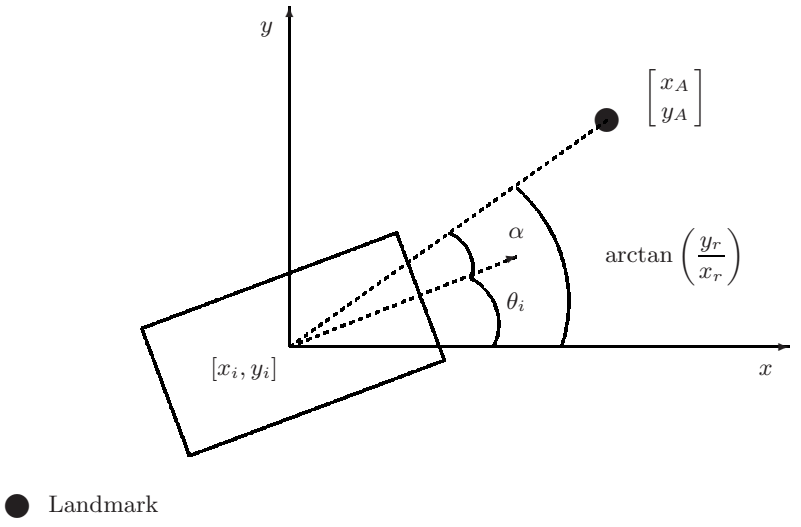


Fig. 3. Robot and landmark notation for the localisation

The calculation of an estimation of state q_i is then obtained in two steps:

1. The calculation of an estimator of the successive derivatives of z (see 3.2.3),
2. The calculation of the expression of the vectorial function K (see 3.2.4).

In the following, this method is described for the kinematics model of the robot (see equation (2)). The robot measures the angle between itself and a landmark, so the measurement model is (see figure 3):

$$z = H(q_i) = \alpha = \arctan\left(\frac{y_r}{x_r}\right) - \theta_i + \xi_\alpha \tag{10}$$

where x_r and y_r are the relative position between the landmark and the robot such as : $x_r = x_A - x_i$, $y_r = y_A - y_i$ with (x_A, y_A) the landmark coordinate and ξ_α an additive noise.

3.2.3 Numerical Differentiation in a Noisy Environment

This algebraic setting for numerical differentiation of noisy signals was introduced in [27] and analysed in [28,23] (see also [29] for interesting discussions and comparisons). The reader may find additional theoretical foundations in [30,31]. Let us note $z(t) = \sigma(t) + \xi(t)$, with $\sigma(t)$ the signal without noise that must be diffentiated at the n^{th} order and $\xi(t)$ an additive noise measurement.

By considering the analytical expression of the signal $\sigma(t)$, it is possible to study, without loss of generality, the convergent Taylor series at $t = 0$

$$\sigma(t) = \sum_{i \geq 0} \sigma^i(0) \frac{t^i}{i!}$$

The truncated Taylor series for $N \geq n$ is:

$$\sigma_N(t) = \sum_{i=0}^{N} \sigma^i(0) \frac{t^i}{i!}$$

which satisfies the differential equation $\frac{d^{N+1}}{dt^{N+1}} \sigma_N(t) = 0$. In the Laplace domain this equation can be written:

$$s^{N+1} \hat{\Gamma}_N(s) = s^N \sigma(0) + s^{N-1} \dot{\sigma}(0) + ... + s^{N-n} \sigma^{(n)}(0) + ... + \sigma^{(N)}(0)$$

where $\hat{\Gamma}(s)$ is the Laplace transform of $\sigma(t)$. A series of algebraic manipulations allows then to obtain an expression of $\sigma^{(n)}(0)$, for every $n \in [0, N]$, by annihiling successively each of the terms $s^{N-i} \sigma^{(i)}(0)$, $i \in [0, N], i \neq n$.

The general expression of the derivative can be declined in continuous version (no error of digital estimate due to the chosen method of integration) and in discrete version (for more details see [23] and [32]). This article presents here both results. The continuous version of the n^{th} derivative is:

$$\sigma^{(n)}(Tt; \kappa; \mu; N) = \int_0^1 g(\tau, \kappa, \mu, N) z_T(t - \tau) d\tau \tag{11}$$

The n^{th} derivative via the discrete approach is obtained as the output of the finite impulse response filter:

$$\sigma^{(n)}(lT_s; \kappa; \mu; N) \approx \sum_{m=0}^{M} W_m g_m(\kappa, \mu, N) z_{l-m} \tag{12}$$

Where $z_m = z(mTs)$, T_s is the sampling period, M is the number of coefficients of the filter, $T = MT_s$ is the first moment of estimation and$c_m = W_m g_m, m = 0, ..., M-1$ is its impulse response. The signal for which we wan to estimate the derivative, is noisy and sampled : it is the input of the filter. The coefficients W_m correspond to a method of integration of a signal between 0 and 1. The trapezoidal method gives the following coefficients:

$$\begin{aligned} W_0 = W_M &= \frac{T_s}{2} \\ W_m &= T_s, m = 1, ..., M-1 \end{aligned} \tag{13}$$

The coefficients g_m are such as $g_m = g(mT_s/T), m = 0, ..., M$, with $\tau = \frac{MT_s}{T}$ where

$$g(\tau) = \sum_{l=0}^{q} \lambda_l h_{\kappa+q-l, \mu+l}(\tau), (\kappa; \mu) \in \mathbb{N}, q = N - n \tag{14}$$

with

$$h_{\kappa,\mu}(\tau) = \frac{(-1)^n \gamma_{\kappa,\mu,n}}{T^n} rect(\tau) \frac{d^n}{d\tau^n} \omega_{\kappa,\mu}(\tau) \tag{15}$$

with

$$rect(\tau) = \begin{cases} 1 \text{ if } \tau \in [0;1] \\ 0 \text{ orthewise} \end{cases} \tag{16}$$

$$\omega_{\kappa,\mu}(t) = t^{\kappa+n}(1-t)^{\mu+n} \tag{17}$$

and

$$\gamma_{\kappa,\mu,n} = \frac{(\mu+\kappa+2n+1)!}{(\mu+n)!(\kappa+n)!} \tag{18}$$

By noting $q = N - n$ and $p = n + \kappa$

$$\lambda_l = (-1)^{q-l} \binom{p+q-l}{p} \binom{p+q+1}{l}, l = 0, ..., q \tag{19}$$

- N allows to complexify the model which tends to reduce the delay of estimation but increases the influence of the measurement noise,
- M is the number of coefficients of the filter. Increasing M is going to reduce the noise of measurement, but increases the noise of modeling error. If $N = n$ then $M \in [60; 80]$, otherwise $M \in [80; 110]$,
- T_s is the sampling period of the signal which, in most cases, cannot be chosen,
- κ corresponds to the number of differentiation in the operational domain generally fixed to 0,
- μ corresponds to the number of integration in the operational domain generally fixed to 0 or 1 which allows to reduce the delay of estimation.

3.2.4 Resolution in the Two Dimensional Case

Theorem 1. *For a non-holonomic robot which has the model described by equation (2) and under the assumption that the model of measurement is defined by equation (10), and that the following data are measured or known:*

- (x_A, y_A, z_A)*, the coordinate of the landmark,*
- v_i*, the linear speed of the robot (possibly noisy),*
- ω_i*, the angular speed of the robot (possibly noisy),*
- θ_i*, its orientation (measured with a compass for example, possibly noisy),*
- α*, angle of the robot with regard to the landmark,*

the position of the robot could be calculated by using the following equation :

$$\begin{bmatrix} x_r \\ y_r \end{bmatrix} = \begin{bmatrix} x_A - x_i \\ y_A - y_i \end{bmatrix} = \begin{bmatrix} \dfrac{u_i \sin(\alpha)\cos(\alpha+\theta_i)}{(\dot{\alpha}+\dot{\omega_i})} \\ \dfrac{u_i \sin(\alpha)\sin(\alpha+\theta_i)}{(\dot{\alpha}+\dot{\omega_i})} \end{bmatrix} \tag{20}$$

where u_i, ω_i, θ_i and α are filtered by using a zero order differentiator described by equations (12)–(19) (with $n = 0$ and $N = 0$) and $\dot{\alpha}$ is calculated by using a first order differentiator (with $n = 1$ and $N = 1$), $M = 80$, $T_s = \frac{1}{25} sec$, $\kappa = \mu = 0$.

The proof is omitted due to the paper limitation.

3.2.5 Simulation Results

The simulations has been implemented on Matlab. The starting points is $[0; 0]$ and the simulation lasts $40sec$. The linear velocity is $0.5m/s$ and the angular velocity is $0rad/s$, so $\theta_i = 0$ during the navigation process. α is calculated with

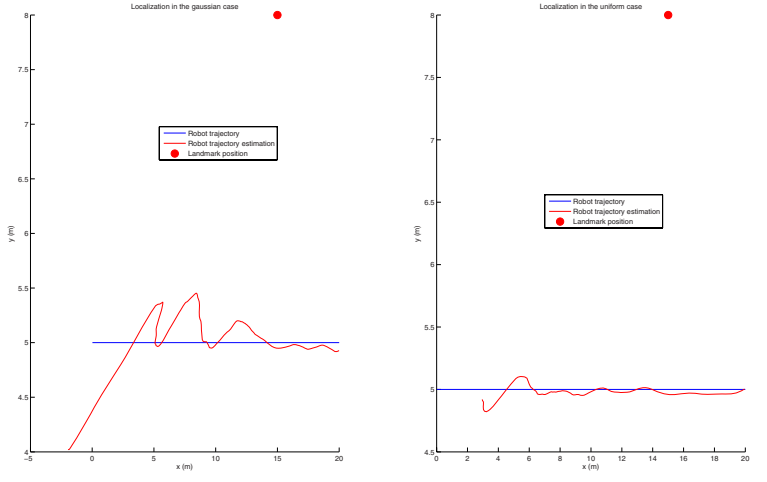


Fig. 4. True trajectory and estimated trajectory

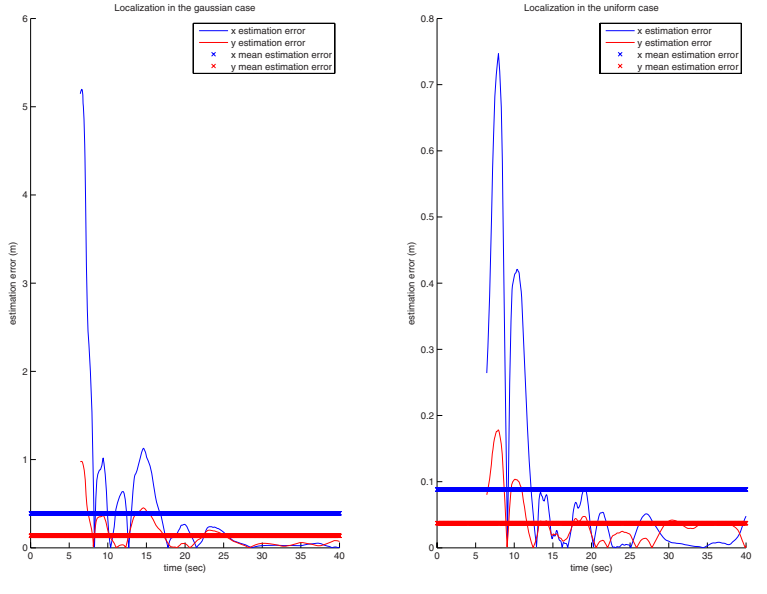


Fig. 5. Estimation error of the localization process

an additive noise. In the first case it is a Gaussian noise and in the second case it is an uniform noise. The landmark is situated in $[9; 5]$. The frequency is $25 frames/sec$ so $T_s = 1/25 sec$. Figure 4 shows the true trajectory of the robot and the estimate trajectory in the gaussian case and in the uniform case. Figure 5 shows the estimation error in both cases. Simulation results show that the method is not sensitive of the noise and gives good results on the robot localization.

4 Conclusion

In this paper a descentralized motion planner for a cooperative robot formation has been presented. The trajectory of each robot was planned by taking into account the maximal linear and angular velocities of the robot and its kinematic model. The planner guarantees a collision free trajectory. Moreover, by using a descending horizon, and the constrained feasible sequential quadratic optimization algorithm this planner can be implemented in real time. Three nonholonomic robots are used for the experiments. The experimental results show the effectiveness of this approach. In order to plan its trajectory the robot needs to know where it is in its environment. A new localization algorithm has been proposed by using a unique landmark. The robot position (x, y) is obtained by using only information about the angle between the robot and the landmark and the robot orientation. The robot also needs information about its angular and linear velocities. The algorithm requires no statistical hypothesis on the measurement noise. The simulation results show that the algorithm is not influenced by the nature of the measurement noise.

References

1. Harmati, I., Skrzypczyk, K.: Robot team coordination for target tracking using fuzzy logic controller in game theoretic framework. Robotics and Autonomous Systems 57(1), 75–86 (2009)
2. Frazzoli, E., Mao, Z.H., Oh, J.H., Feron, E.: Resolution of conflicts involving many aircrafts via semidefinite programming. Journal of Guidance, Control, and Dynamics 24(1), 79–86 (2001)
3. Raghunathan, A., Gopal, V., Subramanian, D., Biegler, L., Samad, T.: 3D conflict resolution of multiple aircraft via dynamic optimization. In: AIAA Guidance, Navigation, and Control Conference and Exhibit, Austin, Texas, USA (August 2003)
4. Borrelli, F., Subramanian, D., Raghunathan, A., Biegler, L.: MILP and NLP techniques for centralized trajectory planning of multiple unmanned air vehicles. In: American Control Conference, 2006, Minneapolis, Minnesota USA (June 2006)
5. Dunbar, W., Murray, R.: Distributed receding horizon control for multi-vehicle formation stabilization. Automatica 42(4), 549–558 (2006)
6. Keviczky, T., Borrelli, F., Balas, G.: Decentralized receding horizon control for large scale dynamically decoupled systems. Automatica 42(12), 2105–2115 (2006)
7. Keviczky, T., Borrelli, F., Fregene, K., Godbole, D., Balas, G.: Decentralized receding horizon control and coordination of autonomous vehicle formations. IEEE Transactions on Control Systems Technology 16(1), 19 (2008)

8. Schouwenaars, T., How, J., Feron, E.: Decentralized cooperative trajectory planning of multiple aircraft with hard safety guarantees. In: Proceedings of the AIAA Guidance, Navigation, and Control Conference, Providence, Rhode Island (August 2004)
9. Richards, A., How, J.: Decentralized model predictive control of cooperating UAVs. In: 43rd IEEE Conference on Decision and Control, 2004. CDC, Paradise Island, Bahamas (December 2004)
10. Bennewitz, M., Burgard, W., Thrun, S.: Finding and optimizing solvable priority schemes for decoupled path planning techniques for teams of mobile robots. Robotics and Autonomous Systems 41(2-3), 89–99 (2002)
11. LaValle, S., Hutchinson, S.: Optimal motion planning for multiple robots having independent goals. IEEE Transactions on Robotics and Automation 14(6), 912–925 (1998)
12. Sanchez, G., Latombe, J.: Using a prm planner to compare centralized and decoupled planning for multi-robot systems. In: Proceedings of IEEE International Conference on Robotics and Automation, 2002, Washington, DC, USA (May 2002)
13. Dimarogonas, D., Kyriakopoulos, K.: Decentralized navigation functions for multiple robotic agents with limited sensing capabilities. Journal of Intelligent and Robotic Systems 48(3), 411–433 (2007)
14. Pereira, G., Das, A., Kumar, V., Campos, M.: Decentralized motion planning for multiple robots subject to sensing and communication constraints. Multi-Robot Systems: From Swarms to Intelligent Automata 2, 267–278 (2003)
15. Pereira, G., Kumar, V., Campos, M.: Closed loop motion planning of cooperating mobile robots using graph connectivity. Robotics and Autonomous Systems 56(4), 373–384 (2008)
16. Fliess, M., Levine, J., Martin, P., Rouchon, P.: Flatness and defect of non-linear systems: introductory theory and examples. International journal of control 61(6), 1327–1361 (1995)
17. Defoort, M., Palos, J., Kokosy, A., Floquet, T., Perruquetti, W.: Performance-based reactive navigation for non-holonomic mobile robots. Robotica 27(02), 281–290 (2008)
18. Kokosy, A., Defaux, F., Perruquetti, W.: Autonomous navigation of a nonholonomic mobile robot in a complex environment. In: IEEE International Workshop on Safety, Security and Rescue Robotics, 2008. SSRR 2008, Sendai, Japan (October 2008)
19. Defoort, M., Kokosy, A., Floquet, T., Perruquetti, W., Palos, J.: Motion planning for cooperative unicycle-type mobile robots with limited sensing ranges: A distributed receding horizon approach. Robotics and Autonomous Systems (2009)
20. Borenstein, J., Everett, H., Feng, L.: Where am I? Sensors and methods for mobile robot positioning. University of Michigan (1996)
21. Borenstein, J.: Control and kinematic design of multi-degree-of freedom mobile robots with compliant linkage. IEEE Transactions on Robotics and Automation 11(1), 21–35 (1995)
22. Negenborn, R.: Robot localization and Kalman filters. PhD thesis, Utrecht University, Netherlands (2003)
23. Mboup, M., Join, C., Fliess, M.: Numerical differentiation with annihilators in noisy environment. Numerical Algorithms 50, 439–467 (2009)
24. Diop, S., Fliess, M.: Nonlinear observability, identifiability, and persistent trajectories. In: Proceedings of the 30th IEEE Conference on Decision and Control, Brighton, England (December 1991)

25. Leiva, H., Siegmund, S.: A necessary algebraic condition for controllability and observability of linear time-varying systems. IEEE Transactions on Automatic Control 48(12), 2229–2232 (2003)
26. Daafouz, J., Fliess, M., Martin, P., Rouchon, P.: Une approche intrinseque des observateurs lineaires a entrees inconnues. CIFA (2006)
27. Fliess, M., Join, C., Mboup, M., Sira-Ramírez, H.: Compression différentielle de transitoires bruités. C.R. Acad. Sci. 339, 821–826 (2004)
28. Mboup, M., Join, C., Fliess, M.: A revised look at numerical differentiation with an application to nonlinear feedback control. In: The 15th Mediterrean Conference on Control and Automation - MED 2007, Athènes Grèce (June 2007)
29. Nöthen, C.: Beiträge zur rekonstruktion nicht direkt gemessener größen bei der silizium-einkristallzüchtung nach dem czochralski-verfahren. Diplomarbeit, Technische Universität Dresden (2007)
30. Fliess, M.: Analyse non standard du bruit. C.R. Acad. Sci. I(342), 797–802 (2006)
31. Fliess, M., Sira-Ramírez, H.: An algebraic framework for linear identification. ESAIM Control Optim. Calc. Variat. 9, 151–168 (2003)
32. Liu, D.: Methodes algebriques d'estimation de parametres pour les systemes dynamiques. Master's thesis, Universite des sciences et technologies de Lille, France (2007)

Author Index