Ngoc Thanh Nguyen
Aleksander Zgrzywa
Andrzej Czyzewski (Eds.)

Advances in Multimedia and Network Information Sytem Technologies

Springer

Advances in Intelligent and Soft Computing 80

Editor-in-Chief: J. Kacprzyk

Advances in Intelligent and Soft Computing

Editor-in-Chief

Prof. Janusz Kacprzyk
Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6
01-447 Warsaw
Poland
E-mail: kacprzyk@ibspan.waw.pl

Further volumes of this series can be found on our homepage: springer.com

Ngoc Thanh Nguyen, Aleksander Zgrzywa,
and Andrzej Czyżewski (Eds.)

Advances in Multimedia and Network Information System Technologies

Springer

Editors

Prof. Ngoc Thanh Nguyen
Institute of Informatics
Wroclaw University of Technology
Wyb. Wyspianskiego 27
50-370 Wroclaw
Poland
E-mail: Ngoc-Thanh.Nguyen@pwr.wroc.pl

Prof. Andrzej Czyżewski
Multimedia Systems Department
Gdansk University of Technology
ul. Narutowicza 11/12
80-233 Gdansk
Poland
E-mail: ac@pg.gda.pl

Prof. Aleksander Zgrzywa
Institute of Informatics
Wroclaw University of Technology
Wyb. Wyspianskiego 27
50-370 Wroclaw
Poland
E-mail: Aleksander.Zgrzywa@pwr.wroc.pl

Preface

The growth of knowledge, unparalleled in the history of the human race, results in the rapid development of technology. The solutions that until quite recently remained in the domain of science-fiction now become a part of our everyday life. Information systems and their technologies enter all the spheres of human's existence. Their influence is multiplied by network connections and by multimedia presentations and communications. As a result, no matter whether we like it or not, irrespectively of all pros and cons, the Multimedia and Network Information Systems shape our future.

These are the roots of the ever increasing interest in the current research and developments in the domain of multimedia and network information systems technologies. Our intention was to offer to the readers of this monograph a very broad review of the recent scientific problems in that area. Searching for their solutions had became a principal task of numerous scientific teams all over the world. In the book we have gathered and presented carefully selected the most representative – in our opinion - investigations, solutions and applications submitted by different scientific teams working in nine European countries.

Content of the book has been divided into five parts:

1. Multimedia information technology
2. Data processing in information systems
3. Information system applications
4. Web systems and network technologies
5. E-learning methodologies and platforms.

Part one contains eight chapters that discuss new methods of visual, audio and video data processing. The attempts and solutions described in particular chapters present dynamically growing applications of artificial intelligence techniques in the domain of multimedia information technology such as sound source localization, dangers sound events recognition, dynamic gesture recognition, and content-based video indexing.

Second part of the book, consisting of six chapters, is devoted to the specific problems of data processing in information systems. Nowadays, data processing struggles to gather, extract and make available information that is often unavailable from the traditional data base retrieval systems. Data mining techniques and social network analysis are some of the examples of new methodologies that are wider and wider used in the extraction of latent information and knowledge investigation.

Even the most significant scientific discoveries remain familiar only to narrow groups of specialists if theory is not followed by practice. Therefore, third part of the book presents five different applications of information systems. Especially worth mentioning in this context is the application of mobile technologies to medical and healthcare systems.

Fourth group of problems discussed in the book relates to the Web systems and network technologies. As the amount of information items commonly available via Internet growths the more difficult becomes the retrieval of the relevant and reliable ones. To manage this problem new methods and techniques are proposed. Their five representatives pertaining different aspects and types of information retrieved are assembled in this part.

The last part is devoted to e-learning systems. This specific type of information systems in recent years has become one of the leading subject matters of scientific research and development. The human society worldwide has experienced total revaluation. Personal vocational success and in consequence well-being depends on the up-to-date knowledge. Today, finishing even the most prestigious university does not guarantee personal success and the term "LLL - Long Life Learning" relates to nearly all professions. E-learning systems have to facilitate learning and knowledge acquiring in the way that is the most convenient for both: employers and employees. Up-to-date e-learning methodologies, platforms, systems, and technologies are presented in five chapters that form this part.

We belief that the monograph will fulfill many expectations of its readers. We will be also very pleased if the book inspires the research community working on the Multimedia and Network Information Systems domain. If so, the goal that motivated authors, reviewers, and editors will be achieved. It will be also the greatest prize for our joint efforts.

Ngoc Thanh Nguyen
Aleksander Zgrzywa
Andrzej Czyżewski

Contents

Part IV - Web Systems and Network Technologies

List of Contributors

František Babič
Centre for Information Technologies,
Technical University of Kosice,
Slovakia

Sheila Becker
University of Luxembourg
6 rue R. Coudenhove-Kalergie,
L-1359 Luxembourg
sheila.becker@uni.lu

Peter Bednár
Centre for Information Technologies
Technical University of Kosice,
Slovakia

Michael Boronowsky
TZI Technologiezentrum Informatik
und Informationstechnik
Universitaet Bremen
Am Fallturm 1, 28359 Bremen,
Germany
worono@tzi.de

Krzysztof Brzostowski
Institute of Informatics
Wroclaw University of Technology
Wybrzeże Wyspiańskiego 27,
50-370 Wrocław, Poland
Krzysztof.Brzostowski@pwr.wroc.pl

Krzysztof Cabaj
Warsaw University of Technology
Nowowiejska 15/19, 00-665 Warsaw,
Poland
kcabaj@elka.pw.edu.pl

Kazimierz Choroś
Institute of Informatics,
Wroclaw University of Technology
Wybrzeże Wyspiańskiego 27,
50-370 Wrocław, Poland
kazimierz.choros@pwr.wroc.pl

Michał Ciesielczyk
Institute of Control and Information
Engineering
Poznan University of Technology,
Poland
Michal.Ciesielczyk@put.poznan.pl

Petr Czekaj
VSB Technical University of Ostrava
17. Listopadu 15, 70833 Ostrava
Poruba, Czech Republic
Petr.Czekaj@vsb.cz

Andrzej Czyżewski
Multimedia Systems Department
Gdansk University of Technology
Narutowicza 11/12, 80-233 Gdańsk,
Poland
ac@sound.eti.pg.gda.pl

Marek Dziubiński
Multimedia Systems Department
Gdansk University of Technology
Narutowicza 11/12, 80-233 Gdańsk,
Poland

Marcin Gorawski
Institute of Computer Science
Silesian University of Technology
Akademicka 16, 44-100 Gliwice,
Poland
Marcin.Gorawski@polsl.pl

Michal Gorawski
Institute of Computer Science
Silesian University of Technology
Akademicka 16, 44-100 Gliwice,
Poland
Michal.Gorawski@polsl.pl

Peter Hoffmann
TZI Technologiezentrum Informatik
und Informationstechnik
Universitaet Bremen
Am Fallturm 1, 28359 Bremen,
Germany
phoff@tzi.de

Agnieszka Indyka-Piasecka
Institute of Informatics
Wroclaw University of Technology
Wybrzeże Wyspiańskiego 27,
50-370 Wrocław, Poland
agnieszka.indyka@pwr.wroc.pl

Dalibor Janckulik
VSB Technical University of Ostrava
17. Listopadu 15, 70833 Ostrava
Poruba, Czech Republic
Dalibor.Janckulik@hotmail.com

Czesław Jędrzejek
Institute of Control and Information
Engineering
Poznan University of Technology,
Poland
Czeslaw.Jedrzejek@put.poznan.pl

Ireneusz Jóźwiak
Institute of Informatics
Wroclaw University of Technology
Wybrzeże Wyspiańskiego 27,
50-370 Wrocław, Poland
ireneusz.jozwiak@pwr.wroc.pl

Pawel Jureczek
Institute of Computer Science
Silesian University of Technology
Akademicka 16, 44-100 Gliwice,
Poland
Pawel.Jureczek@polsl.pl

Przemysław Kazienko
Institute of Informatics
Wroclaw University of Technology
Wybrzeże Wyspiańskiego 27,
50-370 Wrocław, Poland
Przemyslaw.Kazienko@pwr.wroc.pl

Insaf Kerkeni
Laboratoire d'Informatique
Signal et Image de la Côte d'Opale
Université de Lille Nord de France
50, rue Ferdinand Buisson, BP. 719,
62228 Calais Cedex, France
kerkeni@lisic.univ-littoral.fr

Anita Keszler
Computer and Automation
Research Institute
MTA SZTAKI, Budapest,
Hungary
keszler@sztaki.com

Jerzy Kisilewicz
Computer Systems and Networks
Wroclaw University of Technology
Wybrzeże Wyspiańskiego 27,
50-370 Wrocław, Poland
jerzy.kisilewicz@pwr.wroc.pl

Akos Kiss
Computer and Automation
Research Institute
MTA SZTAKI, Budapest,
Hungary
akos.kiss@sztaki.com

Marek Kopel
Institute of Informatics
Wroclaw University of
Technology
Wybrzeże Wyspiańskiego 27,
50-370 Wrocław, Poland
marek.kopel@pwr.edu.pl

Eugeniusz Kornatowski
Faculty of Electrical Engineering
Department of Signal Processing and
Multimedia Engineering
West Pomeranian University of Technology
26 Kwietnia 10 St.,
71-126 Szczecin, Poland
korn@zut.edu.pl

Bożena Kostek
Multimedia Systems Department
Gdansk University of Technology
Narutowicza 11/12, 80-233 Gdańsk,
Poland
bozenka@sound.eti.pg.gda.pl

Józef Kotus
Multimedia Systems Department
Gdansk University of Technology
Narutowicza 11/12, 80-233 Gdańsk,
Poland
joseph@sound.eti.pg.gda.pl

Ondrej Krejcar
VSB Technical University of Ostrava,
17. Listopadu 15, 70833 Ostrava
Poruba, Czech Republic
Ondrej.Krejcar@remoteworld.net

Dariusz Król
Wroclaw University of Technology
Institute of Informatics
Wybrzeże Wyspiańskiego 27,
50-370 Wrocław, Poland
Dariusz.Krol@pwr.wroc.pl

Marek Landowski
Quantitative Methods Institute
Maritime University of Szczecin
Waly Chrobrego 1-2,
70-500 Szczecin, Poland
m.landowski@am.szczecin.pl

Michał Lech
Multimedia Systems Department
Gdansk University Of Technology
Narutowicza 11/12, 80-233 Gdansk,
Poland
mlech@sound.eti.pg.gda.pl

Kuba Łopatka
Multimedia Systems Department
Gdansk University of Technology
Narutowicza 11/12, 80-233 Gdansk,
Poland
klopatka@sound.eti.pg.gda.pl

Zygmunt Mazur
Institute of Informatics
Wroclaw University of Technology
Wybrzeże Wyspiańskiego 27,
50-370 Wrocław, Poland
Zygmunt.Mazur@pwr.wroc.pl

Hanaa Mazyad
Laboratoire d'Informatique Signal et
Image de la Cˆote d'Opale
Universit´e de Lille Nord de France
50 rue Ferdinand Buisson, BP. 719,
62228 Calais Cedex, France
mazyad@lisic.univ-littoral.fr

Pawel Meller
Institute of Computer Science
Technical University of Lodz
ul. Wolczanska 215, 90-924 Lodz,
Poland
meller@stud.ics.p.lodz.pl

Tomasz Merta
Multimedia Systems Departament
Gdansk University of Technology
Narutowicza 11/12, 80-233 Gdansk,
Poland
tommerta@sound.eti.pg.gda.pl

Grzegorz Mikołajczak
Faculty of Electrical Engineering
West Pomeranian University of
Technology, 26 Kwietnia 10,
71-126 Szczecin, Poland
grzegorz.mikolajczak@zut.edu.pl

Jolanta Mizera-Pietraszko
Institute of Informatics
Wroclaw University of Technology
Wybrzeże Wyspiańskiego 27,
50-370 Wrocław, Poland
Jolanta.Mizera-Pietraszko@pwr.wroc.pl

Leona Motalova
VSB Technical University of Ostrava
17. Listopadu 15, 70833 Ostrava
Poruba, Czech Republic
Leona.Motalova@gmail.com

Sebastian Palus
Institute of Informatics
Wroclaw University of Technology
Wybrzeże Wyspiańskiego 27,
50-370 Wrocław, Poland

Jacek Panachida
Faculty of Computer Science and
Management
Wroclaw University of Technology
Wybrzeże Wyspiańskiego 27,
50-370 Wrocław, Poland
Jacek.Panachida@pwr.wroc.pl

Ján Paralič
Department of Cybernetics and
Artificial Intelligence
Technical University of Kosice,
Slovakia

Jakub Pęksiński
West Pomeranian University of
Technology
Faculty of Electrical Engineering
26 Kwietnia 10, 71-126 Szczecin,
Poland
jakub.peksinski@zut.edu.pl

Andrzej Piegat
Faculty of Computer Science and
Information Systems
West Pomeranian University of
Technology
Zolnierska 49, 71-210 Szczecin,
Poland
apiegat@wi.zut.edu.pl

Michal Raček
PÖYRY Forest Industry Oy, Finland

Witold Rekuć
Institute of Informatics
Wroclaw University of Technology
Wybrzeże Wyspiańskiego 27,
50-370 Wrocław, Poland
Witold.Rekuc@pwr.wroc.pl

Andrzej Siemiński
Institute of Informatics
Wroclaw University of Technology
Wybrzeże Wyspiańskiego 27,
50-370 Wrocław, Poland
Andrzej.Sieminski@pwr.wroc.pl

Damian Sikora
Gdansk University of Technology
Narutowicza 11/12, 80-233 Gdansk,
Poland,
Damian.Sikora@zak.eti.pg.gda.pl

Maciej Smiatacz
Gdansk University of Technology
Narutowicza 11/12, 80-233 Gdansk,
Poland,
slowhand@eti.pg.gda.pl

Janusz Sobecki
Institute of Informatics
Wroclaw University of Technology
Wybrzeże Wyspiańskiego 27,
50-370 Wroclaw, Poland
Janusz.Sobecki@pwr.wroc.pl

Michał Szczepanik
Institute of Informatics, Poland
Wroclaw University of Technology
Wybrzeże Wyspiańskiego 27,
50-370 Wroclaw, Poland
michal.szczepanik@pwr.wroc.pl

Krzysztof Szczypiorski
Warsaw University of Technology
Nowowiejska 15/19, 00-665 Warsaw,
Poland
kszczypi@elka.pw.edu.pl

Tamas Sziranyi
Computer and Automation Research
Institute
MTA SZTAKI, Budapest, Hungary
sziranyi@sztaki.com

Andrzej Szwabe
Institute of Control and Information
Engineering
Poznan University of Technology,
Poland
Andrzej.Szwabe@put.poznan.pl

Jakub Tomczak
Institute of Informatics
Wroclaw University of Technology
Wybrzeże Wyspiańskiego 27,
50-370 Wroclaw, Poland
Jakub.Tomczak@pwr.wroc.pl

Konrad Wiklak
Institute of Informatics,
Wroclaw University of Technology
Wybrzeże Wyspiańskiego 27,
50-370 Wroclaw, Poland
konrad.wiklak@gmail.com

Adam Wojciechowski
Institute of Computer Science
Technical University of Lodz
Wolczanska 215, 90-924 Lodz, Poland
adamwoj@ics.p.lodz.pl

Tanja Woronowicz
TZI Technologiezentrum Informatik
und Informationstechnik
Universitaet Bremen
Am Fallturm 1, 28359 Bremen,
Germany
worono@tzi.de

Aleksander Zgrzywa
Department of Information Systems
Institute of Informatics
Wrocław University of Technology,
Wybrzeże Wyspiańskiego 27,
50-370 Wroclaw, Poland
Aleksander.Zgrzywa@pwr.wroc.pl

Paweł Zwan
Multimedia Systems Department
Gdansk University of Technology
zwan@sound.eti.pg.gda.pl

Part I
Multimedia Information Technology

Chapter 1
Interseum - From Physical to Virtual Showrooms

Tanja Woronowicz, Peter Hoffmann, and Michael Boronowsky

Abstract. Following the successful proven concept of small and specialized exhibitions (so called showrooms) run by research institutions as windows to scientific innovation, the network BONITA (a project financed by the INTERREG IV B – Baltic Sea Region) extends the physical showrooms to virtual ones. While the basic idea of the physical showroom is to have an attractive exhibition area for demonstrating cutting edge-technologies in a tangible and accessible fashion and transmitting technological knowledge between science and a region, the main idea of the virtual showroom is to have centralized access to several exhibits located in different places resulting in distributed knowledge and bridging the gap between the physical and virtual world of museums and showrooms and between the expert and the visitor. The presentation of what is now technically feasible should be just one aspect of the showroom. It should also create a connection to what is technically imaginable, whereby the visionary aspects of the technology are communicated. The combination of tangible benefits and interdisciplinary visions for the future is an exceeding interesting one. Firstly it allows specific innovations to find their way to market more quickly, since they gain a higher profile and are in the public eye. On the other hand, long-term trends can also be created interactively and discussed within different target groups.

1.1 Introduction

Many people are fascinated and attracted by new technologies probably while they are continuously faced with technologically sophisticated decisions that have direct impacts on their lives. Often, their only opportunity to experience innovative technologies "live" is at exhibitions – which probably explain why there is such a crush around some stands at for example the CeBIT fair. As Susana Hornig Priest

Tanja Woronowicz, Peter Hoffmann, and Michael Boronowsky
TZI Technologiezentrum Informatik und Informationstechnik
Universitaet Bremen, Am Fallturm 1, 28359 Bremen, Germany
e-mail: {worono,phoff,mb}@tzi.de

[5] concludes in her recent introduction essay on science's contemporary audiences, the job of communicating (technological) science might be to help non-scientists feel they are not excluded as opposed to always included; that they can join in if they want, rather than that there is a necessity to spend their lives engaging. At the same time, nowadays while everything and everybody is connected to the internet, museums might be seen as a concept from the past.

Whereas the internet has evolved from static pages to interactive social networks, the concept of the museum omitted this step. While the number of interactive elements used in for museums websites is growing most of those approaches are more or less game-based. Only few approaches try to use social and even intellectual interaction. Projects like the "Wikiseum of Bad Movies" show, that such ways might be a successful solution to bridge the gap between museum experts and visitors in quite an attractive way [1, 2]. This works not only in the internet and from outside of the museum but also in the exhibition it self, as long as the special characteristics of this environment is considered [3].

Particularly small and specialized exhibitions (so called showrooms) run by research institutions are highly affected by this gap between experts and visitors. The limited availability to only one physical location and the lack of active involvement of the visitor are the main causes. Seifert [7] suggests that as technological issues become more complex they require "special cognitive effort from laypeople to be properly understood and debated". Emerging technologies might be too complicated for many—including policy- and other decision-makers—to understand without some infusion of relevant scientific or technical knowledge: A visitor may feel the need to gather knowledge about a certain exhibit in advance to have a more intense experience when seeing it live. Also after seeing the exhibit, the visitor may become more interested and would like to know more details or even share his own opinion and knowledge about it with others. At the moment, this may only be achieved by expert forums on the internet. These are neither connected to the real objects, nor the exposing institutions, nor the researchers and experts working on those exhibits. Other sorts of knowledge and information—particularly about how science is conducted including the institutional arrangements of the scientific enterprise—will always condition or moderate people's understanding and use of scientific information. Also, the efforts of negotiating one's own "social identity" influence how members of the public view and respond to scientific knowledge [9]. On the one hand, the institution misses out on the transfer of their work to the general public and industry partners. In particular we consider use of demonstration pilots tailored to the needs of additional players. On the other hand, the researchers and experts miss out on the knowledge which could be gained from a broader network of involved stakeholders. The successful transfer of their scientific knowledge into practice is an important building block to come from a pilot to innovation and is an attempt made to overcome the expert/lay divide through this special form of science communication.

We present the BONITA (Baltic Organisation and Network of Innovation Transfer Associations) physical and virtual showroom concept and possible applications to the interested public where visitors could experience firsthand what is possible with information and communication technology and how this might

translate into real world applications. The basic idea of the showroom is to have an attractive exhibition area for demonstrating technologies in a tangible and accessible fashion as well as a meeting point for workshops, seminars or lectures. They have to be flexible so that they can be used for different activities and purposes. Parallel to the physical showrooms, there are virtual showrooms to ensure the connectivity between a network of physical showrooms and to support the exchange of exhibits between the different showrooms. The main idea of the virtual showroom is to have centralized access to several exhibits located in different places meaning distributed knowledge and bridging the gap between the physical and virtual world of museums and showrooms and between the expert and the visitor. These exhibits could either be real exhibits like prototypes or intangible exhibits like software demos, videos and sketches. In order to make the virtual showroom even more attractive, there is a user interface supporting multi-touch technology by multi-touch tables and screens. This means presentations to be used in the showrooms are not done in the traditional way using power-point but rather a web-based application. This application contains a template to prepare presentations where videos, pictures, texts, software demos, etc. can be introduced in an easy way.

1.2   The Showroom Concept

1.2.1   The Physical Showroom

The showrooms as permanent exhibitions allow the diversity of information and communication technology to be made accessible to a wider public. They present mobile solutions and technology which are reachable, touchable and testable. They will be an essential role in the value-added chain in the near future. Talking of applications like wearable computing, smart clothing, mobile sensory networks and other new mobile interactive concepts, the showrooms support technology transfer in both directions – as push and as pull of technologies. The motto of such an exhibition will be "Hands-on research". On the one hand a university as public funded research organization is presenting technologies and potential applications in an attractive way. On the other hand it is an innovative form of communicating science with its outputs and acting persons. Results from current, application-oriented research projects and innovative products from the field are on show. By allowing the general public access to normally restricted content, the showroom enables a concrete dialog with various stakeholders regarding the potential take up of these technologies. In addition to technically interested laypeople, the target group might be anyone who can gain a genuine benefit from the technologies on show.

The need to support not only collaboration and exchange but also events requires a flexibility of the infrastructure to allow this multifunction. A showroom therefore provides meeting facilities like conference tables, interactive whiteboards, video conferencing etc. At the same time, it is a place for demonstrating technologies in a tangible and accessible fashion. It is a flexible exhibition space for showing technologies or presenting solutions in a professional way. The

standard of the exhibition imperatively has to be state-of-the-art to enable for example decision-makers from industry to obtain in-depth information about the technology's current status and to experience it by means of a "look and feel" approach. The presentation of prototypes or systems which have already undergone field testing is particularly informative – visitors can gain experience and find out how such systems "feel". A number of workable solutions are already available, however very few people have, until now, had the opportunity to try them out for themselves. The exhibition is intended to create a more serious, but still playful, approach to these technologies, thus boosting their acceptance. Experts of various disciplines as visitors of the showrooms can get inspired by new technologies and its applications as well and can address needs and problems of their own application domains. The showrooms are rooting universities to a region and supporting the exchange with industries but also with politics. Last but not least the improved science communication in terms of "science interaction" of current research to business partners implies the potential increase of funding for universities in the meaning of marketing.

TZI at University of Bremen as Lead Partner of the BONITA network already has been having good experiences with a showroom as a window to scientific innovation that transforms abstract research to understandable demonstrators for more than five years now. The TZI showroom concept is a central element for implementing successful information technology transfer. It serves as a highly successful as a communication tool to present advances in regional technology & innovation with a double purpose:

- diffusion of knowledge, making it accessible to people, showing the usability and benefits from technologies in practical applications to improve the social acceptance;
- serve as a forum, an encounter point for experts, to attract technology & innovation developers, users and beneficiaries, linking researchers, entrepreneurs and policy makers.

In Bremen, Wearable Computing – computers worn by the user on their body – is one example of technological exciting topics in a showroom. Technical advances in this field have produced developments whose results and application potential are generally still only known to experts. The developments often employ a strongly visual component such as for example miniaturized, radio-networked high performance computers or head mounted displays. This visibility provides Wearable Computing with a certain "realness" which can be easily implemented into interesting exhibits. Of far more interest are, however, those aspects transcending the obvious. The simple, tangible communication of such a potential intrigues many people since, as a rule, the benefits of such computer systems are often obscured by a wide variety of misinterpretations and prejudices. Wearable Computing requires more in-depth explanation if it is to be perceived as more than a desktop computer worn on a belt – whereby this aspect is actually only a marginal element of the true potential.

Successful research into Wearable Computing has been going on for a considerable time now and the technology has put down deeper roots. By now Wearable

Computing information technologies can be employed to support working processes which have previously only been able to benefit slightly or not all from the use of computer systems. The topic also provides an interesting long-term perspective, since a whole spectrum of important innovations will accompany the path of future developments, even after the technology's initial breakthrough. This means that the technology also has great economic potential. The link between technological developments and an exhibition has another interesting inherent potential. The high profile public presentation of topics should open up good access to sponsors. Sponsors are able to provide products for show or to support specific research activities by assisting with the development of new exhibits. These new technologies could then in turn be profitably implemented within the company concerned.

Concentrating technologies, solutions to problems and expertise on a topic in one place facilitates the creation of a competence region. As a logical consequence, people or companies who are interested in this topic will come to the region because of the facilities bundled there. And, since the corresponding experts will be available at the showroom, problems or ideas can be discussed with knowledgeable partners on site. Integration of the showroom into a region's business world is intended to allow local-based companies to also (or even in particular) benefit from this pull.

An important aspect of the concept described here is the ongoing evolution of the exhibition. Since the technologies concerned are continuously developing, the exhibition continually has to reinvent itself. The single showroom, on the one hand, is part of a conventional, interdisciplinary research institution carrying out independent research and developing innovative solutions in cooperation with industry. On the other hand, within a network of showrooms exist a strong European partnership and cooperation with institutions, companies and other research bodies. They are the route via which diverse research findings, prototypes or products find their way into the exhibition. To fulfill this function all of the showrooms to be set up in Riga, Vilnius, Tampere, Lulea, Odense and Bremen should share common elements. They should be attractive places for meetings and workshops with stakeholders of the innovation chain. Therefore the showroom acts as mentioned before as the place for exchange between academia and industry to transfer knowledge and to demonstrate technologies developed and to identify technologies needed based on regional context and international network.

The concept to establish showrooms inside different European regions is the key in sharing innovation experiences. They force science and politics to consider a strong customer orientation of new technologies and hence lead to economic development of edge technology providers.

The presentation of what is now technically feasible should be just one aspect of the showroom. It also creates a connection to what is technically imaginable, whereby the visionary aspects of the technology are communicated. The combination of tangible benefits and visions for the future is an interesting one. On the one hand, it will allow specific innovations to find their way to market more quickly, since they gain a higher profile and are in the public eye. On the other hand, long-term trends can also be created and discussed. For this reason, the concept is to be
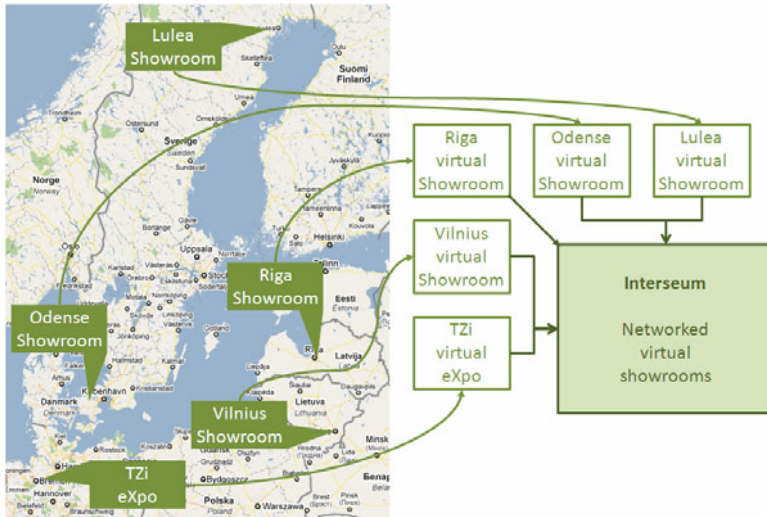
Fig. 1.1 From real showrooms to a networked showrooms: Interseum

seen less as a purely museum-based exhibition but rather far more as an innovative concept for fast-growing technological research institutions.

This institution therefore has many fascinating things to show; offers know-how relating to an exciting topic for the future and has an "open day" every day. Furthermore the physical showroom will allow the access to the virtual exhibits. Therefore the extension of the physical exhibition to a sort of parallel "Interseum" is obvious for the presented showroom concept. The term Interseum is derived from the classical "museum", enhanced by several key features.

1.2.2 Concept of the Virtual Showroom "Interseum"

A given thematic focus of the showrooms goes far beyond merely investigating, for example Wearable Computing. It also includes an examination of innovative aspects of mobile information processing. TZI's showroom is seen as a scientific arm of Bremen's ICT research cluster. The direct effect for business for 5 years now is its function as a contact for local and national industry. Expanding the showroom into virtuality is auspicious, since it allows different audiences to actively engage in enhancing the knowledge about the exhibits– both in the scientific and the commercial sense. These exhibits are media based interactive presentations of research results that are provided by all partners of the showroom network. Following the proposal of having a pilot group using multi-touch technology, the Universities of Bremen, Vilnius and Sothern Denmark will purchase a Multi-Touch-Table, and the Lulea University of Technology will purchase a Multi-Touch-LCD-Screen for gaining access to the virtual showroom in an intuitive way within the physical showroom. This supports the transregional collaboration of the partners and generates an important European value. Due to strategic

connection between the showrooms it is also planned to share individual physical exhibits between them. This makes showrooms within a region more attractive by a higher variety in exhibits. The showrooms are on the one hand side the transmission belts between science and the region and on the other hand side they are operative connectors of the regions for concrete transnational cooperation.

Need for internet-accessibility: Still the accessibility of a physical showroom is usually restricted to a single location with space for a very limited number of exhibits. In case new exhibits arrive, old ones may lose their place. Visitors need to make appointments and when they do so, a professional tour guide who has knowledge about all exhibits might be short in supply. The virtual showroom makes all the information about current, former, and future exhibits available in the internet. According to the National Science Foundation's (NSF) Science and Engineering Indicators 2006, the internet became the second highest source for science news selected by individuals seeking science news. Thereby, the range of also virtual showroom visitors is broadened to a great extend since the information is available from almost everywhere in the world at any time. Obviously "the growth of the Internet offers unique opportunities for science to establish additional channels of communication with the public. Science topics and information will be there, but the question is, will the scientific community have a prominent role in disseminating it?" asks Suleski et al, [8] with good cause. As Kua et al. [4] advise, scientists must learn to translate research both in "language and in idiom." The Internet presents a forum, but the message must still be catered to be understood by its potential audience. Adopting this advice the virtual showrooms offer different types of representation of exhibits. The representation may vary from institution to institution, be it as a graphically intense, Flash-animated virtual tour or simply a more fact-oriented, wiki-like hypertext structure. This offers the added value of a more holistic explanation of the theory and research behind the exhibits. Thereby, the public presentation of the institution itself benefits to a great extent, creating a positive impact on the transfer to and from possible industry partners.

Need for interconnection: Precondition of the operative connection of the showrooms from different institutions is to introduce a unified protocol to describe exhibits and exchange associated information. Only few museums already offer some form of a virtual museum on their website where the user can browse through a number of exposed exhibits. However, the problem is if each one uses its own format, making them stand-alone application. Exhibits of the showrooms should be described by their field of research, age, target group, current location, etc., as well as visually represented by multimedia objects such as videos or pictures. Thereby, new showrooms can easily be integrated into a network of already existing ones. The system can be seen as a non-centralized, global showroom with overlapping networks of local showrooms and their exhibits. Showrooms with relatively similar content may form a network and link subsets of their contents to each other. That way, a recommendation engine can be incorporated to suggest new exhibits to the (virtual) visitor, i.e. the visitor may see a recommendation such as "If you liked this exhibit, you may also like Exhibit ABC in Showroom XYZ." In the area which is only accessible to researchers, the recommendations and
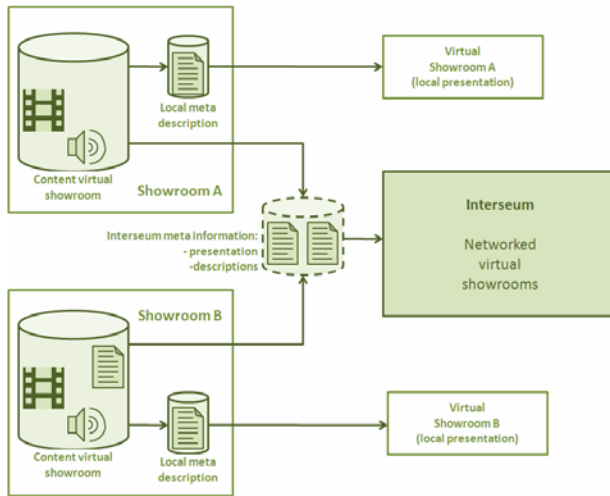
Fig. 1.2 Networked virtual showrooms: Interseum

manually created links to other research material may be even more complex. This feature can also be used to coordinate research on an exhibit. For instance, all currently involved researchers could state their progress and arrange to split up into non-overlapping research directions. People can start to generate thematic maps or user generated tours that put distributed research regarding a certain topic into a common context. The main idea is to use the BONITA network as a starting point that is providing the initial critical mass for the virtual showrooms. However, the main intention is to invite people to join this idea and to grow the number of projects that are part of a virtual exhibition.

Need for interaction: Based on the former research under the name "Wikiseum" as a web-based presentation environment and collaborative authoring system with the option for social and intellectual interaction [1, 2], Interseum will allow the interactive involvement of the users with the content. It represents the exhibits of a showroom in a dialogue with its visitors, which will ultimately benefit both parties. Visitors can incorporate their knowledge and thus to enrich the existing content. Optionally, they can describe their impressions and opinions on the issue. The interactivity includes visitors, institutions, and researchers, all having certain roles with different permissions, to access and edit the content. The information being available via virtual tour for example is usually written by technical experts and prepared for the presentation. However, different types of showroom visitors might have sometimes even a well-established expertise on a subject that is close to in depth expert knowledge.

For the future, Interseum will offer the user more interaction opportunities for influencing the presentation as such. The mentioned above selection of information, navigation through the information service, etc. restricts the user in the dealing with the system to socially and intellectually passive information consumer. This argument against passivity has strong grounds for the network of showrooms:

First, is the existing knowledge, which bears a visitor to a showroom in itself, is lost. Second, the intellectual passivity builds a barrier between the visitor and the showroom and its researchers are in. This barrier will be reduced significantly if the visitor gets the opportunity to participate with their own contributions to the presentation content. However, Interseum goes beyond the point of an enhanced multi-wiki since different target groups as well as institutions may require different views. Therefore, the presentation of the information will be fully flexible. For instance, a visitor from the general public would probably like a more visual representation where he may comment on the exhibits or ask general questions. Thereby, he is able to interact with the professionals, allowing researchers to gather new knowledge concerning their research. Also the institution – especially the persons responsible for the showroom – may gain knowledge which of the exhibits people like or dislike. This allows them to improve their marketing and the presentation of the content to attract a broader audience. While that type of interaction is more focused on the exhibition itself, researchers may use a restricted content area of the system to communicate and share insights about the exhibits and the associated research. Currently this interaction is mostly achieved with special email-lists, expert forums or personal connections. However, to access the research about an exhibit, the starting points are mostly the associated research topics. On the contrary, Interseum will allow research to be more centered on the actual exhibits. Furthermore, another important advancement to a wiki will be an automatic quality assurance module. Whereas in a regular wiki basically everybody may edit everything every time, here we not only have a set of restrictions on editing the content, typically assigned to different roles such as technical administration, marketing, researchers, visitors, and so forth. In addition to those restrictions, the quality assurance will automatically deduce required actions from statistical data. For instance, in case a certain piece of information is always edited immediately after someone is visiting it, the quality is probably very low.

Need for inter-adaptation: The designer-term inter-adaption refers to content presentation not only being adapted to the individual user by means of personalization. In this case, it also means the adaptation mechanism also includes information from other showrooms and is able to distinguish between different target groups. For instance, a virtual visitor from the general public may be provided a virtual tour. Therefore a sequence of different subsets of all available information is presented to him. If one showroom has good experiences with a certain format of a virtual tour, this format might directly be incorporated into another showroom. Furthermore, experts who visit the virtual showroom and directly prefer to skip the visual representation can continue with a more fact- and text-based one. The system can adapt to this behavior and store general user profiles as well as user profiles for those who are registered and have a user-account.

Need for internationalization: A local showroom usually exposes the information about its exhibits such as marketing material, the website, virtual tours, etc. in the local language. This makes it rather impossible to share the resources with institutions, researchers and general visitors from other countries. The virtual showroom should have a built-in language module, requiring all material to be at least

published in English, too. The problem that pages in certain languages are outdated will be prevented by the system by certain functionality, such as required inputs or reminders. This will be achieved by the integrated quality assurance module mentioned earlier.

Need for interdisciplinary: Usually when experts of different research communities work together – i.e. archeology in combination with computer science – communication and understanding can become tough since every discipline has its own language. It becomes even harder when the available information is fragmented into research papers from both communities. The virtual showrooms avoid discipline specific gaps of communication by having a centralized pool for all information, establishing a common language from the start. Furthermore, Interseum allows different views on the information. Thereby, users from all educational and cultural backgrounds and of different generations may only see the information which fits best their needs.

1.3 Conclusions

The showroom as a specified and topic-centered technical exhibition is a proven concept for transmitting technological knowledge and science and a within a region. Furthermore, It is used as operative connectors of the partner regions of BONITA for concrete transnational cooperation gaining access to the virtual showroom in an intuitive way within the physical showroom. This supports the transregional collaboration of experts and generates an important European value. The virtual extension with a unified protocol to describe exhibits and exchange associated information bridges the gap between the physical and virtual world of museums and showrooms and between the expert and the visitor with main benefits in:

- centralized access to locally distributed knowledge
- active engagement of different audiences in enhancing the knowledge about exhibits
- optimized interdisciplinary
- improved marketing of current research to business partners
- improved communication of science and scientist.

References

1. Hoffmann, P., Herczeg, M.: Wiki meets Museum - Die soziale Interaktion als Attraktivitätsgewinn für Web-Präsenzen im Kulturbetrieb. 1. Österreichische Wikiposium in Wien (2005) In: Stockinger, J., Leitner, H. (Hrsg.) Wikis im Social Web. Wikiposium 2005/06, Österreichische Computer Gesellschaf (2007)

2. Hoffmann, P., Herczeg, M.: Attraction by Interaction: Wiki Webs As A Way To Increase The Attractiveness Of Museums' Web Sites. In: Trant, J., Bearman, D. (eds.) Museums and the Web, Proceedings, Toronto: Archives & Museum Informatics (2005) (Published March 31, 2005)
    `http://www.archimuse.com/mw2005/papers/hoffmann/`
    `hoffmann.html`
3. Hoffmann, P., Herczeg, M.: Soziale Interaktion mit Dinharazade - Kommentierung in einem interaktiven mobilen Audiosystem in Museen und Ausstellungen. 2. Österreichisches Wikiposium (2006). In: Stockinger, J., Leitner, H. (Hrsg.) Wikis im Social Web. Wikiposium 2005/06, Österreichische Computer Gesellschaft (2007)
4. Kua, E., Reder, M., Grossel, M.J.: Science in the News: A Study of Reporting Genomics. Public Understanding of Science 13, 309–322 (2004)
5. Priest, S.H.: Reinterpreting the audiences for media messages about science. In: Holliman, R., et al. (eds.) Investigating Science Communication in the Information Age: Implications for Public Engagement and Popular Media, pp. 223–236. Oxford University Press, Oxford (2009)
6. Public Understanding of Science, 19, 115 (2010); originally published online (March 31, 2009)
7. Seifert, F.: Local Steps in an International Career: A Danish-style Consensus Conference in Austria. Public Understanding of Science 15, 73–88 (2006)
8. Suleski, J., et al.: Scientists are talking, but mostly to each other: a quantitative analysis of research represented in mass media
9. Wynne, B.: Public Understanding of Science Research: New Horizons or Hall of Mirrors? Public Understanding of Science 1(1), 37–43 (1992)

Chapter 2
The Synchronization of the Images Based on Normalized Mean Square Error Algorithm

Jakub Pęksiński and Grzegorz Mikołajczak

Abstract. As it is known, to transform an analogue image into digital form it is necessary to undergo the processes of sampling and quantification. The first of them consists of downloading at defined intervals data from analogue image the second one approximates analogue levels of brightness due to the closest digital levels. Both processes are the reason of an errors formation. Those errors have significant influence on the fields of the digital images transformation, in which it is necessary to synchronize images. This problem becomes particularly significant when we use the images gained from two different sources (scanner, digital camera). Anyone who uses the terms concern to images' transformation, knows that bad synchronization can lead to wrong results. In his article authors present the algorithm which eliminates problem of bad images adjustment. The paper features the method of determination of the rotation angle and axis based on computation of the Normalized Mean Square Error (NMSE) coefficient.

2.1 Introduction

Since the start of development of the fields related to the digital image processing, it has always been a problem to find a criterion for the objective evaluation of the digital images in comparison with a reference image. It is required that the objective evaluation of a criterion is consistent with the subjective evaluation. The searching process of an appropriate criterion triggered creating various quality measures including common criteria such as [1], [2]:

Jakub Pęksiński and Grzegorz Mikołajczak
West Pomeranian University of Technology
Faculty of Electrical Engineering
26 Kwietnia 10, 71-126 Szczecin, Poland
e-mail: {jakub.peksinski,grzegorz.mikolajczak}@zut.edu.pl

- **M**ean **S**quare **E**rror (MSE)

$$MSE = \frac{1}{MN} \sum_{x=1}^{M} \sum_{y=1}^{N} \left[f_{in}(x,y) - f_{out}(x,y) \right]^2 \qquad (2.1)$$

- **A**verage **D**ifference (AD)

$$AD = \frac{\sum_{x=1}^{M} \sum_{y=1}^{N} \left[f_{in}(x,y) - f_{out}(x,y) \right]}{M \cdot N} \qquad (2.2)$$

- **U**niversal **I**mage **Q**uality **I**ndex (Q)

$$Q = \frac{4\mu_{f_{in},f_{out}} \overline{f_{in}} \cdot \overline{\overline{f_{out}}}}{\left(\mu_{f_{in}}^2 + \mu_{f_{out}}^2 \right) \cdot \left[\left(\overline{f_{in}} \right)^2 + \left(\overline{f_{out}} \right)^2 \right]} \qquad (2.3)$$

where: M, N- size of the matrix featuring an image;
$f_{in}(x,y)$ – reference image;
$f_{out}(x,y)$- tested image.

The above specified comparison criteria are only a few examples of the quality measures, however, the analysis proves that the numerical value of the criteria strongly depends on the image content. This feature is extremely important, and thereby the criteria can be applied for comparison of the images with a reference image.

The following example proves that the quality measures featured by formulas 2.1 to 2.3 are efficient.

Example 1. Figure 2.1 presents a reference image (Fig. 2.1a) subject to processing by means of a few techniques applied in the digital image processing:

- High Pass (HP) filtration according to figure 1b;
- Gaussian Blur (GB) filtration according to figure 1c;
- Filtration by means of the Glass (G) filter according to figure 1d;

a)   b)   c)   d)

Fig. 2.1 The reference image (a) after: High Pass filtering (b), Gaussian Blur filtering (c), Glass filtering (d).

The processed images (Fig. 2.1b to 2.1c) were compared with the reference image (Fig. 2.1a) by means of the criteria according to formulas 2.1 to 2.3. At the same time the images were subject to the subjective evaluation performed by the representative group of ten persons. Each person assigned a mark from 1 to 10 (whereas 10 is assigned if the images are identical). On the basis of single evaluations, the evaluation of the images was performed according to Formula 2.4.

$$S = \frac{s1 + s2 + \cdots + s30}{30} \qquad (2.4)$$

where: $s1, s2, ..s30$ – single evaluations.

Table 2.1 The values of the criteria for the images after the Digital Image Operations

	MSE	*AD*	*Q*	*S*
Reference	0	0	1	10
High Pass	3283.51	-11,7 4	0.32	1.51
Gaussian Blur	51.34	0.39	0.27	8.75
Glass	436.15	0.67	0.51	3.89

The Table 2.1 features the results of the measurements. During analysis of the results from Table 2.1 it can be found that the above specified criteria can perform a purpose featuring determination of a difference of an image in relation to the reference image, and the objective evaluation is consistent with the subjective evaluation.

However, all criteria have a general drawback featuring a strong dependency on rotation of a tested image in relation to the reference image. The effect of rotation on the values of the criteria is shown in example no. 2.

Example 2. The images from Figure 1 rotated by 1° in relation to the reference image from Fig. 2.1a are featured in Fig. 2.2.

a)   b)   c)   d)
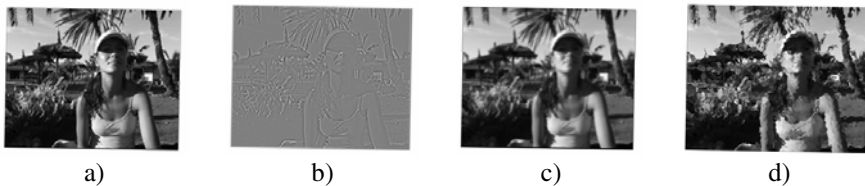
Fig. 2.2 The reference image rotated by 1° (a) after: High Pass filtering (b), Gaussian Blur filtering (c), Glass filtering (d).

The images were subject to the objective and the subjective evaluation in relation to the reference image in the same way as in example no. 1. The results of the measurements are shown in Table 2.2.

Table 2.2 The values of the criteria for the images after the Digital Image operations and rotation by 1

	MSE	*AD*	*Q*	*S*
Reference image rotated by 1	879,21	0.17	0.78	10
High Pass rotated by 1	2661.13	-15,50	0.211	1.51
Gaussian Blur rotated by 1	556.18	7.60	0.23	8.75
Glass rotated by 1	1511,10	9.03	0.29	3.89

During comparison of the results featured in Table 2.2 with the results featured in table 1 it can be noticed that the image rotation has a significant effect on the values of the quality criteria despite no change in the subjective evaluation (the persons performing the evaluation found only that the images were rotated). Obviously, this problem is not so important if there is a big difference between a reference and a tested image. However, if no difference (as for the images from fig. 1a and 2a) or a very small difference between the evaluated images is observed, the problem is particularly important because a quality measure introduces big errors and its value is not consistent with the subjective evaluation of the observers. Thereby, the criterion is irrelevant and its value is unreliable.

Example 3. Another example of errors resulting from the lack of synchronization is the learning process of the neural network, aiming at generating the optimal filter FIR correcting noices in matrice CCD [11]. Inaccurate matching of images, it causes the amplitude characteristics of the filter is distorted. Neural network tries to correct the image rotation.

Figure 2.3 shows the amplitude characteristics of the filters. 2.3a perfectly matched images, 2.3b rotated by 1 degree, 2.3c rotated by 5 degrees relative to the image test.

a)   b)   c)
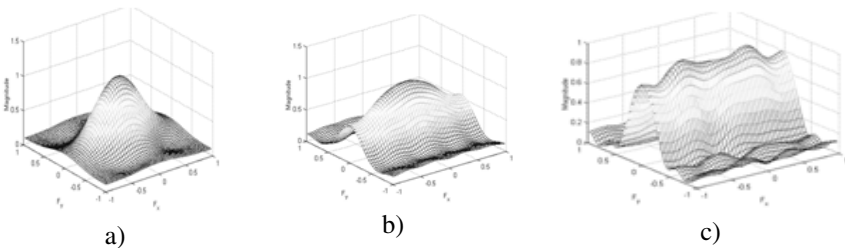
Fig. 2.3 Filter amplitude characteristic: (a) - images adjusted, (b) - images rotated by 1°, (c) - images rotated by 5°

As is clear from the examples fit the image is an important process. if an angle "α" and an axis of rotation between the images is known, the isometry according to Formula 5 (rotation of an image towards the centre of the coordinate system), it is illustrated on Fig. 2.4 or Formula no. 2.6 (rotation of an image towards the point A or B (fig.2.5) can be performed [3].
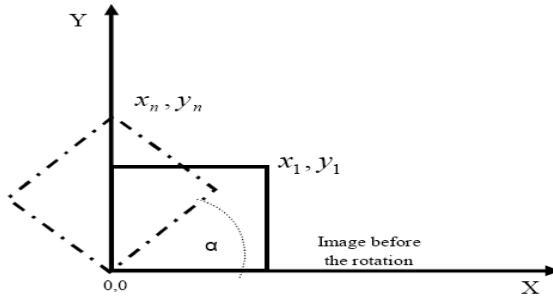
Fig. 2.4 Image's rotation towards the beginning of the coordinates' axis
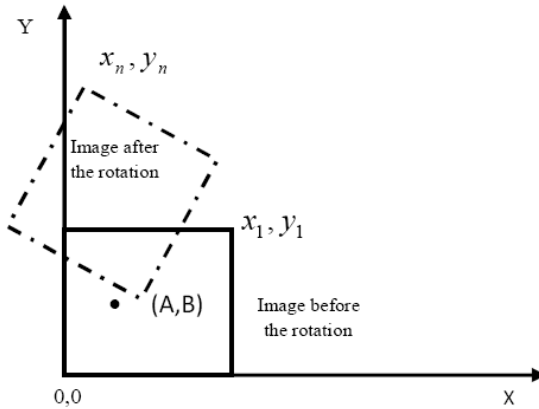
Fig. 2.5 Image's rotation towards the whichever point.

$$x_n = x_1 \cdot \cos(\alpha) - y_1 \cdot \sin(\alpha)$$
$$y_n = x_1 \cdot \sin(\alpha) - y_1 \cdot \cos(\alpha)$$

(2.5)

$$x_n = A + (x_1 - A) \cdot \cos(\alpha) - (y_1 - B) \cdot \sin(\alpha)$$
$$y_n = B + (x_1 - A) \cdot \sin(\alpha) - (y_1 - B) \cdot \cos(\alpha)$$

(2.6)

where: x_n, y_n - pixel's value after the shift; x_1, y_1 - pixel's value before the shift; (A,B) - rotation's point; α - rotation's angl

   If neither an axis of rotation nor an angle is known, it is necessary to determine these values. The paper features the method of determination of the rotation axis and angle between the images based on the NMSE coefficient test [1], [4] according to the following Formula 2.7.

$$NMSE = \frac{\sum\limits_{x=1}^{M}\sum\limits_{y=1}^{N}\left[f_{in}(x,y)-f_{out}(x,y)\right]^2}{\sum\limits_{x=1}^{M}\sum\limits_{y=1}^{N}\left[f_{in}(x,y)\right]^2} \qquad (2.7)$$

Where: M, N - size of the matrix featuring an image;
$f_{in}(x,y)$ - reference image;
$f_{out}(x,y)$ - tested image.

2.2 The Principle of the Algorithm of Determination of the Rotation Axis and Angle Based on the NMSE

The algorithm of determination of the rotation axis and angle between the digital images is performed in two steps:

- Step 1. Determination of the rotation axis between the images;
- Step 2. Determination of the rotation angle between the images.

The first step is to find the rotation's axis. For this purpose the authors of the article have used algorithm, which functioning's principle is based on the NMSE error counting between images and it is as follows:

- We take the window in size 20 per 20 pixels (it can have a different size) and shift it starting from lower edge of the image (point 0, 0) by every one pixel to the right. In every window's position like that we count the NMSE error between images;
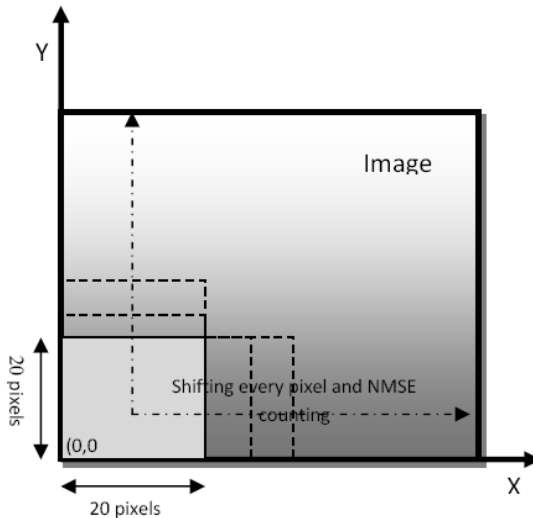
Fig. 2.6 Image's rotation towards the which ever point

- At the moment of the access to the edge of the image, we shift window to the beginning;
- We lift the window by one pixel up and start the whole process from the beginning.

In the range, where the value of the NMSE error will be minimal, there is the searched rotation's axis. This process is presented on the Fig. 2.6. Exemplary surface gained in effect of functioning this part of the algorithm is presented on the Fig. 2.7. As it is presented, the smallest NMSE error is situated close to the edge of the image. It means that the rotation's axis is located precisely there. It is correct because exemplary images were rotated precisely towards the one of the image's edges.
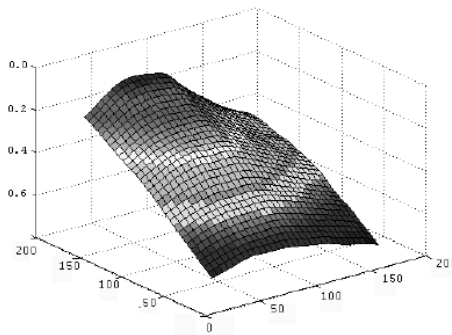
Fig. 2.7 Sample map of distribution of the correlation coefficient

After finding the rotation's axis of the images towards themselves, finding the rotation's angle still remains. The procedure of the rotation's angle searching is following (Fig. 2.8):

- We take the furthest perpendicular edge away from the rotation's axis (marked as A);
- Along this edge we use window in size 5 per 5 pixels by every one pixel. In every position of this kind we need to count the NMSE error.

The process is finished when the NMSE error achieves the minimum value. In this way we received the length of the second edge (marked as B)

After the right-angled triangle is formed, the angle of rotation (α) between the images can be determined by means of the trigonometric relation according to Formula 2.8.

$$\alpha = tg\left(\frac{b}{a}\right) \qquad (2.8)$$

Block diagram of the searching algorithm of the rotation's axis and angle is presented on the Fig. 2.9.
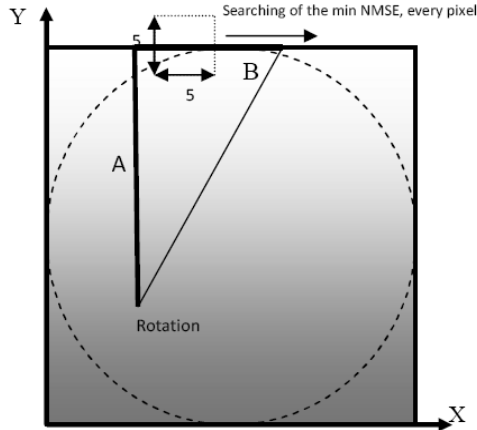
Fig. 2.8 Sample map of distribution of the correlation coefficient
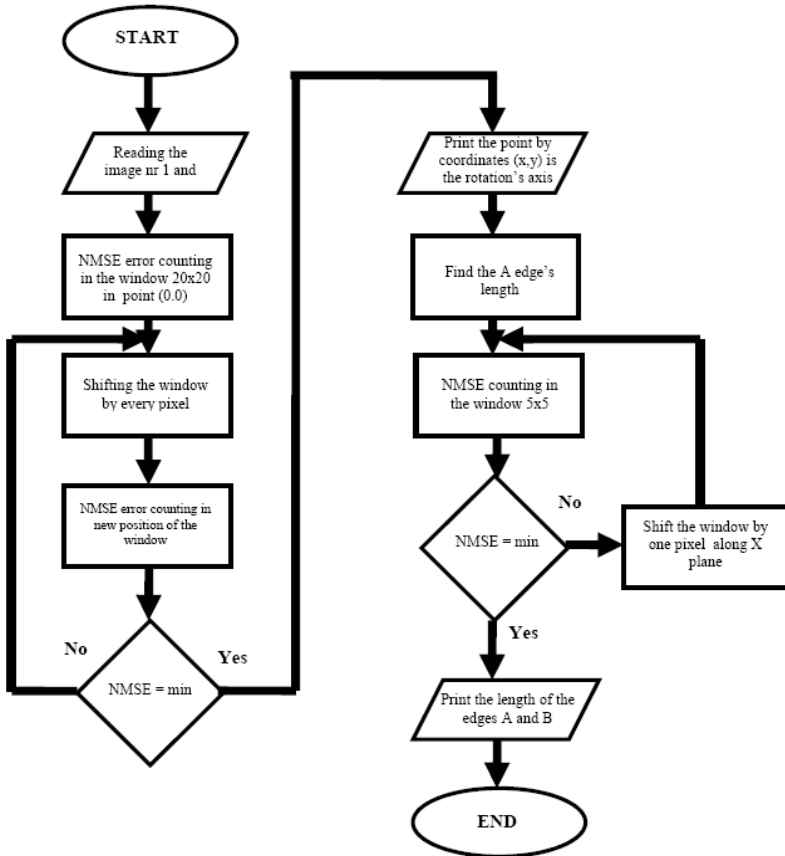
Fig. 2.9 Block diagram, the algorithm rotation's angle and axis.

2.3 The Experimental Results

In order to verify the efficiency of the method of determination of the rotation axis and angle between the images, the author performed a number of the tests. The idea of the tests was that a few reference images of various size (640x480, 1024x768 and 1600x1200 pixels) and 300 dpi resolution were purposely rotated by a given angle (α) around a given rotation axis (O) (so that the new images rotated in relation to the reference image were obtained).

   The pairs consisting of the reference and the rotated image were subject to the algorithm featured in section 2 of the paper. The results of the tests are presented in Tables 2.3 to 2.5.

Table 2.3 The results of the tests for the images of 640x480 size and 300 dpi resolution

Given rotation axis		Determined rotation axis		Given rotation angle	Determined rotation angle
X	Y	X	Y		
150	150	151	149	$1°$	$1°$
150	150	151	149	$2°$	$2°$
150	150	151	149	$3°$	$3°$
100	100	97	101	$1°$	$1°$
100	100	97	101	$2°$	$2°$
100	100	97	101	$3°$	$3°$
200	150	202	152	$1°$	$1°$
200	150	202	152	$2°$	$2°$
200	150	202	152	$3°$	$3°$

Table 2.4 The results of the tests for the images of 1024x768 size and 300 dpi resolution

Given rotation axis		Determined rotation axis		Given rotation angle	Determined rotation angle
X	Y	X	Y		
500	500	502	500	$1°$	$1°$
500	500	502	500	$2°$	$2°$
500	500	502	500	$3°$	$3°$
250	250	247	249	$1°$	$1°$
250	250	247	249	$2°$	$2°$
250	250	247	249	$3°$	$3°$
750	550	751	553	$1°$	$1°$
750	550	751	553	$2°$	$2°$
750	550	751	553	$3°$	$3°$

J. Pęksiński and G. Mikołajczak

Table 2.5 The results of the tests for the images of 1600x1200 size and 300 dpi resolution

Given rotation axis		Determined rotation axis		Given rotation angle	Determined rotation angle
X	Y	X	Y		
800	800	798	803	$1°$	$1°$
800	800	798	803	$2°$	$2°$
800	800	798	803	$3°$	$3°$
400	400	403	399	$1°$	$1°$
400	400	403	399	$2°$	$2°$
400	400	403	399	$3°$	$3°$
1000	850	1003	854	$1°$	$1°$
1000	850	1003	854	$2°$	$2°$
1000	850	1003	854	$3°$	$3°$

2.4 Conclusions

The analysis of the results shown in tables 2.3, 2.4 and 2.5 proves the efficiency of the algorithm proposed by the author of the paper.

Even though the paper features the problem of lack of synchronization of images exclusively on the basis of its effect on the values of the quality criteria, the problem is also encountered in other scientific fields (e.g. in the neural network learning processes [5], [11]).

Using this adjusting algorithm before issues of digital image processing, many errors caused by this can be eliminated.

Therefore, the algorithm elaborated by the author has a practical value, because due to its simplicity and high efficiency, it can be applied in all fields requiring synchronization of the images.

References

1. Kornatowski, E.: Probabilistyczna miara wierności odwzorowania sygnału, Kwartalnik Elektronika i Telekomunikacja, vol. 45 (1999)
2. Wang, Z., Bovik, A.C.: A Universal Image Quality Index. IEEE Signal Processing Letters 9(3), 81–84 (2002)
3. Pęksiński, J., Mikołajczak, G., Kowalski, J., Kornatowski, E.: Filtracja liniowa i nieliniowa obrazów dyskretnych, Wydawnictwo Hogben, Szczecin (2005)
4. Zieliński, T.P.: Cyfrowe przetwarzanie sygnałów, Wydawnictwo Komunikacji i Łączności, Warszawa (2005)
5. Osowski, S.: Sieci neuronowe do przetwarzania Informacji, Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa (2000)
6. Gayle, D., Mahlab, H., Ucar, Y., Eskicioglu, A.M.: A Full-Reference Color Image Quality Measure in the DWT Domain 13th European Signal Processing Conference (EUSIPCO 2005), Antalya, Turkey, September 4-8 (2005)

7. Kornatowski, E.: Analiza właściwości miar jakości obrazów cyfrowych XXVII IC-SPETO 2004, pp. 427–430 (2004)
8. Kornatowski, E.: Miara oceny jakości odwzorowania sygnału V Konferencja Naukowo-Techniczna Zastosowanie Komputerów w Elektrotechnice, Poznań/Kiekrz 2000, pp. 85-88 (2000)
9. Eskicioglu, A.M.: Quality Measurement For Monochrome Compressed Images. In: The Past 25 Years, Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) Conference, Istanbul, Turkey, June 5-9, vol. 4, pp. 1907–1910 (2000)
10. Thong, T.: Digital image processing test patterns. IEEE tr. on ASSP 31 (June 1983)
11. Pęksiński, J., Mikołajczak, G.: Generation of a FIR filter by means of a neural network for improvement of the digital images obtained using the acquisition equipment based on the low quality CCD structure. In: Nguyen, N.T. (ed.) ACIIDS 2010, Part I. LNCS (LNAI), vol. 5990, pp. 190–199. Springer, Heidelberg (2010)

Chapter 3
Evaluation of the Separation Algorithm Performance Employing ANNs

Marek Dziubiński and Bożena Kostek

Abstract. The objective of the presented study is to show that it is possible to effectively separate harmonic sounds from musical sound mixtures for the purpose of automatic sounds recognition, without any prior knowledge of the mixed instruments. It has also been shown that a properly trained ANN enables to reliably validate separation results of mixed musical instrument sounds, and the validation corresponds with subjective perception of the separated sounds quality. A comparison between the results obtained with the use of the ANN-based recognition, subjective evaluation of the separation performance and the energy-based evaluation is provided.

3.1 Introduction

Internet services make vast amounts of multimedia data available for exchange and browsing, thus this induces the need for effective and unambiguous ways of data description based on small fractions of information [5,8]. In the case of audio content, a description of musical phrases and the focus on the discovery and recognition of melodic lines, regardless of the instruments playing the phrase is one of the important tasks. Identification of various instruments in the mix, as well as retrieval of melodic lines is the task of automatic music transcription (AMT) systems, whose goal is to obtain a complete description of a recorded musical piece. This type of applications belongs to the Music Information Retrieval (MIR) domain [5]. Complex AMT systems analyze mixed sounds and trace several pitch tracks (melody lines) [4,6,7,12]. This can be carried out in two independent ways. The first way is usually based on representing a musical signal in the time-frequency domain, and employs pitch tracking directly on such a representation. The second way is to couple a single pitch detection with separation algorithms,

Marek Dziubiński and Bożena Kostek
Multimedia Systems Department
Gdansk University of Technology, Narutowicza 11/12, 80-233 Gdańsk, Poland
e-mail: `bozenka@sound.eti.pg.gda.pl`

and then to detect all melody lines in an iterative procedure. Multi-pitch AMT systems of the first type remain more successful in obtaining reliable pitch contours, and they can deal with simple two or three instrument phrases assuming that the pitch relations of the mixed sources are not harmonic.

The objective of this article is to present a complex method of processing, which provides successful automatic recognition of separated sounds. Automatic recognition supported by a properly trained Artificial Neural Network validates the separation result through the ANN's ability to properly recognize the separated sources. As an extension to this study, a subjective evaluation is carried out to assess whether the results obtained by the ANN and the energy-based separation error evaluation correspond with the listeners' perception of the separation sounds quality.

3.2   AMT System Engineered at the MSD

Fig. 3.1 shows the proposed AMT system with a single pitch detection algorithm coupled with a separation algorithm. The system is extended by two blocks: the parametrization unit and the ANN-based automatic recognition decision system. The AMT system from Fig. 3.1 detects pitch contour of a single instrument and passes the results to the separation block. The separation algorithm removes the sound from the mixture, and then the reduced mix feeds the input of the single pitch detector again. The iterative process is repeated until all pitch contours are estimated and all sources separated. It is obvious that the very important part of the system is pitch detection algorithm (PDA). Overall, pitch detection is one of the most difficult and important tasks in speech processing and automatic music transcription. As such, it has been exploited in numerous publications for many years. Problems related to pitch detection arise from the non-stationary nature of audio signals. In addition, often significant noise disturbances are contaminating analyzed signals making pitch estimation even more difficult. Thus far, a universal PDA solution has not been proposed.

An effective frequency domain-based PDA engineered and published [2] by the authors, called the Spectrum Peak Analysis (SPA), is based on the analysis of peaks representing harmonics of a processed signal in the frequency domain. The underlying concept is that it is relatively easy to determine pitch by investigating a signal spectrum, especially when focus is on analyzing intervals between partials present in the spectrum. This is regardless of the fact that some harmonics may be absent or somehow obscured by the background noise. An assumption should be made however, that at least some of the harmonic peaks are larger than the energy of the background noise.

Sound Separation Algorithms and Separation Procedures

The separation block of the extended AMT system (see Fig. 3.1) plays an essential role in the overall AMT system performance. Therefore this Section shortly explores the separation algorithms used in the study. In general, due to the complexity of the problem, methods that deal with it are computationally inefficient, and
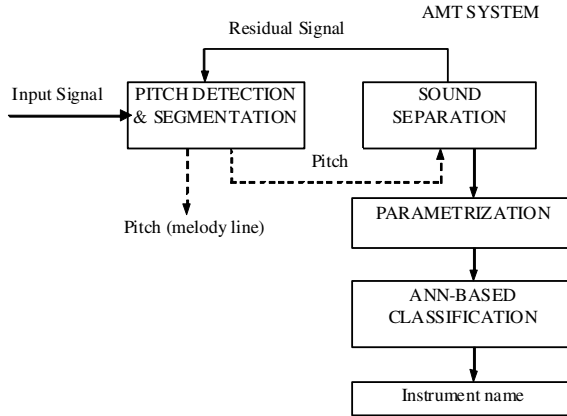
Fig. 3.1 Extended AMT with the ability to recognize instrument sounds in the mix

so far do not provide satisfying audible results. Nevertheless, in recent years, an extensive research has been conducted on this subject, and resulted in interesting ideas and solutions, among which the most promising is sinusoidal modeling approach (SM) [11] extensively exploited over the last decade.

Algorithms engineered by the authors, described thoroughly in one of their publications [1] overcome several limitations of typical SM based algorithms, where higher frequency trajectories cannot be effectively detected as they have relatively low energy and exhibit significant frequency deviations. The engineered algorithms [1,3] utilize a concept of time-frequency sampling with the use of cisoids exactly matching time-frequency characteristics of the harmonic components of the modeled sounds. Two algorithms (A1 and A2) are based on the SM approach with regard to the signal synthesis. The major difference between algorithms A1 and A2 stems from the fact that the size of the computing window depends on the average frequency of a synthesized harmonic, i.e., with the increase of the harmonic order, the window size decreases, while in algorithm A2, the size of the window is constant for all harmonics. Based on the instantaneous pitch tracks of to the harmonic sources in the mixture, average fundamental frequencies representing each block are determined and used for synthesis of the sinusoidal partials. Algorithms A3 and A4 represent a novel approach to harmonic sound separation, where the separated sound is represented by complex spectrum components, calculated with the use of cissoids with time varying frequency. Algorithm A3 differs from the SM approach, and is based on *frequency* sampling of the harmonic content of signals. Algorithm A4 utilizes both: the overlap-add method (as in algorithm A2), where harmonic partials of the signals are represented by the overlapping sinusoidal sequences, and *spectral* representation with frequency varying cisoids (as in A3). To preserve smoothness on the block edges, synthesized signal segments overlap. Based on the instantaneous pitch tracks of the mixed sounds, instantaneous fundamental frequencies representing each block are obtained and used for synthesizing all of the harmonic partials.

It is important to note that sounds from the mix can be synthesized in various ways, i.e., separated sources may be removed from the mix, or preserved in the mix before the next iteration (separation) step. Moreover, the separation process may be performed starting with the lowest frequency harmonic components, or with the highest frequency components, etc. Therefore several separation procedures have been explored by the authors:

Procedure I: Each synthesized sound is subtracted from the mix, and the reduced mix creates a new input signal for the next separation stage (separation of the consecutive instrument). Separation order is designated by the average pitch of each instrument, starting from the sound with the lowest fundamental frequency. The last separated sound has the highest pitch.

Procedure II: All instruments are synthesized, based on the complete mix (sounds are not subtracted from the input), i.e., the complete sound mixture is used as the input for the separation of each sound.

Procedure III: The separated sound is not synthesized but created by subtracting all other instruments from the mix. For example, let us consider the case of four sounds in the mix, from which we want to separate instrument No. 1. Firstly, sound No. 2 is synthesized and subtracted from the complete mix. Then, the resulting residuum is used to synthesize sound No. 3. After subtracting sound No. 3, the residual signal becomes the input for sound No. 4 synthesis. Finally, after subtracting sound No. 4, the residual signal, which is assumed to contain only partials of sound No. 1, is used in the recognition process. In the next stage (separation of instrument 2), the procedure is similar, i.e., all instruments are subtracted from the mix and the residuum becomes the chosen sound, etc.

Procedure IV: This procedure is similar to Procedure III, however, in this case the signal used for recognition is synthesized based on the residuum that was created through subtraction of all other sounds. The synthesized signal contains only harmonic content, while the final residual signal obtained in procedure III contains also inharmonic components that remain after all other sounds separation.

Calculating the Separation Error

Separation algorithms performance evaluation is problematic in the case of separation algorithms. It may depend on the type of sources, and also on the information describing the sources which is used to estimate the separation error. The simplest way to evaluate the separation method is to calculate the energy-based error. This evaluation method is based on the energy ratio of the error signal and the input signal used in creating the mix (*SRR*) [11]:

$$E_{SEP} = \sqrt{\frac{1}{N} \sum_{n=1}^{N} \left(S_{ORIG}[n] - S_{SYNTH}[n] \right)^2} \tag{3.1}$$

$$SRR = \sqrt{\frac{1}{N} \sum_{n=1}^{N} S_{SYNTH}^{2}[n]} \Big/ E_{SEP} \tag{3.2}$$

where: N – signal length (in samples), S_{ORIG} - original signal, S_{SYNTH} - separated signal (synthesized).

Although many researchers follow such an approach, *SRR* has several limitations. The first very important issue is that this method assumes the existence of the original signal used in creating the mix, which may not be available. Secondly, properties of the separated signal should be analyzed in terms of the separation purpose. For example, transients, which are important from a listener's point of view contain a relatively small amount of energy. Thus, their subjective quality may rapidly decrease when they become significantly disturbed by the separation process, but at the same time the *SRR* may indicate an effective separation. Conversely, phase inversion of the harmonic components is negligible from the listener's point of view, but produces low *SRR*.

3.3 Experiment Lay-Out

The Artificial Neural Network (ANN) algorithm was employed to validate the performance of the sound source separation algorithms. A three-layer feed-forward NN having a structure described below was used in the experiments. The number of neurons in the initial layer was equal to the number of elements in the feature vector (FV), the number of neurons in the hidden layer was equal to the number of neurons in the initial layer, each neuron in the output layer was mapped to a different class of musical instrument; neurons in the initial and output layers had log-sigmoid transfer functions, while the hidden layer uses tan-sigmoid transfer functions.

ANNs were trained with FVs derived from the original sound samples selected from the Catalogue of Musical Instrument Sounds created in MSD, GUT [8]. The feature vectors were randomly divided into two sets: training and testing. Each set contained 50% of all FVs. An error back-propagation algorithm was used to train the ANN. The process of training was considered as finished after the cumulative error of the network responses to the set of the training vectors had dropped below a defined threshold, or when the cumulative error of the network responses to the set of the validation vectors had been rising for more than 10 cycles in a row [1]. It should be stressed that the testing set was constructed of the remaining 50% of the instrument sounds used for mixing. The class of the instrument was determined by the set of the highest values at the ANN output.

All experiments utilized a set of five musical instrument sounds of the same musical articulation (i.e. B flat clarinet, oboe, French horn, violin, trumpet; the whole musical scale). Altogether, 361 sound samples with a sampling rate of 44.1 kHz were used. 40 sound samples from five groups of musical instruments were chosen for mixing and separation (eight sound samples from each group, i.e., each instrument was represented by eight sounds of different pitch). Every mix contained from two to four sounds, each belonging to a different class. Sounds from each class were chosen randomly and served to create 50 mix examples. Altogether, the mixes contained 115 sounds, which were later separated and classified. Each mix was formed only from sounds with different pitch. For all mixtures,

separation was carried out for four processing algorithms and procedures described in Section 2. Recognition experiments were performed for various configurations of features forming the feature vectors (FVs). Parameters redundant in terms of separability [8] were excluded from the FVs, based on the Fisher criterion.

Feature Vectors

The key issue in the automatic classification of instrument sounds is their parametrization. Since descriptors used in parametrization, also those formulated within the MPEG-7 standard, were thoroughly reviewed in many studies [5,8,9,10] thus they are only listed here.

Feature vectors employed in experiments are shown below:

FV_1={*KeyNum, ev, LAT, HSC, HSD, HSDv, HSS, HSSv, ASC, ASS, ASSv, $ASE_{2\div5,8,9,18,21,23\div34}$, $ASEv_{5\div9,21,31,34}$, $SFM_{13\div22,24}$*}

FV_2={$D_{1\div10}$ $S_{1\div10}$ $H_{1,2\div4,5,6\div10}$}

FV_3={*KeyNum, ev, LAT, HSC, HSD, HSDv, HSS, HSSv, ASC, ASS, ASSv, $ASE_{2\div5,8,9,18,21,23\div34}$, $ASEv_{5\div9,21,31,34}$, $SFM_{13\div22,24}$ $D_{1\div10}$ $S_{1\div10}$ $H_{1\div10}$*}

FV_4={*KeyNum, $D_{1\div10}$, $S_{1\div10}$, $H_{1\div10}$*}

FV_5={*KeyNum, $D_{1\div10}$, $S_{1\div10}$, $H_{1\div10}$, HSC, HSD, HSS, HSV* }

FV_6={*KeyNum, $D_{1\div10}$, $S_{1\div10}$, $H_{1\div10}$, ev, Tri_1, Tri_2, Tri_3*}

where [1]: *KeyNum* – parameter related to sound pitch, *ev* - even harmonic content (for *N* harmonics), *D* - mean value of differences between amplitudes of a harmonic in adjacent time frames, *H* - mean value of amplitudes of a harmonic over time, *S* - standard deviation of amplitudes of a harmonic over time, *LAT - Log Attack Time* defined as the logarithm (decimal basis) of time between a signal start and the moment it reaches its sustained part, $Tri_{1,2,3}$ – so-called *modified Tristimulus*, describing relations in harmonics energy, *ASE* (*Audio Spectrum Envelope*) describes the short-term power spectrum of the waveform. The mean values and variances of each coefficient over time are denoted as $ASE_1…ASE_{34}$ and $ASEv_1…ASEv_{34}$ respectively, *ASC* (*Audio Spectrum Centroid*) describes the center of gravity of the log-frequency power spectrum; the mean value and the variance are denoted as *ASC* and *ASCv* respectively, *ASS* (*Audio Spectrum Spread*); the mean value and the variance over time are denoted as *ASS* and *ASSv* respectively, *SFM* (*Spectral Flatness Measure*) calculated for each frequency band; the mean values and the variances are denoted as $SFM_1…SFM_{24}$ and $SFMv_{1v}…SFMv_{24}$, parameters related to discrete harmonic values: *HSD* (*Harmonic Spectral Deviation*), *HSS* (*Harmonic Spectral Spread*), *HSV* (*Harmonic Spectral Variation*).

ANN-Based Experiment Results

The ANN trained with FV_1 correctly classified 179 out of 181 original sound samples of the testing set, and reached the effectiveness of 98.9%. The training process for FV_1 was very quick and the training goal was achieved within less than 35 epochs. However, the classification results were not satisfactory for the separated sounds. The ANN trained with FV_2 correctly classified 145 out of 181 original sound samples of the validation set, and reached the effectiveness of 80.1%.

The training process for FV_2 was relatively slow and the training goal was not achieved. Since the classification error for the validation set did not decrease for over 10 epochs, the training process was terminated. Interestingly, classification performance for the separated sounds improved significantly in comparison to FV_1 (see Table 3.1). However, if the ANN incorrectly classifies the original sound, it would most likely incorrectly classify the separated sound as well. Thus, in order to achieve better results of the classification for both: original and the separated sounds, feature vectors of subsequent experiments consisted of the descriptors from both FV_1 and FV_2.

Although the effectiveness of the original sound classification was high again (98.8%), the performance of the separated sounds classification dropped. Similarly to FV_1, the training goal was also reached very quickly (approx. 35 epochs). Like FV_1, feature vector FV_3 contained MPEG7 parameters, while features in FV_2 were only related to harmonics energy behavior. Better performance of the separated sounds classification with the use of FV_2 shows that MPEG7 descriptors alone are not suitable for separated sources recognition. Therefore, further experiments focused on descriptors related directly to harmonic components of the sounds.

The consecutive FV consisted of the descriptors from FV_2 complemented with *KeyNum* which carries additional information about the relation between harmonic energy and pitch. Despite the worsened effectiveness of the original sound recognition (92.3%), the results of the separated sounds recognition were quite satisfactory. In this case, the training goal was not reached, because the classification error of the validation set did not decrease for 10 epochs in a row. Thus, the training was assumed to be finished. The best algorithm-procedure combination turned to be A3(IV) (see Table 3.1), and it reached the performance of 93.9%. In order to improve this achievement, the next FV consisted of the descriptors from FV_4 complemented with the remaining FV_1 descriptors (namely *HSC*, *HSD*, *HSS* and *HSV*). The ANN correctly classified 94.5% of the original sound samples. However, the results obtained for separated sounds were worse than before. Although *Harmonic Spectral* descriptors include information on the harmonic content of the signal, they do not carry data on each harmonic separately. This may have constituted the reason for worsened results. Thus, the last FV examined consisted of the descriptors from FV_4 and some simple spectral attributes, such as: *ev*, Tri_1, Tri_2 and Tri_3, which describe relations between harmonics components of a signal. For this case the performance reached 93.4% for the original sounds and 97.4% for the separated sounds.

Table 3.1 presents the classification results for every algorithm, procedure and feature vector combination. A Roman number in brackets, preceded by the separation algorithm name represents, the procedure utilized by that algorithm.

As seen in Table 3.1, the best performance was achieved for A3(IV) (algorithm A3 and procedure IV), which means that it is important to remove all but the separated sound from the mix, and use the residual signal for synthesis. This specific algorithm/procedure combination enables the ANN to correctly classify 97.4% of the sound samples (classifying erroneously only three times.)

Overall, this result is very encouraging. Three unsuccessfully recognized signals were French horn sounds. The erroneous classification may have been caused

Table 3.1 Classification results for FV_{1-6}

Algorithm/ procedure	No. of sep.sounds	Performance [%] / No. of incorrectly recognized sounds					
		FV_1	FV_2	FV_3	FV_4	FV_5	FV_6
A1(I)	115	50.43/57	66.09/39	59.13/47	80.00/23	78.26/25	78.26/25
A1(II)	115	45.22/63	67.83/37	56.52/50	75.65/28	77.39/26	77.39/26
A1(III)	115	45.22/63	37.39/72	41.74/67	34.78/75	53.04/54	40.87/68
A1(IV)	115	54.78/52	68.70/36	66.09/39	81.74/21	80.00/23	81.74/21
A2(I)	115	60.00/46	72.17/32	71.30/33	86.96/15	82.61/20	87.83/14
A2(II)	115	60.00/46	71.30/33	71.30/33	86.96/15	82.61/20	87.83/14
A2(III)	115	50.43/57	53.91/53	48.70/59	55.65/51	63.48/42	61.74/44
A2(IV)	115	63.48/42	73.04/31	71.30/33	88.70/13	83.48/19	89.57/12
A3(I)	115	60.87/45	81.74/21	73.04/31	91.30/10	78.26/25	95.65/5
A3(II)	115	61.74/44	83.48/19	73.04/31	89.57/12	78.26/25	94.78/6
A3(III)	115	48.70/59	51.30/56	52.17/55	51.30/56	65.22/40	60.87/45
A3(IV)	115	65.22/40	83.48/19	74.78/29	93.91/7	80.87/22	**97.39/3**
A4(I)	115	60.00/46	74.78/29	69.57/35	87.83/14	80.87/22	87.83/14
A4(II)	115	58.26/48	77.39/26	69.57/35	90.43/11	81.74/21	89.57/12
A4(III)	115	57.39/49	58.26/48	51.30/56	63.48/42	73.91/30	78.26/25
A4(IV)	115	63.48/42	77.39/26	71.30/33	93.04/8	85.22/17	93.91/7
Total	1840	56.6/799	68.6/577	63.8/666	78.2/401	76.6/431	81.5/341

by the fact that these sounds had the lowest (or second lowest) fundamental frequency in the mix, and the difference in pitch between the lowest and the highest pitched sound (in the mixture) was significant.

3.4 Comparative Analysis of the ANN-Based Recognition Results, Subjective Tests and Energy-Based Separation Error Evaluation

Subjective Tests and ANN Recognition

The aim of the subjective analysis experiment was to compare the ANN algorithm responses with the results obtained in the subjective tests, therefore to determine whether the ANN system results were generally consistent with the listeners' answers.

Subjective tests involved 25 experts, who were asked to evaluate 120 pairs of the sounds in two test series. Two test series allowed for verifying consistency of a listener's answers. Based on a statistical paired test [8] for $n=120$ (number of pairs) and the significance level at 5%, the critical value of inconsistent answers was calculated revealing that all experts passed the test. Each pair contained one successfully and one unsuccessfully separated sound chosen randomly. Each pair was preceded by a reference sound (instrument used for mixing).

The next step was to compare the ANN results to each listener's preferences with the use of the χ^2 Pearson's conformity test [8]. For the critical value, obtained from statistical tables for the assumed significance level, 21 experts'

answers were consistent with the ANN results, for 4 remaining experts the χ^2 value slightly exceeded the critical value.

In general, 84% of votes for all instruments given by all experts were nearly identical to the ANN results. However, in the case 11% of the sounds experts definitely disagreed with the recognition system. Only in 4% cases, votes were divided, experts were not certain of their choice. Based on the tests, it can be stated that experts did not have significant problems in deciding which instruments in the created pairs sounded better.

Analysis of Energy-Based Separation Error

The *SRR* parameter related to the energy-based error (Eqs. 1, 2) was also studied with reference to the ANN results and experts' preferences. Fig. 3.2 presents parameter D_{SRR} that describes the difference between SRR log values calculated for successfully/ unsuccessfully separated sounds taken from the set of 120 sound pairs.

$$D_{SRR} = 20 \cdot \log_{10}(SRR_S) - 20 \cdot \log_{10}(SRR_U) \qquad (3.3)$$

where: SRR_S - original to separated sound signal energy ratio (successfully separated sound), SRR_U - original to separated sound signal energy ratio (unsuccessfully separated sound).

According to the energy-based error analysis concept [7], *SRR* for successfully separated sounds should be greater than *SRR* for unsuccessfully separated ones. Thus, in the case of the inconsistency between the ANN and the SRR results, D_{SRR} should be smaller than 0, which is indicated by red bars in Fig. 3.3. Dark blue bars indicate cases where *SRR* analysis and the ANN results were consistent. Yellow
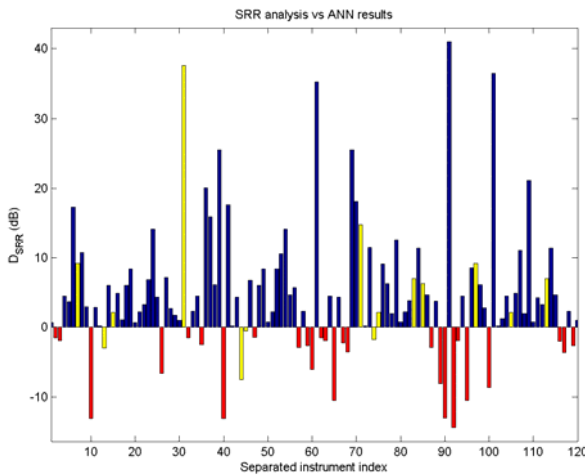
Fig. 3.2 SRR analysis results vs ANN recognition results and experts' preferences

bars indicate pairs for which experts did not agree with the ANN. It can be observed, that yellow bars of values below 0 concern four cases. In these cases experts' opinion was consistent with the SRR analysis results.

Overall, for 84 pairs out of 120, *SRR* analysis results were consistent with experts' opinions (70%), while for the ANN based recognition the consistency was equal to 84%. This shows that the ANN recognition-based separation algorithm evaluation is more similar to subjective perception than to the simple energy-based error analysis.

3.5   Conclusions

Separation methods have been developed to enable to retrieve source sounds. Artificial Neural Network has been used to evaluate effectiveness of the proposed separation algorithms. It has been shown that careful choice of features used to train the recognition system, and then utilized to identify signals enable to use the ANN to recognize the separated sounds effectively (i.e. 97.39%). Also, subjective listening tests were performed in order to determine whether the ANN-based results correspond with experts' preferences. Results contained in this paper show that experts preferred successfully separated sounds over the unsuccessfully resolved ones. In the case of 84%, a significant majority of experts agreed with the decision of the recognition algorithm. Moreover, it has been shown that the ANN results are much more consistent with experts' preferences than the energy-based error analysis results. For the set of sounds used in the presented tests, *SRR* analysis results are consistent with experts' preferences in only 70%. This allows a conclusion that the separation effectiveness in the case of the ANN-based systems, and a chosen set of features is similar to perceptual quality assessment, so it may be used to determine the performance of separation algorithms.

It may be then concluded that it is possible to effectively separate harmonic sounds from musical instrument sound mixtures without any prior knowledge of musical instruments mixed constituting the sound mixture. In addition, a properly trained ANN enables to name the separated instruments.

References

1. Dziubiński, M., Dalka, P., Kostek, B.: Estimation of Musical Sound Separation Algorithm Effectiveness Employing Neural Networks. J. Intel. Inform. Systems 24(2), 133–157 (2005)
2. Dziubiński, M., Kostek, B.: Octave Error Immune and Instantaneous Pitch Detection Algorithm. J. New Music Research 34, 273–292 (2005)
3. Dziubiński, M.: Musical Instrument Sound Separation Methods Supported by Artificial Neural Network Decision System, Ph.D. thesis, MSD, GUT (2006)

4. Gillet, O., Richard, G.: Transcription and separation of drum signals from polyphonic music. IEEE Transactions on Audio, Speech and Language Processing 16, 529–540 (2008)
5. http://ismir2009.ismir.net (Intern. Conference on Music Information Retrieval website)
6. Klapuri, A.: Multipitch analysis of polyphonic music and speech signals using an auditory model. IEEE Trans. Audio, Speech and Language Processing 16(2), 255–266 (2008)
7. Klapuri, A.: Multipitch estimation and sound separation by the spectral smoothness principle. In: Proc. IEEE ICASSP 2001, Salt Lake City, USA, pp. 3381–3384 (2001)
8. Kostek, B.: Perception-Based Data Processing in Acoustics. Springer, Berlin (2005)
9. Kostek, B., Czyzewski, A.: Representing Musical Instrument Sounds for their Automatic Classification. J. Audio Eng. Soc. 49, 768–785 (2001)
10. Kostek, B.: Applying computational intelligence to musical acoustics. Archives of Acoustics 32(3), 617–629 (2007)
11. Quatieri, T.F., McAulay, R.J.: Magnitude-only reconstruction using a sinusoidal speech model. In: Proc. IEEE ICASSP 1984, vol. 2, pp. 27.6.1–27.6.4 (1984)
12. Tzanetakis, G.: Signal Processing Methods for Music Transcription Computer Music J., 32(4), 86-88 (2008)

Chapter 4
Localization of Sound Source Direction in Real Time[*]

Eugeniusz Kornatowski

Abstract. The paper describes options to use a single point surround microphone to determine a direction of the sound source localization (SSL). A "soundfield" microphone with four transducers (capsules) characterized by cardioid responses. A unique mechanical design of the transducer results in its omnidirectional response. The microphone enables 3D sound acquisition in so called A-format. Upon further processing of signal it is possible to determine e.g. a direction of a sound source within space. The conducted experiments prove that a simple calculation algorithm is in particular feasible for a real time operation, and application of the soundfield microphone significantly simplifies mechanical design of the SSL system.

4.1 Introduction

There are advanced techniques available for recognition of surrounding environment and objects applied to the design of present day robots (including the mobile robots). Those techniques are based on application of digital processing algorithms and image analysis, but also, more and more frequently, acoustic signal analysis is utilized for this purpose [1], [2]. References [3] and [4] are particularly interesting in this respect. Each of the suggested solutions uses a matrix of microphones featuring individual design and properties. The solutions described in references are, according to their authors, showing good properties and enable relatively accurate localization of sound generating objects. Unfortunately, technical application of the published experiment results in real life are seriously limited: as in order to obtain the described and achieved results, it would be necessary

Eugeniusz Konratowski
West Pomeranian University of Technology, Faculty of Electrical Engineering,
Department of Signal Processing and Multimedia Engineering,
26 Kwietnia 10 St., 71-126 Szczecin, Poland
e-mail: `korn@zut.edu.pl`

to use microphones showing exactly the same parameters as those used by the authors in their original (prototype) solutions.

This paper shows the possibility to use typical "soundfield" microphones in order to localize sound generating objects. Such microphones are made by several manufacturers, and experiments done by the author were conducted with a use of "TetraMic Single Point Surround Sound Microphone" [5]. Standard application of "soundfield" type microphones encompasses audio signal recordings destined to play in multi-channel spatial sound systems (surround sound systems, "home theatre"). "Soundfield" microphone may be used in two ways:

- as a monophonic microphone having any directional characteristic pattern,
- as a matrix of four microphones to enable signal recording in so called A-format.

Possibility to use the latter option is shown below in this paper.

4.2 A and B Formats

The "soundfield" microphone comprises four capsules having cardioid response characteristics. An example of such microphone is illustrated in Fig. 4.1a. Microphone capsules are mounted in configuration resembling walls of a regular tetrahedron oriented as shown in Fig. 4.1b.

A microphone configuration designed in this way allows to record all needed information pertaining to the sound field in a given point of space. This microphone produces four independent outputs which could be processed further in any

a)   b)
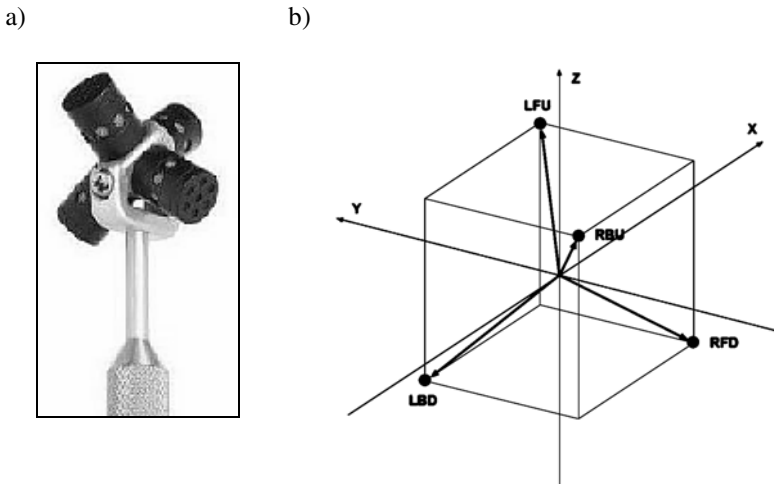
Fig. 4.1 "Soundfield" microphone: a) sample illustration of "TetraMic Single Point Surround Sound Microphone", b) spatially oriented microphone capsules

chosen way. A-format is a direct signal derived from the microphone outputs. On the other hand, B-format is substantially a studio format and a base for so called ambisonic sound system applied in the sound engineering field. It is the signal that is applied in a localization of sound source direction.

A-format signal includes four monophonic signals: LFU – left front top, RBU – right rear bottom, LBD – left rear bottom and RFD – right front bottom; all being direct outputs from the microphone capsules. In order to receive B-format signals, the A-format signals are arithmetically processed, resulting in X, Y, Z as well as W signals [6]:

$$W = v_{LFU} + v_{RBU} + v_{LBD} + v_{RFD} \qquad (4.1)$$

$$X = v_{LFU} - v_{RBU} - v_{LBD} + v_{RFD} \qquad (4.2)$$

$$Y = v_{LFU} - v_{RBU} + v_{LBD} - v_{RFD} \qquad (4.3)$$

$$Z = v_{LFU} + v_{RBU} - v_{LBD} - v_{RFD} \qquad (4.4)$$

where v_{LFU} is an output signal from LFU capsule, and so on.

These calculated WXYZ signals can be interpreted as follows: W is proportional to acoustic pressure in a central point of the microphone illustrated in Fig. 4.1a, while X, Y and Z signals are proportional to acoustic pressure values present on appropriate axes of a 3-D coordinate system: X – front - back, Y – left - right, Z – top - bottom.

"A – B format" conversion is simply a change of coordinate system from "tetrahedron" into XYZ system. Additionally, an acoustic pressure information can be obtained for a center point of the XYZ coordinate system (W).
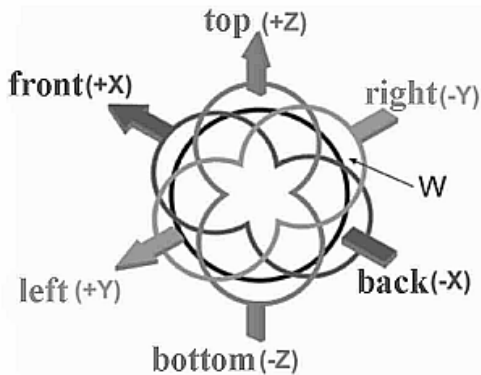
Fig. 4.2 Spatial orientation of B-format signals.

4.2.1 Analysis of "Soundfield" Microphone Operation Principle

Formulas (4.1) to (4.4) define conversion of formats A to B. Taking into consideration a geometry of tetrahedron and locations of the capsules, let us define a set of vectors describing their localization in space:

$$\mathbf{u}_{LFU} = \frac{1}{\sqrt{3}} \cdot [1 \quad 1 \quad 1]^T, \; \mathbf{u}_{RBU} = \frac{1}{\sqrt{3}} \cdot [-1 \quad -1 \quad 1]^T \tag{4.5}$$

$$\mathbf{u}_{LBD} = \frac{1}{\sqrt{3}} \cdot [-1 \quad 1 \quad -1]^T, \; \mathbf{u}_{RFD} = \frac{1}{\sqrt{3}} \cdot [1 \quad -1 \quad -1]^T$$

Assuming that a distance of each capsule from XYZ coordinate system equals $\mathbf{r}$, the locations of individual capsules are as follows:

$$\mathbf{x}_{LFU} = r \cdot \mathbf{u}_{LFU}, \quad \mathbf{x}_{RBU} = r \cdot \mathbf{u}_{RBU} \tag{4.6}$$

$$\mathbf{x}_{LBD} = r \cdot \mathbf{u}_{LBD}, \quad \mathbf{x}_{RFD} = r \cdot \mathbf{u}_{RFD}$$

If $\mathbf{d}$ is a unit vector pointing to a direction of incident wave, or, in another words a versor, and "k" is a wavelength number ($k = 2\pi/\lambda$), then one can define a wave vector in Cartesian coordinate system for acoustic three dimensional waveform as equal to:

$$\mathbf{k} = k \cdot \mathbf{d} = k \cdot \begin{bmatrix} \cos(\theta) \cdot \cos(\varphi) \\ \sin(\theta) \cdot \cos(\varphi) \\ \sin(\varphi) \end{bmatrix} \tag{4.7}$$

where θ- is an elevation angle, and φ - azimuth angle.

For acoustic wave incident from a direction $\mathbf{d}$, at point $\mathbf{x}_m$ of a 3-D space, there will be the following acoustic pressure magnitude:

$$p_m = A \cdot e^{j \cdot \omega \cdot t + \mathbf{k} \cdot \mathbf{x}_m} \tag{4.8}$$

where A is an waveform amplitude with pulsation ω, and t is time.

One could easily note the value of acoustic pressure within the centre of capsule matrix equal to $p_0 = Ae^{j\omega t}$, since $\mathbf{x}_0 = [0 \; 0 \; 0]^T$. Each capsule will immediately record the following signal:

$$p_{LFU} = A \cdot e^{j \cdot (\omega \cdot t + \mathbf{k} \cdot \mathbf{x}_{LFU})} = p_0 \cdot e^{j \cdot \mathbf{k} \cdot \mathbf{x}_{LFU}},$$

$$p_{RBU} = A \cdot e^{j \cdot (\omega \cdot t + \mathbf{k} \cdot \mathbf{x}_{RBU})} = p_0 \cdot e^{j \cdot \mathbf{k} \cdot \mathbf{x}_{RBU}},$$

$$p_{LBD} = A \cdot e^{j \cdot (\omega \cdot t + \mathbf{k} \cdot \mathbf{x}_{LBD})} = p_0 \cdot e^{j \cdot \mathbf{k} \cdot \mathbf{x}_{LBD}}, \tag{4.9}$$

$$p_{RFD} = A \cdot e^{j \cdot (\omega \cdot t + \mathbf{k} \cdot \mathbf{x}_{RFD})} = p_0 \cdot e^{j \cdot \mathbf{k} \cdot \mathbf{x}_{RFD}}$$

It is known from the reference [7] that a single equation defining an output signal is valid for a single point pressure gradient type microphone (or single capsule) response:

$$v = G \cdot \frac{1}{a+b} \cdot [a + b \cdot \mathbf{u} \cdot \mathbf{d}] \cdot p \qquad (4.10)$$

where p – acoustic pressure of wave incident from the direction $\mathbf{d}$, while G, a, b – constants characteristic for microphones with a defined directional response.

Thus, the equations determining A-format outputs for microphones from Fig. 4.1 will be:

$$v_{LFU} = G \cdot \frac{1}{a+b} \cdot [a + b \cdot \mathbf{u}_{LFU} \cdot \mathbf{d}] \cdot p_{LFU}, \qquad (2.11)$$

$$v_{RBU} = G \cdot \frac{1}{a+b} \cdot [a + b \cdot \mathbf{u}_{RBU} \cdot \mathbf{d}] \cdot p_{RBU},$$

$$v_{LBD} = G \cdot \frac{1}{a+b} \cdot [a + b \cdot \mathbf{u}_{LBD} \cdot \mathbf{d}] \cdot p_{LBD},$$

$$v_{RFD} = G \cdot \frac{1}{a+b} \cdot [a + b \cdot \mathbf{u}_{RFD} \cdot \mathbf{d}] \cdot p_{RFD}$$

Those equations for B-format, while considering equations (4.1) and (4.11), are:

$$W = G \cdot \frac{1}{a+b} \cdot [a \cdot (p_{LFU} + p_{RBU} + p_{LBD} + p_{RFD}) + \\ + b \cdot (p_{LFU} \cdot \mathbf{u}_{LFU} + p_{RBU} \cdot \mathbf{u}_{RBU} + p_{LBD} \cdot \mathbf{u}_{LBD} + p_{RFD} \cdot \mathbf{u}_{RFD}) \cdot \mathbf{d}] \qquad (4.12)$$

and further on:

$$W = G \cdot \frac{1}{a+b} \cdot [4 \cdot a \cdot p_0 + b \cdot p_0 \cdot (\mathbf{u}_{LFU} + \mathbf{u}_{RBU} + \mathbf{u}_{LBD} + \mathbf{u}_{RFD}) \cdot \mathbf{d}] = \\ = G \cdot \frac{1}{a+b} \cdot [4 \cdot a \cdot p_0 + b \cdot p_0 \cdot \mathbf{0} \cdot \mathbf{d}] = \\ = \frac{4 \cdot G \cdot a}{a+b} \cdot p_0 \qquad (4.13)$$

From (4.2) and (4.11) we may obtain the following equations for signal X:

$$X = G \cdot \frac{1}{a+b} \cdot [a \cdot (p_{LFU} - p_{RBU} - p_{LBD} + p_{RFD}) + \\ + b \cdot (p_{LFU} \cdot \mathbf{u}_{LFU} - p_{RBU} \cdot \mathbf{u}_{RBU} - p_{LBD} \cdot \mathbf{u}_{LBD} + p_{RFD} \cdot \mathbf{u}_{RFD}) \cdot \mathbf{d}] \qquad (4.14)$$

and finally:

$$X = G \cdot \frac{1}{a+b} \cdot [a \cdot (2 \cdot p_0 - 2 \cdot p_0) + b \cdot p_0 \cdot (\mathbf{u}_{LFU} - \mathbf{u}_{RBU} - \mathbf{u}_{LBD} + \mathbf{u}_{RFD}) \cdot \mathbf{d}] =$$

$$= G \cdot \frac{1}{a+b} \cdot (\mathbf{u}_{LFU} - \mathbf{u}_{RBU} - \mathbf{u}_{LBD} + \mathbf{u}_{RFD}) \cdot \mathbf{d} \cdot p_0 = \qquad (4.15)$$

$$= \frac{4}{\sqrt{3}} \frac{G \cdot b}{a+b} \cdot \cos(\theta) \cdot \cos(\varphi) \cdot p_0$$

Similarly, it is possible to determine signals Y and Z from Equations (4.3), (4.4) and (4.11):

$$Y = \frac{4}{\sqrt{3}} \frac{G \cdot b}{a+b} \cdot \sin(\theta) \cdot \cos(\varphi) \cdot p_0 \qquad (4.16)$$

$$Z = \frac{4}{\sqrt{3}} \frac{G \cdot b}{a+b} \cdot \sin(\theta) \cdot p_0 \qquad (4.17)$$

Pressure gradient capsules were used in the examined "soundfield" microphone. Therefore, based on [7], and assuming $a = b = 1$, the following can be derived from Equations 4.13, 4.15, 4.16 and 4.17:

$$X = \frac{W}{\sqrt{3}} \cdot \cos(\theta) \cdot \cos(\varphi), \quad Y = \frac{W}{\sqrt{3}} \cdot \sin(\theta) \cdot \cos(\varphi), \quad Z = \frac{W}{\sqrt{3}} \cdot \sin(\theta) \qquad (4.18)$$

As it transpires from the latter formulas, it is quite easy to determine a sound source direction (azimuth and elevation) if one knows signal W and a single pair from X, Y, Z set of values.

4.3   Experimental Testing

The experiments were conducted in a 5 x 6 x 3 m room, with a reverb time RT60 equal to 0.8 s. "Soundfield" type microphone was located at about 1.5 m above the floor. A loudspeaker playing signal obtained from a pink noise test tone generator was used as a sound source.Distance to speaker from the microphone was set at around 1.5 m. The A-format signals were registered using a digital Zaxcom DEVA 5.8 recorder.

The 10 s duration time samples were recorded with sampling frequency of 96 kHz and 24 bit resolution for various sound source locations. During the first phase of experiment the sound source was immobilized. Upon conversion from A-format to B-format signals W, X, Y, Z were divided into 250 ms duration time frames. The elevation angle and azimuth was calculated for each set of these frames, and subsequently the results were averaged for a total time of recording (10s). A sample result of one measurement is shown in Fig. 4.4.

Fig. 4.3 A room where the measurements were conducted: located at the Sound and Ambiophonics Engineering Laboratory, West Pomeranian University of Technology, Szczecin, Poland.

Table 4.1 Average of direction errors of one sound source localization (A - azimuth, E - elevation)

Sound Source Direction	$A=0^0$, $E=90^0$	$A=90^0$ $E=45^0$	$A=45^0$ $E=-45^0$	$A=-45^0$, $E=90^0$
Average Direction Errors	$\varepsilon_A=4^0$, $\varepsilon_E=5^0$	$\varepsilon_A=6^0$, $\varepsilon_E=5^0$	$\varepsilon_A=3^0$, $\varepsilon_E=4^0$	$\varepsilon_A=4^0$, $\varepsilon_E=6^0$

The Fig. 4.4 shows directly measurement results. Fig. 4.4b presents an image including thresholding segmentation [8]. This operation was applied in order to increase accuracy of azimuth and elevation angle definitions.

Experimental results for several sound source locations in relation to microphone were presented in Table 4.1.

The results of tests shown in Table 4.1 were accomplished by holding the sound source in one location while the microphone location was changed. For that reason it was possible to determine a real location of the sound source in relation to microphone in relatively accurate way. It should be emphasized that the azimuth and elevation errors were obtained based on numerical calculations and not by analyzing the graphs (Fig. 4.4).

By using the "soundfield" microphone it is also possible to successfully monitor the movement of a sound source in real time. The Fig. 4.5 shows the sound source location change within a timeframe of about 300 s.

a)
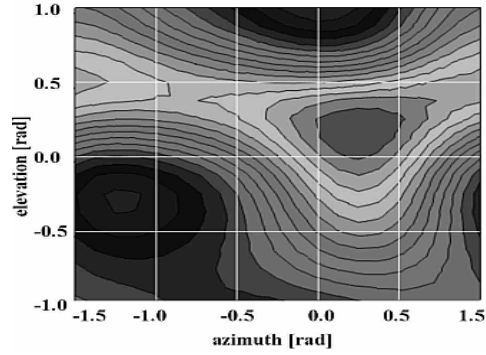
b)
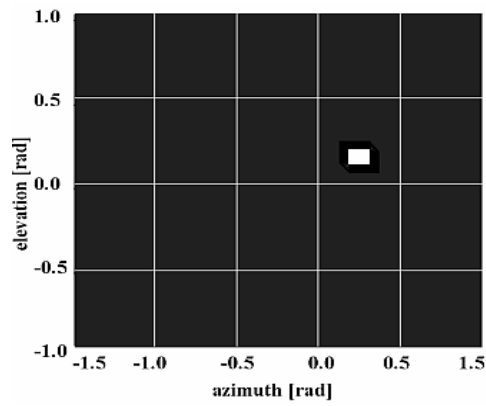
Fig. 4.4 Result for localization of a single sound source in 3-D coordinate system (plan view)
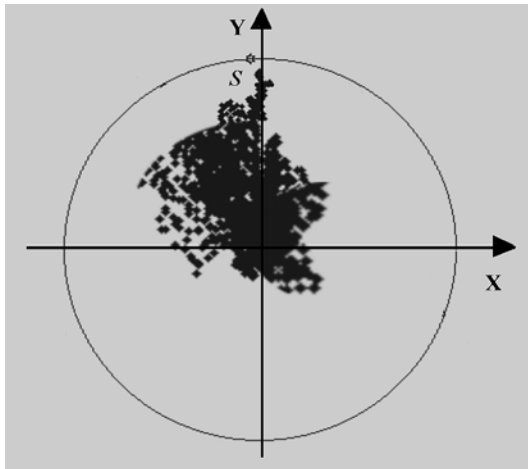
Fig. 4.5 Monitoring of sound source location in horizontal plane

In this case (Fig. 4.5) distances of particular points from a coordinate system centre are proportional to values of acoustic pressure. The graph was plotted for the source moving through a planar surface. "Soundfield" microphone was placed in the centre of coordinate system. Marker "S" denotes a current direction of the sound source. Variations of point distances from XY system centre are reflecting the changes of sound source level or changes of its location in relation to the microphone. The sound source volume level was constant during the experiment, while its distance from the microphone was varied.

4.4 Conclusions

The paper offers a system for determination of sound source direction with a use of soundfield microphone. This solution may be useful in design of robots, e.g. as one of the components for a recognition of surroundings. Obtained results of simulation tests are very encouraging. The errors shown in Table 4.1 prove that the system is able to localize the sound source direction as well as human ear. Based on sense of hearing, a human being is able to determine the azimuth and elevation within accuracy of 6° to 10°. Unfortunately, a single soundfield microphone cannot determine a distance to the sound source. A reason for that is connected to relatively small distances between capsules; therefore application of time delay of arrival (TDOA) method [9] is not effective. However, by adding one "soundfield" microphone to the system (or even a single capsule having a omnidirectional response) one could determine a distance of a sound source within 3-D space. The condition is that the distance of that additional microphone from a soundfield microphone was 15 cm.

References

1. Li, H., Yoshiara, T., Zhao, Q., Watanabe, T., Huang, J.: A spatial sound localization system for mobile robots. IEEE Volume, Issue, Trans. Instrum. and Meas., 1–6 (2007)
2. Valian, J.M.: Robust Sound Source Localization Using a Microphone Array on a Mobile Robot. In: IEEE/RSJ International Conference on Intelligent Robotics and Systems, pp. 1228–1233 (2003)
3. Martinson, E.: Hiding the Acoustic Signature of a Mobile Robot. In: IEEE Int. Conf. on Intelligent Robots and Systems (IROS), San Diego (2007)
4. Martinson, E., Arkin, R.C.: Noise Maps for Acoustically Sensitive Navigation. In: Proceedings of SPIE, vol. 5609 (2004)
5. Core Sound, http://www.core-sound.com/TetraMic/1.php
6. Farrar, K.: Soundfield Microphone. Wireless World 85(1526), 48–50 (1979)
7. Jamroz, A.: The Design and Use of a Double Cardioid Stereophonic Microphone. Journal of the Audio Engineering Society 8(2), 100–104 (1960)
8. Saberi, K., Perrott, D.R.: Lateralization threshold obtained under conditions in which the precedence effect is assumed to operate. J. Acoust. Soc. Am. 87, 1732–1737 (1990)
9. Brandstein, M., Silverman, H.: A practical methodology for speech source localization with microphone arrays. Comput., Speech Ing. 11(2), 91–126 (1997)

Chapter 5
Dangerous Sound Event Recognition Using Support Vector Machine Classifiers

Kuba Łopatka, Paweł Zwan, and Andrzej Czyżewski

Abstract. A method of recognizing events connected to danger based on their acoustic representation through Support Vector Machine classification is presented. The method proposed is particularly useful in an automatic surveillance system. The set of 28 parameters used in the classifier consists of dedicated parameters and MPEG-7 features. Methods for parameter calculation are presented, as well as a design of SVM model used for classification. The performance of the classifier was tested on a set of 372 example sounds, yielding high accuracy.

5.1   Introduction

In automatic surveillance systems sound analysis is often used as a reinforcement of video-based dangerous event detection. It is possible to recognize hazardous events by the analysis of their acoustic representation. To achieve this goal, a feature extraction and classification technique must be established. In the literature several solutions for sound recognition employing different classification algorithms are described [2,6,10,12].This paper focuses on methods for feature extraction and the use of Support Vector classification as an efficient tool for classifying sound events. The results of classification are presented, which proves that the described algorithms are useful in an automatic security surveillance system.

A sample of a sound recognition system, as shown in Fig. 5.1.1 consists of 5 basic elements – event detection, parameter calculation, model training, classification and decision module. It is assumed that the data given at the input of the system is already in digital domain. Originally it may come straight from the microphone or it can be streamed via network. Next, after detection of a sound event, parameters are extracted from the signal. The calculated features are used for the classification. The classifier is trained with the use of examples from the event database. The output of the system gives one of two messages, one being regarding no dangerous situation and the other alarming the operator about a dangerous event being detected.

Kuba Łopatka, Paweł Zwan, and Andrzej Czyżewski
Gdansk University of Technology, Multimedia Systems Department
e-mail: {klopatka,zwan,ac}@sound.eti.pg.gda.pl
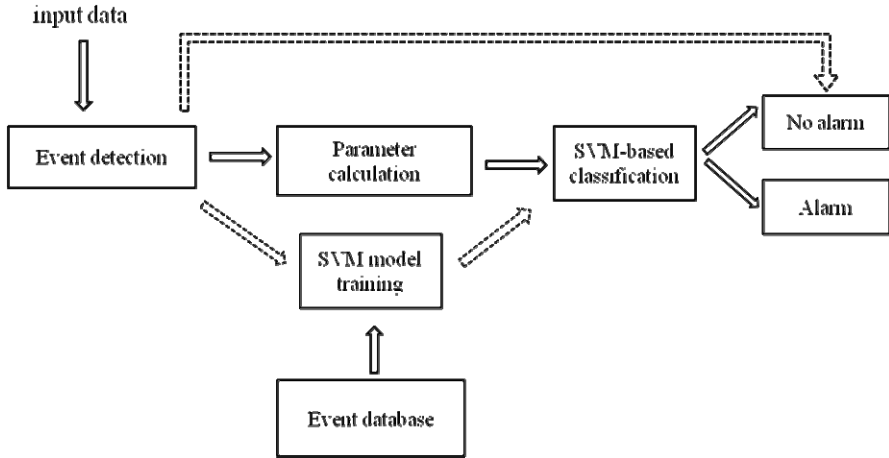
Fig. 5.1.1 Block diagram of the sound recognition system

5.2   Feature Extraction

The described sound recognition system described uses 28 parameters selected experimentally. The set of parameters consists of:
– 9 energy parameters,
– 2 transient-sensitive parameters,
– 17 MPEG-7 features.

The parameters were calculated for all the examples from a training set, consisting of 372 sound files, divided in 5 classes:
– Class 1: explosion – 16 objects,
– Class 2: broken glass – 120 objects ,
– Class 3: gunshot – 157 objects,
– Class 4: scream 26 objects,
– Class 5: other sounds, not related to danger – 51 objects.
The vectors of parameters defined above were used as a training set to build the SVM model.

5.2.1   Energy-Based Parameters

The energy-based parameters were proposed in the literature [13] and are connected to energy in certain sub-bands of the spectrum. The parameters are calculated according to the formula (5.1):

$$p = \frac{\sum_{n_{f1}}^{n_{f2}} P(n)}{\sum_{n_{f1'}}^{n_{f2'}} P(n)} = \frac{energy\ in\ band\ 1}{energy\ in\ band\ 2} \tag{5.1}$$

where:

n_{f1}, n_{f2}, n_{f1}', n_{f2}' – indices of respective band limits,

$P(n)$ – power spectrum of the signal,

For some of the features the denominator of Eq. 5.1 relates to the entire spectrum. In this case, a parameter expresses, how much of the signal's energy is cumulated in a given frequency sub-band. The boundary frequencies of the bands were set after series of experiments and examinations on the example signals from the event database.

5.2.2 Transient-Sensitive Parameters

There are some dangerous events that are related to rapid changes in the signal (e.g. gunshots, explosions, or broken glasses). Thus, it is reasonable to recognize these events using transient-sensitive parameters.

Various algorithms for detection of transient states may be found in literature [4]. There are several issues related to this problem, thus they are discussed below.

1. The algorithm should have a low complexity. Therefore, the detection algorithm should not introduce high delays into the processing scheme
2. The algorithm should be accurate. Too many false positive detections will result in false decisions
3. Many transient detection algorithms operate in the off-line mode and they require access to the complete signal [4], which is not possible in case of real-time processing. Therefore, the detection algorithm should be able to perform decisions using only information obtained from the current and previous signal frames
4. Most of transient detection methods found in literature are dedicated to speech or musical signals, some of them are even named 'musical onset detection algorithms'. These methods are not equally accurate for dangerous sound events detection

Therefore, it is not recommended to simply use a detection algorithm that was proved to be accurate for speech or musical signals and implement it for purposes related to the topic of this paper.

The transient detection algorithm described below, was tested by the authors and selected for use in the proposed application. The algorithm proved to provide sufficiently effective and accurate method for transient detection in signals related to security threads. Two additional dedicated parameters were defined. First of them, called *transient_length* is calculated according to equation (5.2). This parameter is related to the length of the period in which rapid signal changes are observed. Transient detection is based on finding of the maximum change in signal value by examining the first-order difference of the signal.

$$tr\_length = n_{|d(n)| < d_{thr}} - n_{d(n) = \max(d)} \qquad (5.2)$$

where:

  n – sample's index,
  $d(n)$ – fist-order difference of the signal,
  d_{thr} – threshold value.

If the first-order difference maintains below the threshold value, the rapid changes in the signal are considered terminated. In described studies the threshold was set to 5% of the maximum of the signal's difference. This threshold value was set experimentally using numerous initial experiments.

The second parameter – *transient_rate* – defined by formula (5.3) represents the ratio (in decibels) of the small signal fragment, which contains the detected transient, to the fragment that follows it.

$$ tr\_rate = 10*\log\left(\frac{\sum_{n_{tr}-M/2}^{n_{tr}+M/2-1}|x(n)|^2}{\sum_{n_{tr}+h+-M/2}^{n_{tr}+h+M/2-1}|x(n)|^2} \right) \qquad (5.3) $$

where:

  n_{tr} – index of the detected transient,
  M – length (in samples) of the fragment around the transient,
  h – step to the next analyzed fragment.

Although many classes of sounds related to an alarm represent rapid changes detected as transients, the transient parameters provided enable the system to differentiate between them. For classes like gunshot and explosion the *transient_rate* parameter is greater than for broken glass. However, the time, in which transients are observed, is much shorter.

5.2.3 MPEG-7 Features

MPEG-7 description is a widely used general method of signal parameterization. It is used in many applications related to automatic sound recognition [2, 6, 7, 12]. Therefore, in the research presented those parameters have been also included into a feature vector. The audio descriptors, which could be useful for sound event recognition, were chosen on the basis of performance tests and statistical analysis. The set of 17 MPEG-7 features that were included in the feature vector consists of:

− *Log Attack Time (LAT)*,
− *Audio Spectrum Envelope* calculated in the following bands: ASE_{29}: 6727Hz-8000Hz, ASE_{30}: 8000Hz-9513HZ, ASE_{31}: 9513Hz-13454Hz, ASE_{32}: 11314Hz-13454Hz, ASE_{33}:13454Hz-16000Hz,
− *Spectral Flatness Measure* calculated in the following bands: SFM_{10}: 1189Hz-1414Hz, SFM_{15}: 2828Hz-3363Hz, SFM_{16}: 3363Hz-4000Hz, SFM_{19}: 5657Hz-6727Hz, SFM_{20}: 6727Hz-8000Hz, SFM_{21}: 8000Hz-9514Hz, SFM_{22}: 9514Hz-11314Hz,

- *Spectral Flatness Measure* mean variance (*SFM$_{mv}$*),
- *Audio Spectrum Spread* and *Audio Spectrum Spread* variance (*ASS, ASS$_v$*),
- *Audio Spectrum Centroid* variance (*ASS$_v$*).

The *ASE* parameters are related to the power spectrum of the signal. They are therefore similar to the energy parameters defined in Sec. 5.1.1, but the frequency resolution is better, thanks to narrower bands. The *SFM* descriptors show the ability to reflect harmonicity of the signal, which is a strong basis for distinction between e.g. screams and other sounds.

5.3 Building SVM Model

The Support Vector Machine classifier was used in the experiments. The development of this algorithm needs a learning step. Part of the event database was used for the training of the algorithm, another part for validation of its generalization capacities. This is described in a more in detailed way in next sections. The classifier was built and tested in two environments:

- *WEKA* data mining software [14],
- *C++* program implementing *LIBSVM* [3].

In case of *WEKA,* the algorithm for classification was *Sequential Minimal Optimization*, providing a fast implementation of Machine Learning [9].

5.3.1 Principles of Support Vector Machine Classification

As described in literature [3, 5, 9, 10, 11], the idea of Support Vector Machine classification is to find a hyperplane in $\mathbf{R}^n$ space that separates best the data vectors with positive and negative labels. These vectors represent positive and negative examples included in the training set. If the hyperplane is described by a formula *w·x+b=0,* where *x* is the data vector and *w* represents the weights of the hyperplane, the separation may be described as in Eq. (5.4):

$$\begin{cases} \mathbf{w} \cdot \mathbf{x_i} + b > 1 & \textit{if} \quad y_i = 1 \\ \mathbf{w} \cdot \mathbf{x_i} + b < -1 & \textit{if} \quad y_i = -1 \end{cases} \tag{5.4}$$

The separation of positive and negative examples from the training set is illustrated in Fig 5.3.1. The aim of SVM model training is to find such weight vector **w**, for that the margin between the positive and negative vectors and the hyperplane is maximized. For this task, only the vectors lying closest to the hyperplane are used, i.e. *support vectors.*

The formula (5.2.1) and Fig. 5.2.1 both relate to the ideal case, in which all positive examples from the training set fall into one side of the hyperplane and all negative examples fall to the other one. In most cases, the scenario is more complicated. The first problem one can encounter is that some input vectors fall to the
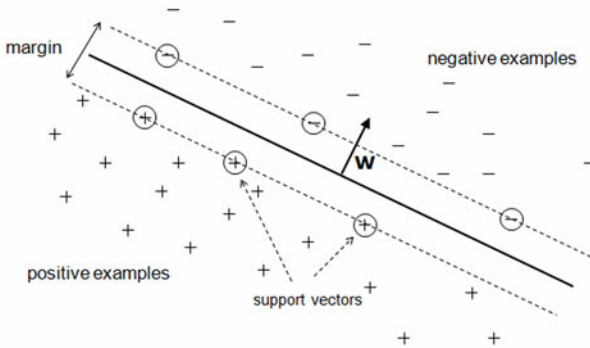
Fig. 5.3.1 Separation of training data with an optimal hyperplane

opposite side of the hyperplane. In this case, a cost factor c is introduced. The greater the cost parameter, the less likely the training algorithm is to ignore vectors, which fail to be separated with the optimum hyperplane. The second problem is non-linear separability. If the data is not linearly separable, no hyperplane can be found to efficiently distinguish between positive and negative examples. The problem can be solved, using kernel methods. Kernel functions are used to transform the coordinates of the data set, so that a non-linear problem is transformed into a linear problem. In other words, the kernel trick is mapping the data vector in $\mathbf{R}^n$ space into another k-dimensional space. The most frequently used kernel functions in SVM applications are:

− radial-basis function (*RBF*),
− polynomial kernel function,
− sigmoid function.

The example use of a kernel function is presented in Fig. 5.3.2. The non-linearly separable set of data is transformed, using a polynomial kernel of 2-nd degree, to a linearly separable set.
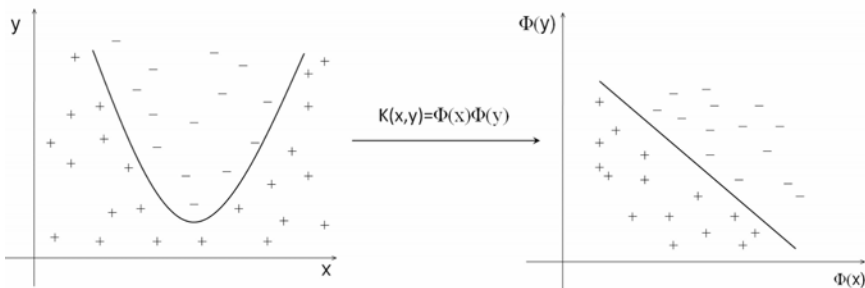
Fig. 5.3.2 Kernel transformation of non-linearly separable data

5.3.2 Parameters of the Model

In this study a polynomial kernel given by the formula (5.2.2) was used. The choice of the kernel function and its parameters was based on the classification accuracy achieved in experiments.

$$\mathbf{K}(\mathbf{x}, \mathbf{y}) = (\gamma \cdot \langle \mathbf{x}, \mathbf{y} \rangle + C_0)^d \qquad (5.5)$$

where: $\langle \mathbf{x}, \mathbf{y} \rangle$ is the dot product of vectors $\mathbf{x}$ and $\mathbf{y}$.

The value parameters of the Equation (5.2.2) that have been used for the training the SVM classifier are:

–   multiplier $\gamma = 1$,
–   degree $d = 2$,
–   constant $C_0 = 5$.

As a result, a multiclass SVM model was built, consisting of *151* total support vectors. It is also a probability model, which provides all outputs with corresponding probability values. . This information can be used to improve the efficiency of the classification.

5.4 Classification Results

The classifier was tested upon the training set in 10-fold cross-validation mode. Therefore, a model was built 10 times, each time leaving 10% of the data out of the training set. In Table 5.4.1 the results of an experiment conducted in *WEKA,* are presented. The average accuracy of the classifier is 97,85%, as only 8 of 372 instances were misclassified.

Table 5.4.1 Confusion matrix of a *SVM* classifier implemented in *WEKA*

explosion	broken glass	gunshot	scream	Other	
93,75%	0%	6,25%	0%	0%	explosion
0%	**96,72%**	3,28%	0%	0%	broken glass
0%	1,27%	**98,73%**	0%	0%	gunshot
0%	0%	0%	**96,15%**	3,85%	scream
0%	0%	0%	0%	**100%**	other

Tests using LIBSVM provided comparably good results. The classifier was again tested in 10-fold cross-validation and yielded 96,49% correctly classified objects.

An additional experiment was carried using LIBSVM to determine the noise influence on the classification efficiency. The model was trained on a training set,

consisting of 80% of the benchmark sound data (chosen randomly). The remaining 65 objects were used as a test set. Next, Gaussian white noise was added to test data at different levels. The results are presented in Table 5.4.2.

The experiment proves that noise influence on the classification efficiency is significant. Addition of the white noise changes the power spectrum of the signal, thus changing values of some parameters. Therefore, the ratio of correctly classified objects decreases along with the decrease of the *SNR ratio*. In a real environment noise is always present. If the noise level is high, it may be necessary to apply noise reduction methods before the classification.

Table 5.4.2 Accuracy of sound recognition for different *Signal-to-Noise-Ratio* values

SNR [dB]	correctly classified	incorrectly classified	**accuracy**
no added noise	62	3	**95,38%**
30	59	6	**90,77%**
25	59	6	**90,77%**
20	59	6	**90,77%**
15	56	9	**86,15%**
10	48	17	**73,85%**

5.5   Conclusions

Support Vector Machine classifier proved itself to be an efficient tool for the recognition of sounds related to danger. However, it must be stressed that a classifier trained on sound files is not the complete solution for a sound recognition system which must successfully operate in real-life conditions. Experiments that were carried out on this research were based on audio files, recorded specially for this purpose. In development of the research, the system has to be prepared for real-time operation in real environment by means of developing robust sound event detection algorithms. Nevertheless, the proposed signal feature extraction method, combined with Support Vector Machine applied to sound classification could serve as a tool for dangerous sound event recognition in surveillance systems.

References

1. Burges, C.: A tutorial on Support Vector Machines for pattern recognition. Data Mining and Knowledge Discovery 2, 121–167 (1998)
2. Casey, M.: General sound classification and similarity in MPEG-7. Organized Sound Archive 6(2), 153–164 (2001)

3. Chang, C., Lin, J.: LIBSVM: a library for Support Vector Machines. Dept. of Computer Science. National Taiwan University, Taipei (2009)
4. Daudet, L.: A Review on Techniques for the Extraction of Transients in Musical Signals. In: Proc. CMMR 2005, Pisa (2005)
5. Hsu, W., Chang, C., Lin, J.: A practical guide to support vector classification. Dept. of Computer Science. National Taiwan university, Taipei
6. Kim, H., Moreau, N., Sikora, T.: Audio classification based on MPEG-7 spectral basis representations. IEEE Trans. on Circuits and Systems for Video Technology 14, 716–725 (2004)
7. Kostek, B., Zwan, P., Dziubinski, M.: Musical Sound Parameters Revisited. In: Music Acoustic Conference, Stockholm, pp. 623–626 (2003)
8. Nitalampiras, S., Potamitis, I., Fakotakis, N.: On acoustic surveillance of hazardous situations. In: IEEE Int. Conf. on Acoustics, Speech and Signal Processing, pp. 165-168 (2009)
9. Platt, J.: Sequential minimal optimization: a fast algorithm for training Support Vector Machines. Technical Report MSR-TR-98-14 Microsoft Research (1998)
10. Rabaoui, A., Kadri, H., Lachiri, Z., Ellouze, N.: Using robust features with multi-class SVMs to classify noisy sounds. In: Int. Symp. on Communications, Control and signal Proc., Malta
11. Vapnik, V.: Statistical Learning Theory. Wiley-Interscience, New York (1998)
12. Wang, J.F., Wang, J.C., Huang, T., Hsu, C.: Home environmental sound recognition based on MPEG-7 features. In: 2003 IEEE Symp. on Mico-NanoMechatronics and Human Science, vol. 2, pp. 682–685 (2003)
13. Zwan, P., Czyzewski, A.: Automatic sound recognition for security purposes. In: Proc. 124th Audio Engineering Society Convention Amsterdam (2008)
14. Waikato Environment for Knowledge Analysis,
    http://www.cs.waikato.ac.nz/ml/weka/

Chapter 6
Noise Tolerant Community Detection Using a Mixed Graph Model*

Anita Keszler, Akos Kiss, and Tamas Sziranyi

Abstract. In this paper a new concept is proposed for finding communities in a social network based on a mixed graph theoretic model of a standard and a bipartite graph. Compared to previous methods the introduced algorithm has the advantage of noise-tolerance and is applicable independently of the size of the clusters in the graph. The cluster core-mining method is based on a modified MST algorithm. Clustering incomplete data is done by using bipartite graphs and fuzzy membership functions.

6.1 Introduction

The interest for social network analysis has recently been increasing radically. The importance of network analysis has grown due to the wide range of application areas. We are trying to understand how our society works, how we form the structure of a network containing millions of nodes. Besides scientifically interesting problems, numerous questions concerning industrial applications also arise. From crime fighting applications [2], to shopping behaviour or web mining [7], network analysis occurs in almost every field of our lives . The most widely used and most powerful techniques are based on graph theory. Linear algebric solutions [7] occur quite often as well, however, graphs have the advantage of transparency and wider possibilities in case of several types of problems.

First we will present a short overview of the used graph models in social network analysis including the most prevalent definitions of a cluster. After that we will present a new algorithm based on a mixed model of a standard and a bipartite graph. The innovation lies partly in the mixed model, and partly in the noise tolerance of

---

Anita Keszler, Akos Kiss, and Tamas Sziranyi
Computer and Automation Research Institute, MTA SZTAKI, Budapest, Hungary
e-mail: keszler@sztaki.com,akos.kiss@sztaki.com,
       sziranyi@sztaki.com

the method. In most applications it is supposed that we have complete information about the society we analyse. In our algorithm we suppose that part of the dataset is damaged or missing. The goal is to group people having common features together based on the available data, and through this, to predict the missing information. Later on we will suppose that the reader is familiar with the basics of graph theory.

6.2 Overview of Previous Methods

In this section we are going to give a brief summary of the typical graph based methods in this area. The overview is based on the questions emerging during the planning and application of network analyzer methods.

6.2.1 What to Model by a Graph?

In case of network analysis problems the used models can mostly be categorized in two main different types.

Some problems can be easily modeled by a graph in which the vertices correspond to the objects we want to work with, and the edges show the connection between them. It might be enough for the edges to symbolize only whether the objects have a connection or not, like in [4] or extended with direction, like in [7].The weights can describe the strength [5] of the connections. Semantic graphs are also useful for modeling social networks. We can get more tools to represent the data structure by giving semantic labels to vertices or edges.

6.2.2 How to Build up the Graph?

If we have information about the connection between any pair of objects and we model these relations, we can build a graph with complete information. This does not necessarily mean that we have all the edges in the graph. For example the information can be complete if we know whether an object has a property or not. However usually only the positive answers will be signed by edges as in [12].

In several tasks it is not necessary to know the relation of each object pair. In case of weighted graphs the most common approach is to use the so called k nearest neighborhood graphs [17]. A different approach is based on the minimum weight spanning tree of the graph [9]. In [15] the authors use the MST of the graph built from the objects for clustering.

6.2.3 How to Define a Group?

In case of clustering problems an obvious decision to be made is to define the *cluster*. In this section we will give the most frequently used aspects of defining similarity in other application areas as well to see their usability in network analysis.

A similar document mining algorithm is introduced in [3] using a weighted bipartite graph. The goal is to find the similar documents to a query object where similarity depends on the co-occurances of terms in the documents. As most similarity relations are not transitive,the objects similar to a query object are not necessarily similar to each other. As a consequence finding pairwise similar objects is different problem.

The previously mentioned hypothesis is that if the objects are similar to each other there is a high chance for the corresponding vertices in a graph to be in connection. Therefore similar objects often appear as a dense region of the graph. There are several ways to define density (and through this the cluster)in a graph, in the following we shall see a few possible variations. When defining a cluster the idea of mining cliques and bicliques arises naturally. However, approximations are usually more appropriate for real world applications, partly because real world samples are often imperfect and partly because most of the problems related to finding cliques or bicliques are NP-hard [6]. In general NP-hard problems may become solvable in polinomial time in case of restrictions on the input, like in [1] for the maximum weight clique problem. However for bipartite graphs it is also proven that for a wide range of edge weights even finding *good* approximations of the maximum weight biclique in polinomial time is impossible to do unless RP=NP [14].

The most frequently used solution is to search for *dense* subgraphs (or *quasicliques*) as clusters. Yet we still have a lot of freedom to define density. A common approach in standard graph models is to define a dense subgraph $G' = (V', E')$ in a graph $G = (V, E)$ in the following way:

$$| E'(G') | \geq \gamma_{dense} \binom{| V'(G') |}{2} \tag{6.1}$$

where γ_{dense} is a density threshold larger than γ, the average density of the graph.

In case of bipartite graphs the corresponding density function is not that obvious. In [12] the authors have a bipartite graph, with objects and their features. They define *quasi-bicliques* as follows. Let $G = (A, B, E)$ be a bipartite graph and $0 \leq \varepsilon \leq 1$. A $G' = (A', B', E')$ subgraph is a quasi-biclique, if

$$\forall v_i \in A', |N'(v_i)'| \geq (1 - \varepsilon) \cdot |B'|; \text{where} N'(v_i) \in B'. \tag{6.2}$$

That is every vertex in A' neighbours at least $(1 - \varepsilon)$ fraction of the vertex set B'. A very similar idea occurs in [10], where each vertex of A' can be disconnected from at most δ percent of the vertices in B' and conversely for B'. So in contrast with the previous one, it is a symmetric restriction.

In social community detection problems a more complex model of a cluster became popular. Besides the dense inner connections, communities can be recognized by the sparse outer connections as well. This property is measured by the *conductance*. The condutance value belongs to a cut in a graph. Let P and R denote the vertex sets on the sides of the cut. The conductance is the ratio of the number of edges crossing the cut to the minimum of the volume of P and R, where the volume of a vertex set is the number of edges with at least one end vertex in the set.

6.3   The New Algorithm

In this section we are going to introduce a new algorithm for efficient society mining
with partially available information. The developed algorithm is based on a mixed
model of a standard and a bipartite graph.

1. Build a standard graph of the people, we have all the information about.
2. Find the potential cores of the clusters in this graph.
3. Build a bipartite graph of the whole dataset.
4. Re-cluster people corresponding to the complete information vectors.
5. Cluster people corresponding to damaged information vectors.

In the following sections we are going to introduce the algorithm in details.

6.3.1   The Model

Let us suppose that the input dataset D looks as follows. We store information about
people that can be represented as a vector $\mu_i = \{\mu_{i1}, \mu_{i2}, ..., \mu_{ir}\}$ for a person $v_i \in V$,
$|V| = n$. V represents the set of people, and r is the dimension of the feature vectors.
Each element of the vector symbolizes a feature, for example whether the given
person has a dog or not. The edge weights ($-1 \leq \mu_{ij} \leq 1$), depending on the appli-
cation, may signify the strength of the connection, or, for instance, the probability
of the existence of the connection. (Joining this notion to semantic graphs, further
possibilities are opened.)

According to this, the whole dataset is represented by a matrix $M = [\mu_1; \mu_2; ...; \mu_n]$,
where each row corresponds to a member of the society.

In case of social networks, the trivial choice is to use a standard graph. However if
we have more detailed data about the members of the society, bipartite graphs offer
a better solution. With the latter model, the application is not necessarily limited
to finding people knowing each other, there is also an opportunity to detect groups
with similar behaviour.

The idea is to use both models parallelly. The weighted version of the previ-
ously introduced standard graph strucutre can be applied simply by using the Eu-
clidean distance of the vectors. Of course, other types of distance metric would
also be appropriate [13], the point is that the smaller the distance, the more sim-
ilar the vectors are. The Euclidean distance of two people is $dist_{wEucl}(v_i, v_j) =$
$\sqrt{\sum_{k=1}^{r} w_k \left(\mu_{ik}^2 - \mu_{jk}^2 \right)}$, where w_i is a weight function corresponding to the ith di-
mension.

In the bipartite graph model nodes represent not only objects, but features or
parameters as well. The set of nodes is divided into two node classes. One contains
the members of the society, the other contains the features.

Part of the advantages of a bipartite model can be presented by typical applica-
tions for protein-protein interaction mining. Most of these algorithms are based on
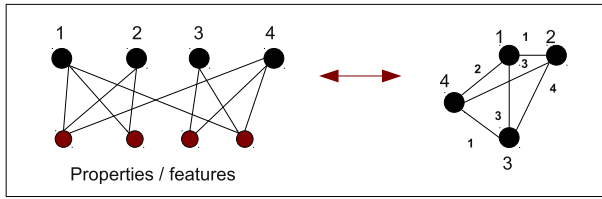
Fig. 6.1 Parallelly used standard and bipartite graph model

bipartite subgraph mining. [16][8] This approach could be an important part of a model to be used in social network analysis as well. The question might arise that if the model given in the [11] corresponds to the model we need to a great extent, why not apply simply the dense subgraph search algorithms described there as well? The answer partly lies in the fact that in that application, connections have a different role, therefore a different type of dense bipartite subgraphs is searched for.

6.3.2   Handling Noisy and Missing Data

An important improvement of this structure is applying a bipartite graph for network analysis problems, however, a more significant advance is a new way to handle missing or noisy data. Suppose we have complete information about only part of the society. In case of the others we would like to predict the missing information with partially available data, using the patterns of the complete data.

In previous models the ideal situation was to search for complete subgraphs or bisubgraphs (in case of bipartite graphs). From this point of view handling noisy data can be equal to *dense* subgraph mining replacing complete ones as in [10]. In our case noisy data means that instead of dataset D we have a partly damaged dataset $D' = [D_{complete}, D_{damaged}]$, and a $M' = [M_{complete}, M_{damaged}]$ will replace the M matrix. The t_{ij} element of M':

$$t_{ij} = \begin{cases} \mu_{ij} & \text{in case of complete data} \\ 0 & \text{in case of damaged data} \end{cases}$$

6.3.3   Finding Cores of Clusters Using Complete Information Vectors

In numerous dense subgraph mining algorithms an essential part is the mining of the cluster cores. A core is a set of vertices that is sufficiently dense compared to parameter γ. Especially in a large graph it is a hard task to find a fast way of recognizing the cores. In [12] the authors present a theoretical approach to solve this problem for bipartite graphs in $O(|V| + r)$ time. Although it is linear in the number of nodes, it is exponential in the parameter showing how close we are to the complete bipartite subgraphs. Another disadvantage is that it is only applicable if the size of a cluster is also $O(|V| + r)$, and it only finds one cluster core.
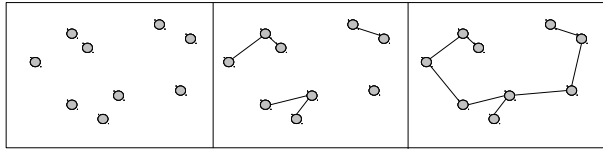
Fig. 6.2 a) Nodes of the graph; b) Inner step of Kruskal-algorithm; c) MST of the graph

Our solution finds the cluster cores in $O(|V_{complete}|^2)$, but it is suitable for any cluster size, and finds all of them parallelly. Using only the complete t_i vectors, we build a standard graph. The nodes represent people we have complete information about, the edgeweights are the Euclidean distance between the corresponding vectors. It takes $O(|V_{complete}|^2)$ to compute the edgeweights. The idea is to use a modified minimum weight spanning tree algorithm. First, the edgeweights should be in increasing order. This can be done in $O(|V_{complete}| \cdot log(|V_{complete}|))$ time. We use the Kruskal-algorithm as a basis. If we run this algorithm until the end, we will get a minimum weight spanning tree. However, if we stop after a few steps, the resulting connected components can be considered as the cluster cores.

An obvious problem is to define the stopping conditions of the algorithm. We will use d_{limit} as a threshold to ensure that in each cluster core the nodes (people) are similar enough. The steps of the algorithm are the following:

Algorithm [ClusterCores]=FINDCORES(M_{comp}, d_{limit})

1. Compute the distance between complete feature vector
2. Increasing order of the distance values: $Dist_{order}$
3. Inicialization: Let $G = (V_{complete}, E)$ be a graph, E=; $i = 1$;
4. $x = Dist_{order}(i)$;
5. if $x \cup E$ contains circle, discard x;
6. else if diameter(Component(x)) $> d_{limit}$, discard x;
7. $E = E \cup x$; $i = i + 1$
8. If there is an edge left, go to step 5.
9. ClusterCores = Connected components

In each step, when the Kruskal-algorithm decides whether an edge should be a part of the spanning tree, we also check the diameter of the evolving component.

6.3.4 Clustering the Nodes Using a Bipartite Graph and Fuzzy Membership Functions

The role of the standard graph was to find the cluster cores. Now we should use the bipartite graph to cluster the remaining nodes.
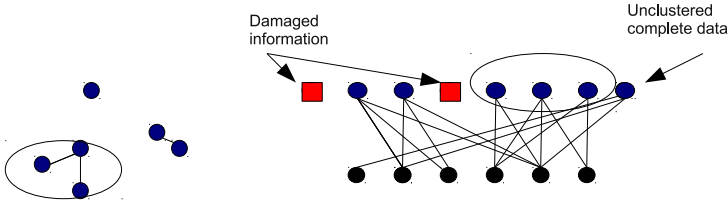
Fig. 6.3 1.fig Cluster cores of the standard graph; 2.fig Cores and damaged data in the bipartite graph

In the bipartite graph, the gained cluster cores form dense bipartite subgraphs, where the density definition is the same as in [12]. The ε parameter can be bounded from below using the d_{limit} parameter from the FINDCORES subalgorithm. Although these cores are dense compared to the average density of the graph, if $\varepsilon \geq \varepsilon_{desired}$ then to get better results the most weekly connected nodes should be removed. The steps of the CLUSTER algorithm are as follows:

Algorithm [MV]=CLUSTER(ClusterCores,M_{comp},M_{dam})

1. For each core in ClusterCores
2. Compute C_{ClCore} cluster core characteristics using CORECHAR()
3. For each vector in M_{comp} compute MV_{comp} membership values
4. Cluster people corresponding to damaged information vectors

---

Using the feature vectors of the nodes in each core, we get the C_{ClCore} characteristics. After this a cluster can be represented as a vector of the weights corresponding to the features.

Algorithm [C_{ClCore}]=CORECHAR(ClusterCores)

1. For each core c_i in ClusterCores
2. $avg(\mu_{jk}) = C_{ClCore}(i,k)$ if $v_j \in c_i$

Clustering usually means deciding to which cluster core the given node belongs to. Instead of this, here we calculate how strong the connection to each core is. In the following we calculate the membership values for each node.

Algorithm [MV_{matrix}]=COMPUTEMV(M_{comp},M_{dam})

1. For each v_i in M_{comp} and $M_d am$
2. For each c_j cluster core if $\mu_{ik} \neq 0$
3. $MV(i,j,k) = tnorm[\mu_{ik},C_{ClCore}(j,k)]$
4. $MV_{matrix}(i,j) = snorm[MV(i,j,k)]_k$

We calculate for every node *i*, how strongly it satisfies the parameters of a cluster c_j for each of the *k* feature. This could be done by an arbitrary fuzzy t-norm, here we have chosen the mininum as a t-norm. As an output we get an MV_{matrix}, where each row corresponds to a node, and each coloumn to a cluster represented by its core. An element of the matrix shows the membership value of a given node to the given cluster. This value was derived from the membership values corresponding to the features by using a normalized sum function as an s-norm.

People we have incomplete or damaged information about have been clustered without difficulties based on the available data.

Let us notice that every node was re-clustered, even the ones used to build up the cluster cores. This way, we can handle overlaps between clusters. Practically it means that taking into consideration each person habits, one might belong to several clusters (communities).

6.4 Test Results

The algorithm was tested on synthetic datasets. A probability model of the relevance of features was used to create the initial data points,to which were added a uniform random noise and features were randomly deleted simulating unknown or missing data. The structure of this dataset is introduced on Fig 4.

The clustering algorithm can be interpreted as a training process. The complete data form the training set, and the damaged ones are the test samples.

The matrix in Fig 5. contains the membership values (modeled by colours) of the samples for each cluster.The order of the subgroups of training and test objects are the same and correspond to the number of the original cluster they belong to. One can see that there is a large overlap between the first two clusters (the cross-membership values are high).

Fig 6. shows histograms of confidence values for tests using different noise and missing data rate. By confidence value we mean the ratio of the membership values corresponding to the best and second best fitting clusters.

From the test results it can be derived that if we increase the noise or the ratio of missing parameters, the two largest membership values for the objects are getting more equal. The ratio of these membership values show how certain the correct cluster can be predict for the objects. Let us notice, that for a part of the object set this ratio is near 1 even if we remove only a small percentage of the data. This can be explained by the large overlapping between part of the clusters.
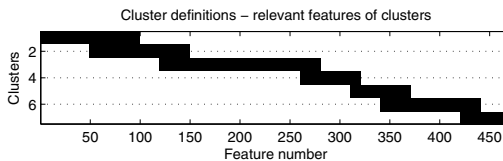
Fig. 6.4 The basis of the training data with overlaps between the relevant features.
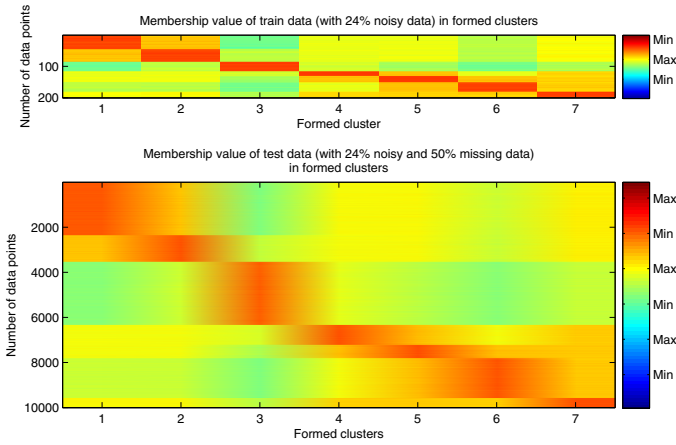
Fig. 6.5 Test results on 10000 objects of 7 overlapping clusters with training and test samples.
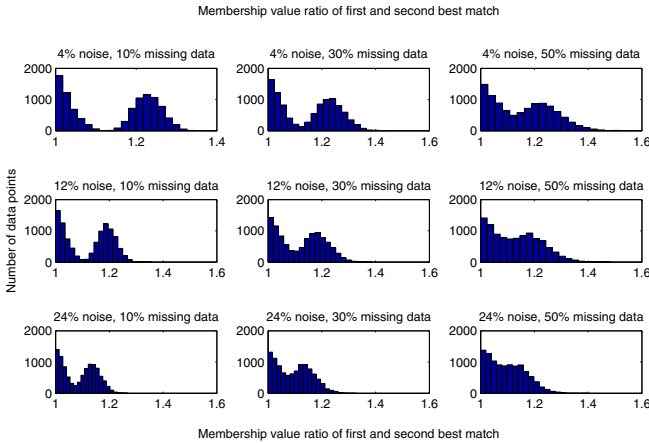
Fig. 6.6 Test results using different noise and missing data rate.

6.5   Conclusions

A new model of community detection was introduced with a highly noise toler-
ance algorithm. The method is suitable for mining groups of people with similar
behaviour, even if part of the information is missing. The algorithm can be used for
predicting the damaged data using the patterns of the cluster cores. If the available
data show that the given person is a member of a community, then there is a high
chance for the missing parameters to match with the ones corresponding to people
belonging to this community.

References

1. Bala, S.E., Chvatal, V., Nesetril, J.: On the maximum weight clique problem. Mathematics of Operations Research 12(3), 522–535 (1987)
2. Basu, A.: Social Network Analysis of Terrorist Organizations in India. In: 2006 Conference of the North American Association for Computational Social and Organizational Science (2006)
3. Ceglowski, M., Coburn, A., Cuadrado, J.: Semantic Search of Unstructured Data using Contextual Network Graphs. National Institute for Technology and Liberal Education (May 16, 2003)
4. Du, N., Wu, B., Pei, X., Wang, B., Xu, L.: Community Detection in Large-Scale Social Networks. In: Zhang, H., Spiliopoulou, M., Mobasher, B., Giles, C.L., McCallum, A., Nasraoui, O., Srivastava, J., Yen, J. (eds.) WebKDD 2007. LNCS, vol. 5439. Springer, Heidelberg (2009)
5. Faloutsos, C., McCurley, K.S., Tomkins, A.: Connection Subgraphs in Social Networks. In: Workshop on Link Analysis, Counterterrorism, and Privacy. SIAM International Conference on Data Mining (2004)
6. Karp, R.M.: Reducibility among combinatorial problems. In: Complexity of Computer Computations, pp. 85–104. Plenum Press, New York (1972)
7. Kleinberg, J.M.: Authoritative sources in a Hyperlinked Environment. Journal of the ACM 46, 604–632 (1999)
8. Klugar, Y., Basri, R., Chang, J.T., Gerstein, M.: Spectral Biclustering of Microarray Data: Coclustering Genes and Conditions. Genome Research 13, 703–716 (2003)
9. Kruskal, J.B.: On the shortest spanning tree of a graph and the traveling salesman problem. In: Proc. Amer. Math. Society, vol. 7, pp. 48–50 (1956)
10. Li, J., Sim, K., Liu, G., Wong, L.: Maximal Quasi-Bicliques with Balanced Noise Tolerance: Concepts and Co-clustering Applications. In: Proceedings of the 2008 SIAM International Conference on Data Mining (2008)
11. Madeira, S.C., Oliveira, A.L.: Biclustering Algorithms for Biological Data Analysis: A survey. IEEE Transactions on Computational Biology and Bioinformatics 1(1) (2004)
12. Mishra, N., Ron, D., Swaminathan, R.: On Finding Large Conjunctive Clusters. In: Schölkopf, B., Warmuth, M.K. (eds.) COLT/Kernel 2003. LNCS (LNAI), vol. 2777, pp. 448–462. Springer, Heidelberg (2003)
13. Schenker, A., Last, M., Bunke, H., Kandel, A.: Comparison of Distance Measures for Graph-Based Clustering of Documents. In: Hancock, E.R., Vento, M. (eds.) GbRPR 2003. LNCS, vol. 2726, pp. 187–263. Springer, Heidelberg (2003)
14. Tan, J.: Inapproximability of Maximum Weighted Edge Biclique and Its Applications. In: Agrawal, M., Du, D.-Z., Duan, Z., Li, A. (eds.) TAMC 2008. LNCS, vol. 4978, pp. 282–293. Springer, Heidelberg (2008)
15. Xu, Y., Olman, V., Uberbacher, E.: Minimum Spanning Trees for Gene Expression Data. Genome Informatics (12), 24–33 (2001)
16. Yang, J., Wang, W., Wang, H., Yu, P.: δ-cluster: Capturing Subspace Correlation in a Large Data Set. In: Proc. 2002 ACM SIGMOD Conf. Management of Data, pp. 394–405 (2002)
17. Zhao, D., Yang, L.: Incremental Isometric Embedding of High-Dimensional Data Using Connected Neighborhood Graphs. IEEE Transactions on Pattern Analysis and Machine Intelligence 31(1) (2009)

Chapter 7
Fuzzy Rule-Based Dynamic Gesture Recognition Employing Camera and Multimedia Projector

Michał Lech and Bożena Kostek

Abstract. In this chapter the system based on camera and multimedia projector enabling a user to control computer applications by dynamic hand gestures is presented. The main objective is to present the gesture recognition methodology which bases on representing hand movement trajectory by motion vectors analyzed using fuzzy rule-based inference. The approach was engineered in the system developed with J2SE and C++ / OpenCV technology. OpenCV was used for image processing and J2SE with jFuzzyLogic package for gesture interpretation. Results of fuzzy rule-based and fixed threshold-based gesture recognition effectiveness are provided. As an example of system usage the so-called Interactive Whiteboard application is presented. Details on the application engineered are provided in the context of fuzzy inference processing.

7.1 Introduction

Recently, an increasing need for development of novel human-computer interfaces (HCI) eliminating use of keyboard and mouse can be observed. Among such solutions gesture recognition interfaces gained much interest. Such interfaces can be classified regarding means used for gesture recognition. For instance, for hand gestures recognition most reliable results can be obtained using motion sensors providing information about all joint angles [1, 2, 3]. Such sensors are often referred to as Datagloves or Cybergloves. Retrieved parameters can be further efficiently interpreted using fuzzy logic [3, 4]. Although the recognition results are very accurate the necessity of wearing gloves all the time when operating can be considered uncomfortable and restrictive even if the device is wireless. Therefore, much attention has been devoted to developing visual-based systems having the ability to recognize gestures effectively. There have been many approaches to this

Michał Lech and Bożena Kostek
Multimedia Systems Dept., Gdansk Univ. Of Technology,
Narutowicza 11/12, 80-233 Gdansk, Poland
e-mail: {mlech,bozenka}@sound.eti.pg.gda.pl

problem which utilize methods of mathematical modelling and artificial intelligence [1, 2]. However, majority of solutions involve a front-facing camera and the user's gestures are recognized from the frames containing static background.

Dynamic gestures are recognized basing on changes of motion parameters such as speed, position and acceleration. Ranges of these parameters can be represented as fuzzy sets. In the paper by Allevard et al. [5] a dynamic gesture recognition process based on fuzzy nominal scales has been presented. The authors of this study utilized however the data glove which provided numerical data transformed later into linguistic variables. Moreover, the study concentrated on mapping finger configuration and orientation, thus this approach cannot be utilized as a straightforward reference to the study presented here. In this case the dynamic gesture recognition uses a resampling technique based on the distance operator introduced by Sawada et al. [6]. Allevard and his colleagues stated that this method works efficiently for the classification of gestures having similar sizes in terms of time. This statement shows that although the algorithm is efficient, it could not be applied to our work. Conversely, hand Dynamic Time Warping seems a promising algorithm providing better time mapping of gestures.

The approach presented herein for dynamic gesture recognition employs fuzzy logic with fuzzy rules describing parameters of local motion vectors, i.e. speed and direction.

In the chapter the human-computer interface based on the multimedia projector-camera configuration is presented. In the proposed solution a camera, coupled with multimedia projector, is placed behind a user. The system has been examined in two work conditions, namely while utilizing the fuzzy-rule based gesture recognition module and without the module applied.

7.1.1 Study Objectives

The main objective of the chapter is to present a concept and the design of employing fuzzy logic and image processing methods used in the proposed projector-camera system configuration for dynamic gestures recognition. Dynamic gestures, defined by changes of controlling the object position within the given time interval, require processing a spatio-temporal image [1, 7]. Therefore, the principal objective of the work was to develop an interface enabling to time effectively control the computer applications using gestures in cases when the content is displayed by the projector and thus, a camera should not be placed front-faced. The assumption was not to use infrared cameras and diodes or other motion sensors. It was also assumed that among possible applications interactive whiteboards and image viewers/browsers, especially helpful in hospitals and photo laboratories, where contactless interaction with a computer is desirable, should be considered. The Interactive Whiteboard application dedicated to the system engineered has already been implemented. In comparison with existing interactive whiteboards based on multimedia projectors the novelty of the interface engineered consists in lack of electronic pens and panels equipped with sensors.

7.2   System Overview

The system was engineered based on both J2SE and C++ / OpenCV environments. All time consuming operations were written in C++ using OpenCV library. The library provided image processing and video stream obtaining and displaying. Also, hand position detection was developed in C++ using OpenCV contour feature with an algorithm of simple chain approximation. Gesture interpretation was engineered in JAVA code and employed the use of jFuzzyLogic package for creating the fuzzy inference system. Due to requirements of jFuzzyLogic package, fuzzy rules were described in separate file using Fuzzy Control Language (FCL). In addition, recalling system events assigned to gestures was contained in JAVA code. JAVA Swing package was used for GUI developing. The C++ / OpenCV code was wrapped with JNI (JAVA Native Interfaces) and compiled to dynamic link library that could be loaded in JAVA code of the gesture interface.

7.2.1   Methodology

10 gestures were defined in the system (see Table. 7.2.1). Gestures made with two hands have been associated with the typical system command-like actions.

Each recognized gesture is interpreted according to the associated system event. For instance, moving left hand down and right hand up is associated with rotating

Table 7.2.1 Gesture set

Gesture name	Description	Default action	Emulated event
Hand steady	Keeping one hand steady	Clicking	Mouse left button click
Left	Moving hand left	Showing previous image / slide or moving the content	key ← pressing
Right	Moving hand right	Showing next image / slide or moving the content	key → pressing
Up	Moving hand up	Showing previous image / slide or moving the content	key ↑ pressing
Down	Moving hand down	Showing next image / slide or moving the content	key ↓ pressing
Hands steady	Keeping both hands steady	-	
Zoom in	Moving hands farther apart	Zooming in	key '+' pressing
Zoom out	Moving hands close together	Zooming out	key '-' pressing
Rotate left	Moving left hand down and right hand up	Rotating left	key 'l' pressing
Rotate right	Moving left hand up and right hand down	Rotating right	key 'r' pressing

left operation when using the engineered system with image viewers or engineered Interactive Whiteboard application. Running operation bases on emulating key 'l' pressing in this case.

7.2.2 Image Processing

The solution bases on subtracting the image extracted from the video stream from the image displayed by the multimedia projector and recognizing gestures in further processed output. In the first step, the effective area of the image captured from the camera is determined. This area contains only the image displayed by the projector and is determined by the user who points out positions of the image corners in the frame. Based on these positions, the projected image is scaled to ensure identical dimensions with the camera frame. In the next step, the perspective correction is performed. To reduce impact of light conditions and distortions introduced by the camera lens, especially vignetting effect, color calibration is performed. During this process five colors (red, green, blue, white and black) of backgrounds are displayed. Each camera frame respectively for each background color is subtracted from particular displayed image. Based on the results, tables of discrete constant values used in the later image processing are created. Gesture recognition iteration begins with subtracting processed camera frames from projected images. Subtraction is done in RGB color space. To each pixel of the output image an appropriate value retrieved from the color calibration tables is added. This value is chosen based on displayed screen's particular pixel component intensity. The result is binary thresholded and median filtered with separable square mask. In the obtained image, thanks to color calibration and filtering, the biggest areas of adjacent pixels represent hands. Their positions are detected using active contours method with a given size threshold to prevent the recognition of accidental aggregation of pixels as hand when in fact no hand is present in an image.

7.2.3 Hand Motion Modeling

The first approach to gesture interpretation, implemented in the system, assumed that when performing a gesture of moving hand in one particular direction there are no local direction changes. Thus, movement was represented as global motion vector over particular number of frames. Also, it was noticed that the number of frames on which such a vector was created should be relatively small because movement can be of a high speed and thus not properly represented within many frames. Taking these assumptions into considerations, global motion vectors each over three adjacent frames were created. However, it was noticed during tests that wide hand movements to left or right side are often performed semi-circular, similarly to a waving gesture. Thus, representing a gesture as a singular change of speed and direction over particular time interval often led to interpreting it as moving hand up - in the beginning phase of the movement - or as moving hand
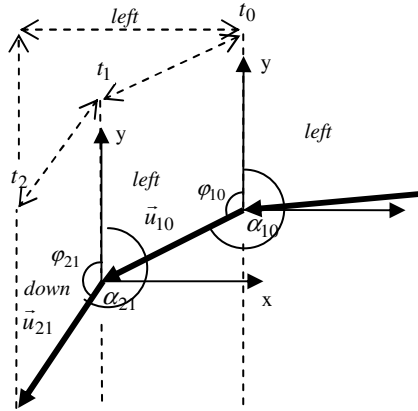
Fig. 7.2.1 Motion vectors created for semi-circular hand movement in left direction

down - in the ending phase. Therefore, the movement trajectory in the second approach has been modeled by motion vectors created for points in time moments t_1 and t_2, in relation to moments t_0 and t_1, respectively, as presented in Fig. 7.2.1 and gestures were analyzed considering a possibility of local change of direction.

Time intervals $t_1 - t_0$ and $t_2 - t_1$, expressed in number of frames retrieved from a camera, depend on camera frame rate. For 15 FPS, obtained during the tests, the value by default equaled 3 frames. The engineered interface provided the possibility of changing this parameter. Six frames obtained within time interval $t_2 - t_0$ constituted in average 378 ms which was the optimum time range of registering representative movement change for most people that took part in the system testing. Such a solution effectively prevented from recognizing gesture as moving hand up or down when waving occurred.

Two parameters of the motion vectors, i.e. speed and direction, were used as a basis for gesture interpretation mechanism. Speed for motion vector within the time interval $t_i - t_{i-1}$, denoted as v_{ij}, where $j = i - 1$, was calculated according to Eq. (7.2.1). Direction for particular motion vector $\vec{u}_{ij} = \left[u_x^{ij}, u_y^{ij}\right]$ was denoted as an angle α_{ij} in relation to angle φ_{ij} between $\vec{u}_{ij}$ with origin at $[0,0]$ and versor of y axis, according to Eqs. (7.2.2) and (7.2.3).

$$v_{ij} = \frac{\sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2}}{t_i - t_{i-1}} \left[\frac{px}{s \cdot 10^{-1}}\right] \tag{7.2.1}$$

where: x_i and x_{i-1} are x positions of hand in time t_i and t_{i-1}, respectively, and y_i, y_{i-1} are y positions of hand in time t_i and t_{i-1}, respectively;

$$\varphi_{ij} = \frac{180° \cdot a_{ij} \cos \dfrac{u_y^{ij}}{|\vec{u}_{ij}|}}{\pi} \ [°] \tag{7.2.2}$$

$$\alpha_{ij} = \begin{cases} \varphi_{ij} & ,u_x^{ij} \geq 0 \\ 360° - \varphi_{ij} & ,u_x^{ij} < 0 \end{cases} \tag{7.2.3}$$

7.2.4 Fuzzy Rule-Based Gesture Recognition

The approach used in the system before applying fuzzy rules was based on fixed thresholds of speed. The angle obtained according to Eqs. (7.2.2-7.2.3) was classified as one of four directions. Classifying movement to a particular gesture class was positive if a speed was above or below the threshold, depending on a gesture type, and hand moved in particular direction. However, using fixed speed thresholds didn't allow for gesture recognition when the speed was not high enough, although hand moved in the proper direction. Further decreasing the threshold was not an appropriate solution since the effectiveness of recognizing 'hand(s) steady' gesture would decrease severely. Using fuzzy sets both for speed and direction and defining gesture classes based on fuzzy rules solved this problem.

Fuzzy rules were created basing on speed and direction of motion vector over time interval $t_2 - t_1$ and $t_1 - t_0$ separately for left and right hand. Eight linguistic variables were proposed, i.e.: speed of left and right hand in time interval $t_2 - t_1$, speed of left and right hand in time interval $t_1 - t_0$, direction of left and right hand in time interval $t_2 - t_1$, direction of left and right hand in time interval $t_1 - t_0$,
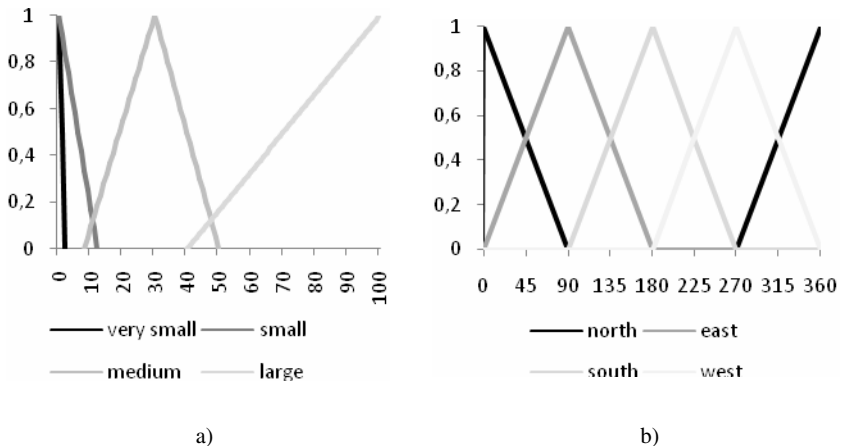
Fig. 7.2.2 Fuzzy sets for linguistic variables speed (a) and direction (b)

denoted as $\upsilon_{21}^L, \upsilon_{21}^R, \upsilon_{10}^L, \upsilon_{10}^R, d_{21}^L, d_{21}^R, d_{10}^L, d_{10}^R$, respectively. Four linguistic terms were used for speed, i.e.: *very small, small, medium and high*, represented by triangular functions as shown in Fig. 7.2.2a. Fuzzy sets were identical for all four variables. For directions the terms used were *north, east, south* and *west* and fuzzy sets were also formed using triangular functions as shown in Fig. 7.2.2b.

The zero-order Takagi-Sugeno inference model which bases on singletons was used to express discrete rule outputs representing gesture classes. The output of the system was the maximum of all rule outputs. When this value was lower than 0.5 a movement was labeled as *no gesture*. This enabled to efficiently solve the problem of classifying meaningless transitions between each two various gestures to one of the gesture classes. The total number of rules equaled 30. Two examples of rules expressed in FCL code are given below.

```
// beginning phase of hand movement in the left direction (for semi-circular motion) for left
hand
RULE 1 : IF directionLt0 IS north AND directionLt1 IS west AND velocityLt0 IS NOT small
AND velocityLt1 IS NOT small AND velocityRt0 IS vsmall AND velocityRt1 IS vsmall THEN
gesture IS g1;
// rotate left
RULE 29 : IF directionLt0 IS south AND directionLt1 IS south AND directionRt0 IS north AND
directionRt1 IS north AND (velocityLt1 IS NOT vsmall AND velocityLt0 IS NOT vsmall) AND
(velocityRt1 IS NOT vsmall AND velocityRt0 IS NOT vsmall) THEN gesture IS g7;
```

The first rule describes the beginning phase of semi-circular left hand movement from right to the left side. Therefore, d_{10}^R is north and d_{21}^R is west. Since the gesture involves left hand only, the speed of the right hand should be very small. If the right hand is not present in an image, 0.0 values are given as an input to the fuzzy inference system for variables υ_{21}^R and υ_{10}^R. The second rule represents the gesture associated with rotating the displayed object. During the gesture performing, the left hand moves down and the right hand moves up. No local change of direction is allowed. For this reason, both d_{21}^L and d_{10}^L are south and d_{21}^R, d_{10}^R are north. While making gestures involving both hands, speed of each hand movement can be lower than when performing a single hand gesture. Therefore, contrary to the first rule the second one allows for small speed.

7.3 Results

20 persons took part in tests. Each person was asked to repeat each gesture 18 times. Among these 18 repetitions 10 middle gesture representations were chosen. Since the system analyzes motion vectors for time intervals $t_2 - t_1$ and $t_1 - t_0$ in relation to each obtained camera frame, among each gesture representation there

were many assignments to particular gesture class. All these classifications were taken into consideration while analyzing the recognition effectiveness. Among the samples obtained four beginning and four ending gesture representations were rejected due to time needed to familiarize with making each gesture and weariness leading to human mistakes. No special constraints like moving hand absolutely straight in particular direction or forming particular shape with hand were put on persons who tested the system. The results of a comparison between fuzzy rule-based recognition and recognition based on fixed thresholds with the analysis of global motion vector change are presented in Tables 7.3.1 and 7.3.2. For gestures composing of one hand movements the presented results are average of recognition effectiveness for left and right hand. The differences in effectiveness for each hand were below 1.0% and could be disregarded. If a particular column presenting misclassifications for all analyzed gesture classes is not present in the table it means that all its values equaled 0.0. For example, one hand gesture was never interpreted as a gesture involving both hands and for this reason there are no classes labeled *zoom in*, *zoom out*, *rotate left*, *rotate right* and *hands steady* in Table 7.3.1.

Table 7.3.1 Gesture recognition effectiveness for the system employing fuzzy inference and without a module of fuzzy inference, for one hand gestures [%]

	With fuzzy logic						No fuzzy logic					
	Left	Right	Up	Down	Hand steady	No gesture	Left	Right	Up	Down	Hand steady	No gesture
Left	95.0	0.0	2.3	2.6	0.0	0.1	89.5	0.0	4.9	5.6	0.0	0.0
Right	0.0	94.2	2.9	2.7	0.0	0.2	0.0	89.6	5.8	4.6	0.0	0.0
Up	0.9	0.5	98.6	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0
Down	2.2	0.9	0.0	96.9	0.0	0.0	0.0	0.0	0.0	99.8	0.0	0.2
Hand steady	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	73.3	16.7

Table 7.3.2 Gesture recognition effectiveness for the system employing fuzzy inference and without a module of fuzzy inference, for gestures involving both hands [%]

G1 – up, G2 – down, G3 – zoom in, G4 – zoom out, G5 – rotate left, G6 – rotate right, G7 – no gesture

	With fuzzy logic						No fuzzy logic						
	G1	G2	G3	G4	G5	G6	G1	G2	G3	G4	G5	G6	G7
Zoom in	0.2	0.1	99.7	0.0	0.0	0.0	0.0	0.0	98.6	0.0	0.0	0.0	1.4
Zoom out	0.6	0.2	0.0	99.2	0.0	0.0	0.0	0.2	0.0	98.9	0.0	0.0	0.9
Rotate left	0.5	0.0	0.0	0.4	99.2	0.0	0.4	0.0	0.0	0.0	98.2	0.0	1.4
Rotate right	0.2	0.3	0.0	0.4	0.0	99.1	0.0	0.2	0.0	0.0	0.0	99.0	0.8

A comparison of the efficacy test results for both solutions shows an improvement in recognition of hand movements that were not ideally straight in particular direction, such as the waving gesture mentioned before. The effectiveness was also increased by using fuzzy speed sets which were especially relevant in recognition of gestures often composed of slow movements, like zooming or rotating. With the solution based on fixed speed thresholds and four directions only, in the case when the speed was below threshold gestures were not classified. Also, when using fuzzy inference there was an improvement in recognizing gesture labeled *Hand steady*. Using fixed thresholds led to no recognition when hand trembled. Conversely, using fuzzy rules caused worse recognition of gestures labeled *Up* and *Down*. These gestures were in some cases recognized as movement from left to right side or inversely. Such a situation occurred when the hand movement was not performed vertically and resulted from this the direction of motion vector at a given moment was *west* or *east*. A membership degree of rule output was in this case slightly above 0.5 and thus the misclassification happened.

7.4 Conclusions

The results obtained have shown that describing motion vectors by fuzzy rules can provide an efficient solution for gesture interpretation, especially considering the possibility of local change of hand movement direction. Applicability of fuzzy inference could be also noticed in reliable differentiation between similar gestures like moving hand very slowly in particular direction and keeping hand steady. Moreover, setting an appropriate membership degree threshold can reassure that meaningless movements, which in particular may occur during transitions between gestures of different classes, are not classified as gestures.

For the purpose of an effective work with the Whiteboard, an additional set of fuzzy rules was created. A gesture associated with scrolling whiteboard contents, hand movement of a speed labeled *high* in one particular direction was considered. High relevance of using fuzzy inference in this case can be noticed. When fuzzy sets are not utilized in the processing this might create a risk of interpreting hand movements such as gesture associated with scrolling the contents during for example option setting. Conversely classifying hand movement gesture based on the rule membership degree can efficiently deal with this issue.

The presented Whiteboard in its current form enables to write and handle the content from distance. Such solution is useful due to the fact that the whiteboard content is not obscured by a user. However, its intuitivity while writing or drawing shapes may be still improved. For this reason a concept of applying Support Vector Machines for differentiating hands from the rest of a body and recognizing particular palm shapes in captured frames has been thought up. The next step would consist in adding SVM to provide possibility of operating in very close proximity.

References

1. Mitra, S., Acharya, T.: Gesture Recognition: A Survey. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 37(3), 311–324 (2007)
2. Vlaardingen, M.: Hand Models and Systems for Hand Detection, Shape Recognition and Pose Estimation in Video (2006),
   `http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.98.`
   `2041` (accessed June 14, 2010)
3. Kim, J.-H., Kim, D.-G., et al.: Hand Gesture Recognition System Using Fuzzy Algorithm and RDBMS for Post PC. In: Wang, L., Jin, Y. (eds.) FSKD 2005. LNCS (LNAI), vol. 3614, pp. 170–175. Springer, Heidelberg (2005)
4. Callejas Bedregal, B.R., Dimuro, G.P., Rocha Costa, A.C.: Interval Fuzzy Rule-Based Hand Gesture Recognition. In: 12th GAMM - IMACS International Symposium on Scientific Computing, Computer Arithmetic and Validated Numerics, SCAN 2006, p. 12 (2006)
5. Allevard, T., Benoit, E., Foulloy, L.: Dynamic gesture recognition using signal processing based on fuzzy nominal scales. Measurement 38, 303–312 (2005)
6. Sawada, H., Ukegawa, T., Benoit, E.: Robust gesture recognition by possibilistic approach based on data resampling. In: Fuzzy Systems and Innovational Computing (FIC 2004), pp. 168–173 (2004)
7. Yeasin, M., Chaudhuri, S.: Visual understanding of dynamic hand gestures. Pattern Recognition 33(11), 1805–1817 (2000)

Chapter 8
Video Structure Analysis and Content-Based Indexing in the Automatic Video Indexer AVI

Kazimierz Choroś

Abstract. Similarly to text, video is hierarchically structured. The analogies of text and video structures are discussed. Then the juxtaposition is presented of two indexing processes, i.e. of text and video indexing based on the content analysis of their structure units. Several frameworks of automatic detection and categorisation of video shots and scenes reporting the sport events in a given discipline in TV sports news have already been proposed. It has been observed that many sport videos such as archery, diving, soccer, and tennis have repetitive structure patterns. In the tests performed using the Automatic Video Indexer AVI shots and then scenes have been detected in tested TV news videos. Experimental results show good performance of the scheme of video scene detection of a given sport discipline in TV sports news. The Automatic Video Indexer is a research project investigating tools and techniques of automatic video indexing for retrieval systems.

8.1 Introduction

New technologies applied in visual retrieval systems allow the storage of a very huge amount of digital video data. Video data have become publicly and relatively easy available. However, the easy and efficient access to desired videos will be possible when appropriate indexing and retrieval methods of these video data are applied. The methods well-known in textual retrieval systems are not satisfactory solutions in this case because users want to query not only technical data of videos which are indexed in traditional ways, but also the content of video clips. Manual indexing is unfeasible for large video collections. Unfortunately, the content-based automatic indexing and retrieval of video data are still processes difficult to be effectively performed. The content is very subjective to be characterized completely. It is related to main objects, second plan, background, domain, context, etc. In consequence content-based access to large video archives is still largely

Kazimierz Choroś
Institute of Informatics, Wrocław University of Technology,
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
e-mail: `kazimierz.choros@pwr.wroc.pl`

unsolved. Content-based indexing of videos has become a research topic of increasing importance, difficult but at the same time fascinating, theoretical, scientific problem as well as practical task. It is necessary to develop methods to organize, index, browse, and retrieve videos in archives using high-level semantic features. At the same time the same procedures could be used to detect and to remove commercial in TV news program. Therefore, we are looking for effective tools to identify the video segments like shots or scenes with a specific content, for example news on weather, sports, science, finances, technology, world travel, national economy, or entertainment news.

The procedure of the detection of shots called temporal segmentation of video is basic processing. It is a crucial process not only in the video indexing process but also in automatic generation of highlights, summarization and content annotation as well as in the process of automatic categorization of scenes detected. Many experiments have been conducted with TV news, mainly TV sports news.

The chapter is organized as follows. The next section describes the main related works in the area of automatic scene detection and selection in TV news videos. Moreover, some recent related research works are cited. The section 8.3 discusses analogies of text and video structures. The juxtaposition is also presented of two indexing processes, i.e. of text and video indexing based on the content analysis of their structure units. The section 8.4 presents the Automatic Video Indexer AVI which is a research project investigating tools and techniques of automatic video indexing for retrieval systems. The temporal segmentation process leading to the partition of a given video into a set of meaningful and individually manageable segments will be discussed in the next section. This process is already well managed and the adequate procedures have been implemented, tested and verified in the AVI system. The section 8.6 presents a automatic detection of the sequence of shots making a scene. The ASD module, i.e. the Automatic Shot Detection module of the Automatic Video Indexer performs this task with adequate efficiency. The experimental results for the classification of scenes in the tested TV sports news videos are reported. The tests performed using the ASA module – Automatic Shot Analyzer have shown that it is possible to detect the soccer news in the sequence of TV sport news. The procedures are now being developed mainly by definition of repetitive patterns which we observe in practice in news editing standards. Many sport videos such as for example archery, diving, soccer, or tennis have repetitive structure patterns. The final conclusions and the future research work areas are discussed in the last section of this chapter.

8.2 Related Works

Last years many investigations in automatic recognition of a content of a video clip [1] have been carried out, many of proposed methods have been tested on TV sport videos. An automatic summarization of sport videos has become a popular application because of its popularity to users and its simplicity due to standard editing style. So, because of a huge commercial appeal sports videos became a dominant application area for video automatic indexing and retrieval.

Many promising frameworks have been proposed for soccer videos. Automatic goals detection method in soccer video using audio/visual keywords has been described in [2]. Two levels of processing have been applied. The first one extracts low-level features of a video such as motion, colour, texture, pitch, etc. to detect video segments boundaries and label segments using audio and visual keywords. Then on the second level two Hidden Markov Models have been used to model the exciting break portions with and without goal event, respectively. The proposed framework has been applied and tested to the detection of goal event in six half matches of soccer videos (270 minutes, 14 goals) from FIFA 2002 and UEFA 2002 and achieve 90% precision and 100% recall, respectively. Whereas, in another paper [3] a method of replay detection also in soccer video has been presented.

Another approach [4] and another kind of analyses has been implemented in a system that performs automatic annotation of soccer videos. It has resulted in detecting principal highlights, recognizing identity of players based on face detection, and detecting contextual information such as jersey's numbers and superimposed text captions.

New methods have been proposed and experiments have been carried out not only with soccer videos but also for example with baseball videos [5], with tennis videos [6], with badminton [7], as well as with other sports.

Furthermore, many experiments have been also performed on sports news classification and many approaches and schemes have been developed. In the scene classification algorithm described in [8] used the DCT (Discrete Cosine Transformation) components extracted from the whole image as the classification features. An other technique [9] is relied upon the concept of "cues" which attach semantic meaning to low-level features computed on the video. A cue detector was defined as a supervised specifically trained classifier. Examples of cue detectors included: grass, swimming pool lanes, ocean, or audio elements like referee whistle, crowd cheer. In [10] it has been shown that the weighting of individual classifier according to their estimated performance gives better results in automatic classifications. A unified framework for semantic shot classification in sports videos has been defined in [11]. The proposed scheme makes use of domain knowledge of specific sport to perform a top-down video shot classification, including identification of video shots classes for each sport. The method has been tested over 3 types of sports videos: tennis, basketball, and soccer. The results ranging from 80~95% have been achieved. In the next research work [12] it has been demonstrated that combining the tiny images and tiny videos datasets improves categorization precision in a wider range of categories.

The main two processes of video indexing and summarization are shot detection (temporal segmentation) and scene detection. A video scene is defined as a sequence of semantically correlated shots near in time or location [13, 14]. One of the promising approach in scene detection is based on the fact that broadcast video clips are usually edited in a standard scheme. Many sports videos such as archery, diving, and tennis have repetitive structure patterns [15].

8.3 Text Structure vs. Video Structure

Text is autodescriptive. The text in any natural language is composed of words, sentences, paragraphs, and chapters. The process of indexing is usually limited to the identification of words or expressions in the text. These words or expressions are used as index terms in retrieval systems. Let's notice that the text has also structural features. These are for example: a language of the text, its length measured in characters, but also in words, lines, or pages.

A digital video clip is also hierarchically structured into a strict hierarchy [16, 17]. It is composed of different structural units such as: acts, episodes (sequences), scenes, camera shots and finally, single frames. The most general unit is an act. So, a film is composed of one or more acts. Then, acts include one or more sequences, sequences comprise one or more scenes, and finally, scenes are built out of camera shots. A shot is a basic unit. A shot is usually defined as a continuous video acquisition with the same camera, so, it is a sequence of interrelated consecutive frames recorded contiguously and representing a continuous action in time or space. The length of shots affects a film. Shots with a longer duration make a scene seem more slower paced whereas shots with a shorter duration can make a scene seem dynamic and faster paced. The average shot length of a film is generally several seconds or more. Depending on the style of video editing shots in a scene are content related but can be temporally separated and/or even spatially disconnected.

The indexing process – it is evident that much more complex – can be undertaken similarly to the text indexing strategy, form the simplest textual units to the most advanced parts. Table 8.1 compares the basic textual units with video structural elements.

Table 8.1 Analogies in the structures of a text and of a video

Text	Video
character	frame
word	shot
sentence	scene
paragraph	episode
chapter	act
book	movie

We see the great analogy between the structural elements of a text and the structural elements of a video clip. The video is a visual representation of the book. It is not a surprising conclusion if we remember that a screenplay of a movie is frequently based on a novel.

Also in indexing processes of text and video we can observe many analogies. The juxtaposition of two indexing processes based on the content analysis of their structure units is presented in Table 8.2.

Table 8.2 Comparison of processes of text indexing and of video indexing

Text Indexing	Video Indexing

character decoding:
    recognition of individual characters, elimination of punctuation symbols
	frame decoding:
	    frame analysis, calculation of frames characteristics (histograms etc.)

word selection in a text:
    morphological analysis, elimination of word using stop-lists, word normalisation, identification of descriptors from a thesaurus, identification of relation between words, calculation of word frequency
	temporal segmentation:
	    shot detection, calculation of shot length, elimination of shots too short, content-based categorisation of shots, detection of objects, faces, lines, words etc. in a shot

morpho-syntactic analysis of sentences:
    identification of multi-word (compound) index terms, syntactical patterns, multi-word expressions, noun phrases
	scene detection:
	    shot clustering, pattern analysis of scenes

semantic analysis of a paragraph:
    semantic and contextual analysis of sentences
	content analysis of episodes:
	    scene clustering

content analysis of chapters and of the whole text (book)	content analysis of acts and of the whole movie

It should be noticed that a book is traditionally a sequential text, in contrast to hypertext which overcomes the traditional linear constraints of written text. Traditional video has also a sequential nature, although semantically the scenes of the same episode in a given video are not necessarily placed one after another. But such a situation is also natural for a written text in a novel. It usually happens that two or several episodes are presented in alternation.

8.4 AVI – Automatic Video Indexer

The AVI – Automatic Video Indexer [18, 19] is a research project investigating tools and techniques of automatic video indexing for retrieval systems. The main goal of the project is to develop efficient techniques of content-based video

retrieval. The process of automatic content-based analysis and video indexing is composed of several stages (Fig. 8.1). Two main modules have been already implemented in the Automatic Video Indexer: the Automatic Shot Detector ASD is responsible for temporal segmentation and shot categorisation, the Automatic Scene Analyser ASA is responsible for shot clustering, scene detection, and content analysis of scenes. The next modules will analyse the scene to classify TV sports news and to detect important events and people, and then to extract interesting highlights, which facilitate browsing and retrieval of sports video.

Generally, the first step of video indexing is a temporal segmentation leading to the segmentation of a movie into small parts called video shots. Automatic video file segmentation in the AVI includes five important steps. The first step is a temporal segmentation process. It leads to the shot boundary detection. The next step is the key frame extraction, the best for depicting the content of corresponding shot or scene. And the third step the content of the shots detected in the video during the temporal segmentation is analysed. In the fourth step the shots with studio are treated as the boundaries of scenes. It leads to shot clustering, and in consequence scene segmentation. Finally after shot content identification and shot clustering processes, the content of a scene is recognized using different approaches. This step is still being developed.

8.5   Temporal Segmentation Process

A video clip is structured into a strict hierarchy and is composed of different structural units. These structural units are: acts, episodes (sequences), scenes, camera shots, and finally single, still frames.

A shot is usually treated as basic video unit. A shot is defined as a continuous video acquisition with the same camera, so, it is as a sequence of interrelated consecutive frames recorded contiguously and representing a continuous action in time or space [14]. Shots in a scene are content related but can be temporally separated and/or even spatially disconnected depending on the style of video editing.

A shot change occurs when a video acquisition is done with another camera. Cuts and dissolves are the most frequently used transitions to perform a change between two shots. Cut takes place when the last frame of the first video sequence is directly followed by the first frame of the second video sequence. Whereas, a dissolve is a transition where all the images inserted between the two video sequences contain pixels whose values are computed as linear combination of the final frame of the first video sequence and the initial frame of the second video sequence.

Cross dissolve describes the cross fading of two scenes. Over a certain period of time (usually several frames or several seconds) the images of two scenes
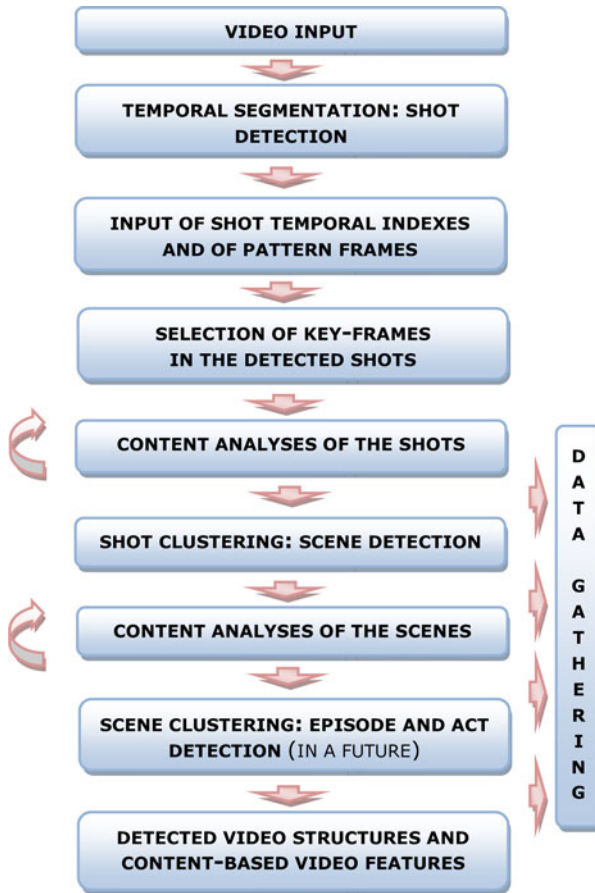
Fig. 8.1 Indexing scheme in the Automatic Video Indexer AVI

overlay, and then the current scene dissolves into a new one. Examples of cuts and dissolves are presented in [17]. Fades are special cases of dissolve effects, where a black frame most frequently replaces the last frame of the first shot (fade in) or the first frame of the second shot (fade out). Whereas, a wipe effect is obtained by progressively replacing the old image by the new one, using a spatial basis.

The Automatic Shot Detector have achieved sufficient level of quality and reliability (Tab. 8.3) to be applied in practice for further research investigations. The detailed description of the segmentation methods applied in the Automatic Video Indexer as well as the detailed presentation of testing results carried out for several categories of videos have been reported in [18]. The aggregated results of temporal segmentation effectiveness are presented in Table 8.3.

Table 8.3 The best results of recall R and the best results of precision P of temporal segmentation methods received (but not necessarily simultaneously) for several categories of video when testing the effectiveness of the Automatic Video Indexer

Results with the best recall and the best precision	Pixel pair differences		Likelihood ratio method		Histogram differences		Twin threshold comparison	
	R	P	R	P	R	P	R	P
TV Talk-Show	1.00	1.00	1.00	0.98	1.00	1.00	1.00	0.89
Documentary Video	0.87	1.00	0.98	1.00	0.89	1.00	1.00	0.86
Animal Video	0.88	1.00	1.00	0.89	0.96	1.00	1.00	0.76
Adventure Video	1.00	0.80	1.00	0.76	0.92	1.00	0.97	0.75
POP Music Video	0.95	1.00	0.85	0.90	0.65	1.00	0.88	0.85

8.6 Automatic Scene Detection

It was observed that TV sports news program has a specific structure. The analyses of TV sports newscast broadcasted in the first national Polish TV channel show that it has its individual standard editing structure. It is composed of several highlights introduced and commented by anchorperson, and often accompanying by numerical results presented in tables.

8.6.1 Shot Clustering

Two different classes of shots can be identified: studio shots and news report shots. The special shots with commentaries and tables will be called studios. They usually point out the boundaries of individual sport scenes. Shots that belong to the same scene are visually similar because they report a given sport event and they are also located closely along the time axis according to the definitions of video structural units defined in the section 8.3. A single scene is formed by all successive news report shots until the next studio shot.

The analysis of the length of shots (Table 8.4) has shown that the average length of studio shots is significantly greater than that of the news report shots. So, the length of a shot was a criterion for the decision which type a given shot is.

In practice studio shots last twice longer than any other and another interesting observation was that the soccer shots – the most frequently used in experiments - are generally the shortest ones. It has been assumed that all shots between two studio or table shots belong the same semantic video scene. A scene was detected as a series of shots separated by these two kinds of relatively easily detected shots: studio and table.

Table 8.4 Average length of a shot for a given sport discipline in the training set of TV sports news programs [19]

Category/Discipline	Average shot length [in frames]
Soccer	91
Golf	125
Speed way	145
Cross-country skiing	189
Basketball	239
Commentary/Studio	413
Table	438

8.6.2 TV Sports News Categorization

The basic test collection used in the experiments described in this chapter is a set of 20 TV sports newscasts. These videos were broadcasted on different days in the first national Polish TV channel, digitised to DV format (720 x 576 pixels, full colour), and then introduced to the database for training and evaluation. Seven videos have been used as a training set, the next 13 are used in an automatic content scene analysis. The experiment has concerned the scene categorization, i.e. the recognition of a sport category (discipline). The training set has also been used to choose the most representative still frames for a given sport discipline [19].

The content analysis of scenes is carried out in two steps. The first step is based on histograms which are commonly used to classify images in content-based image retrieval systems. The distance of histograms is measured of pattern frames set and the key-frames, which are the most representatives frames indicated by the strategy applied in the Automatic Video Indexer.

The standard distance D of quantized colour histograms H of two images I_A and I_B has been applied in the AVI experiment and measured as follows:

$$D_H(I_A, I_B) = \sum_{k=1}^{n} \left| H^k(I_A) - H^k(I_B) \right|,$$
(8.1)

where n is the number of colours in the colour space.

In the second step of content analysis a colour coherence vector [19] has been measured to improve the results obtained using the histogram distance measure.

A colour coherence vector measure improves global histogram matching and takes into account spatial information in colour images. Colour coherence vector indicates if pixels belong to a large region of similar colour. Each pixel is classified as either coherent or incoherent to a given colour. If every pixel in the set has at least one pixel of the same colour among its eight closest neighbours, such a set is called a maximal set. The size of a maximal set must exceed a given threshold, then a whole region is classified as coherent. The total number α of coherent and the total number β of incoherent pixels are computed for each colour of n colours in a discretized set of colours. The colour coherence vector V_C of an image is defined as:

$$V_C = [(\alpha_1, \beta_1), (\alpha_2, \beta_2), ..., (\alpha_n, \beta_n)].$$ (8.2)

Two colour coherence vectors of two images *A* and *B* can be compared according to the following distance formula:

$$D_C(I_A, I_B) = \sum_{k=1}^{n} \left(\left| \alpha_{A_k} - \alpha_{B_k} \right| + \left| \beta_{A_k} - \beta_{B_k} \right| \right).$$ (8.3)

The examples of results of the procedure applied in the Automatic Video Indexer are presented in Table 8.5. Standard measure of recall has varied from 0.51 to 1 whereas precision have the values between 0.60 and 0.82.

Table 8.5 The extreme recall values and precision for two videos of the automatic content-based recognition of soccer scenes [19]

Discipline	Recall	Precision
Video 1		
Commentary/studio	1	0.38
Soccer	1	0.60
Video 2		
Commentary/studio	1	0.26
Soccer	0.51	0.82

8.6.3 Scene Repetitive Patterns

The automatic detection and categorisation of video shots and scenes reporting the sport events in a given discipline in TV sports news is one of the area of recent developments. The analyses of newscasts concerning given sport disciplines have shown that many sport videos such as archery, diving, soccer, and tennis have repetitive structure patterns.

For example, a story unit in diving video in the majority cases can be defined as a sequence of shots with predictable contents. The diving scene is composed of four following shots: the player standing on the dive platform, the actions of taking off, diving, and finally entering water [15].

In our experiments tennis TV news have been analysed and the repetitive pattern has also been observed. A strong majority of tennis highlights in TV sport news are of a standard structure of six or seven shots: first player, second player, tennis court, serve, return ball or balls, and zoom presenting two players shaking hands over a tennis net.

The recognition of scene editing patterns leads to the optimisation of other processing of digital videos such as face detection, scene classification, goal or serve detection, etc.

8.7 Final Conclusion and Further Studies

A framework of the automatic content based analysis of TV sport news broadcasted from the first Polish national TV channel has been proposed. The Automatic Video Indexing has been designed and implemented to investigate tools and techniques of automatic video indexing for retrieval systems. The main tasks of the AVI Indexer are: temporal segmentation of videos, key-frames selection, shot analysis and clustering, soccer shot retrieving, repetitive editing pattern identification.

In further research the representative frames for other sport disciplines will be selected. Then, the specific, repetitive structures of scenes for other sport discipline will be determined. Also new computing techniques will be developed for the Automatic Video Indexer. Its functionality will be extended by introducing an automatic extraction of video features and objects like faces, lines, texts, etc., as well as its application will be extended to other kinds of TV shows.

References

1. Lew, M.S., Sebe, N., Djeraba, C., Jain, R.: Content-based multimedia information retrieval: State of the art and challenges. ACM Transactions on Multimedia Computing, Communications, and Applications, 1–19 (2006)
2. Kang, Y.-L., Lim, J.-H., Kankanhalli, M.S., Xu, C., Tian, Q.: Goal detection in soccer video using audio/visual. In: Proceedings of the ICIP, pp. 1629–1632 (2004)
3. Yang, Y., Lin, S., Zhang, Y., Tang, S.: A statistical framework for replay detection in soccer video. In: IEEE International Symposium on Circuits and Systems ISCAS 2008, pp. 3538–3541 (2008)
4. Bertini, M., Del Bimbo, A., Nunziati, W.: Automatic annotation of sport video content. In: Lazo, M., Sanfeliu, A. (eds.) CIARP 2005. LNCS, vol. 3773, pp. 1066–1078. Springer, Heidelberg (2005)
5. Lien, C.-C., Chiang, C.-L., Lee, C.-H.: Scene-based event detection for baseball videos. Journal of Visual Communication and Image Representation, 1–14 (2007)
6. Delakis, M., Gravier, G., Gros, P.: Audiovisual integration with segment models for tennis video parsing. Computer Vision and Image Understanding, 142–154 (2008)
7. Yang, Y., Lin, S., Zhang, Y., Tang, S.: Statistical Framework for Shot Segmentation and Classification in Sports Video. In: Yagi, Y., Kang, S.B., Kweon, I.S., Zha, H. (eds.) ACCV 2007, Part II. LNCS, vol. 4844, pp. 106–115. Springer, Heidelberg (2007)
8. Ariki, Y., Sugiyama, Y.: Classification of TV sports news by DCT features using multiple subspace method. In: Proceedings of Fourteenth International Conference on Pattern Recognition, vol. 2, pp. 1488–1491 (1998)
9. Messer, K., Christmas, W., Kittler, J.: Automatic sports classification. In: Proceedings of the 16th International Conference on Pattern Recognition, pp. 1005–1008 (2002)
10. Vakkalanka, S., Mohan, C.K., Kumaraswamy, R., Yegnanarayana, B.: Combining multiple evidence for video classification. In: Proceedings of the International Conference on Intelligent Sensing and Information Processing, pp. 187–192 (2005)
11. Ling-Yu, D., Min, X., Qi, T., Chang-Sheng, X., Jin, J.S.: A unified framework for semantic shot classification in sports video. IEEE Transactions on Multimedia, 1066–1083 (2005)

12. Karpenko, A., Aarabi, P.: Tiny videos: a large dataset for image and video frame categorization. In: Proceedings of the 11th IEEE International Symposium on Multimedia ISM 2009, pp. 281–289 (2009)
13. Wengang, C., Xu, D.: A novel approach of generating video scene structure. In: TENCON 2003. Conference on Convergent Technologies for Asia-Pacific Region, vol. 1, pp. 350–353 (2003)
14. Choroś, K.: Video shot selection and content-based scene detection for automatic classification of TV sports news. In: Internet – Technical Development and Applications, Book Series: Advances in Soft Computing, vol. 64, pp. 73–80. Publisher Springer, Heidelberg (2009)
15. Liu, C., Huang, Q., Jiang, S., Zhang, W.: Extracting story units in sports video based on unsupervised video scene clustering. In: IEEE International Conference on Multimedia and Expo, pp. 1605–1608 (2006)
16. Zhang, Y.J., Lu, H.B.: A hierarchical organization scheme for video data. Pattern Recognition 35, 2381–2387 (2002)
17. Choroś, K.: Digital video segmentation techniques for indexing and retrieval on the Web. In: Advanced Problems of Internet Technologies. Academy of Business, Dąbrowa Górnicza, pp. 7–21 (2008)
18. Choroś, K., Gonet, M.: Effectiveness of video segmentation techniques for different categories of videos. In: New Trends in Multimedia and Network Information Systems, pp. 34–45. IOS Press, Amsterdam (2008)
19. Choroś, K., Pawlaczyk, P.: Content-based scene detection and analysis method for automatic classification of TV sports news. In: Szczuka, M. (ed.) RSCTC 2010. LNCS (LNAI), vol. 6086, pp. 120–129. Springer, Heidelberg (2010)

Part II
Data Processing in Information Systems

Chapter 9
Acoustic Radar Employing Particle Velocity Sensors

Józef Kotus and Andrzej Czyżewski

Abstract. A concept, practical realization and applications of a passive acoustic radar for automatic localization and tracking of sound sources were presented in the paper. The device consists of the new kind of multichannel miniature sound intensity sensors and a group of digital signal processing algorithms. Contrary to active radars, it does not emit a scanning beam but after receiving surroundings sounds it provides information about the directions of incoming acoustical signals. Practical examinations of the sensitivity and accuracy of the developed radar were also presented and discussed. The sensitivity of the realized acoustic radar was examined in a free sound field. Several kinds of sound signals were used, such as: pure tone from 125 to 16000 Hz, one third octave band noise in the same frequency range and impulsive sounds. The obtained results for every kind of signal groups were presented and discussed. Results from experiments show that in some cases even a small value of the signal to noise ratio was sufficient to localize the sound source correctly. A video camera can be pointed automatically to the place where the detected acoustical source is localized. Hence, information about the sound event direction can be used for the automatic and remote control of the PTZ (Pan Tilt Zoom) cameras. The automatic and continuous tracking in real time of the selected sound source movement is also possible. The proposed solution can significantly improve the functionality of traditional surveillance monitoring systems.

9.1   Introduction

A concept, practical realization and applications of a passive acoustic radar for automatic localization and tracking of sound sources were presented below. Contrary to active radars, it does not emit a scanning beam but after receiving

Józef Kotus and Andrzej Czyżewski
Multimedia Systems Department, Gdansk University of Technology,
Narutowicza 11/12, 80-233 Gdańsk, Poland
e-mail: {joseph,andcz}@sound.eti.pg.gda.pl

surroundings sounds it provides information about the directions of incoming acoustical signals. The device consist of the new kind of multichannel acoustic vector sensor (AVS) invented by the Microflown company [1] and a group of digital signal processing algorithms developed in the Multimedia System Department, Gdansk University of Technology [2]. Concerning the acoustic properties, beam forming arrays have lower frequency limitations and a line (or plane) symmetry. Data from all measurement points have to be collected and processed first in order to obtain correct results. The acoustic vector sensor approach is broad banded, works in 3D, and has a better mathematical robustness [3]. The ability of a single AVS to rapidly determine the bearing of a wideband acoustic source is of essence for numerous passive monitoring systems.

9.2 Acoustic Particle Velocity Sensors

The single acoustic vector sensor measures the acoustic particle velocity instead of the acoustic pressure which is measured by conventional microphones, see e.g. [4]. It measures the velocity of air flowing across two tiny resistive strips of platinum that are heated to about 200°C, see Fig. 1. It operates in a flow range of 10 nm/s up to about 1 m/s. A first order approximation shows no cooling down of the sensors, however particle velocity causes the temperature distribution of both wires to alter. The total temperature distribution causes both wires to differ in temperature. Because it is a linear system, the total temperature distribution is simply the sum of the temperature distributions of the two single wires. Due to the convective heat transfer, the upstream sensor is heated less by the downstream sensor and vice versa. Due to this operation principle, the Microflown can distinguish between positive and negative velocity direction and it is much more sensitive than a single hot wire anemometer and because it measures the temperature difference, the sensitivity is (almost) not temperature sensitive [5].
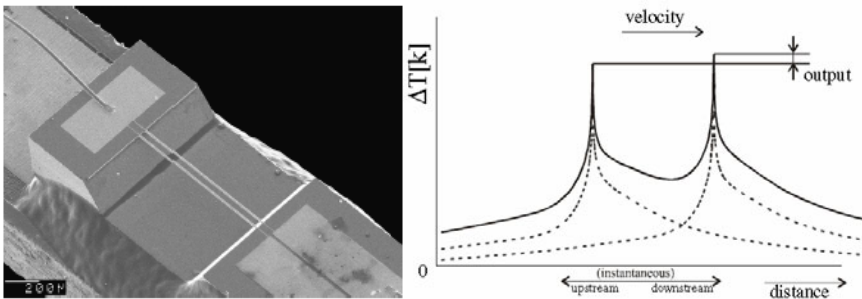
Fig. 1 (Left) A microscope picture of a standard Microflown. (Right) Dotted line: temperature distribution due to convection for two heaters. Both heaters have the same temperature. Solid line: sum of two single temperature functions: a temperature difference occurs [5]

Each particle velocity sensor is sensitive in only one direction, so three orthogonally placed particle velocity sensors have to be used. In combination with a pressure microphone, the sound field in a single point is fully characterized and also the acoustic intensity vector, which is the product of the pressure and particle velocity, can be determined [6]. This intensity vector indicates the acoustic energy flow. With a compact probe as given in Fig. 2, the full three dimensional sound intensity vector can be determined within the full audible frequency range of 20 Hz up to 20 kHz.
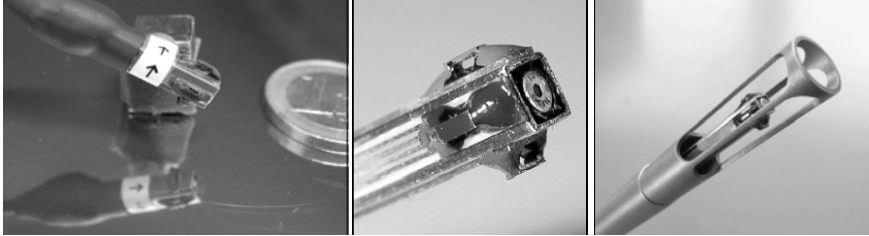
Fig. 2 A standard, three dimensional sound probe (three orthogonally placed Microflowns and a 1/10" sound pressure microphone in the middle). For size comparison one Euro is shown [5]

9.3 The Algorithm of the Acoustic Radar

The algorithm of the passive acoustic radar is based on 3D sound intensity component determination. Its diagram was presented in Fig. 3. In the first step, the particular acoustic signals are captured and prepared for frequency analysis. In the second step, the dominant frequency of the sound is estimated based on the FFT coefficients and using Quinn's First Estimator [7]. Next, the frequency value is used to design the narrow-band recursive filter [8]. The result of the recursive filtration is finally used to compute the particular sound intensity components.

Sound intensity is the average rate at which sound energy is transmitted through a unit area perpendicular to the specified direction at the point considered. The intensity in a certain direction is the product of sound pressure (scalar) $p(t)$
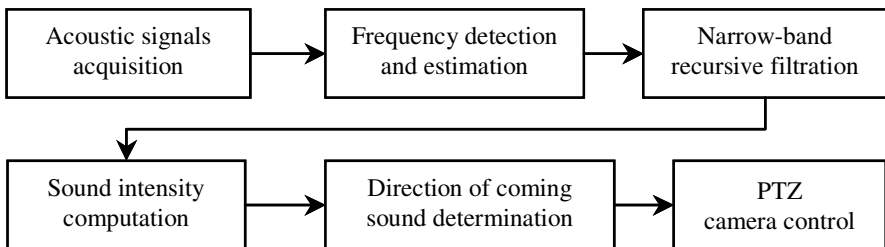
Fig. 3 The block diagram of the passive acoustic radar algorithm

and the particle velocity (vector) component in that direction $u(t)$. The time averaged intensity I in a single direction is given by Eq. 1 [5]:

$$I = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} p(t)u(t)dt \tag{9.1}$$

It is important to emphasize that using the presented 3D AVS, the particular sound intensity components can be simply obtained just based on Eq. 1. The sound intensity vector in three dimensions is composed of the acoustic intensities in three orthogonal directions (x,y,z) and is given by Eq. 2 [9]:

$$\vec{I} = I_x\vec{e}_x + I_y\vec{e}_y + I_z\vec{e}_z \tag{9.2}$$

In the presented algorithm the time average T was equal to 4096 samples (sampling frequency was equal to 48000 Hz). It means that the direction of the sound source was updated more than 10 times per second.

9.4 Practical Evaluation of the Acoustic Radar

The practical examinations of the sensitivity and accuracy of the developed radar were conducted in an anechoic chamber (free field). Several kinds of sound signals were used, such as: pure tone from 125 to 16000 Hz, one third octave band noise in the same frequency range and impulsive sounds. The set-up of the measuring system was presented in Figs. 4 and 5.

The acoustic sound pressure level was additionally independently determined using a Bruel&Kjær PULSE measurement system type 7540 with microphone type 4189, calibrated before the measurements using acoustic calibrator type 4231. The intensity probe and measuring microphone were located in the same place to ensure an identical acoustic field condition. The pure tone and noise test signal for a particular frequency were presented twice. The first time only the test signal from one loudspeaker was emitted. The test signal has two phases: the starting phase with a constant sound level of the sound and the decay phase in which the
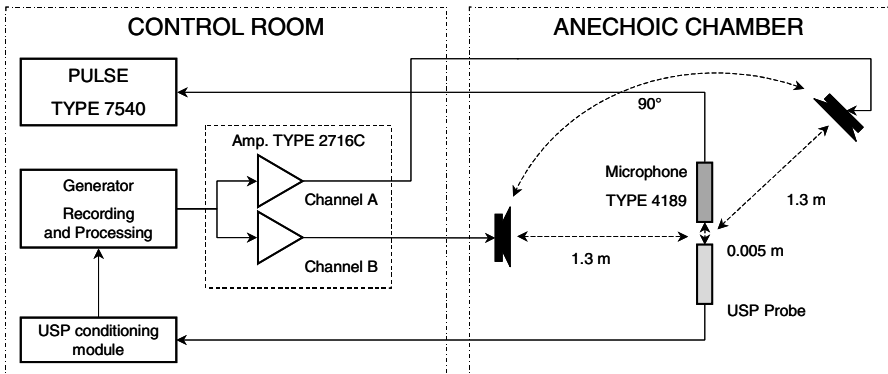
Fig. 4 Block diagram and equipments used during the measurements
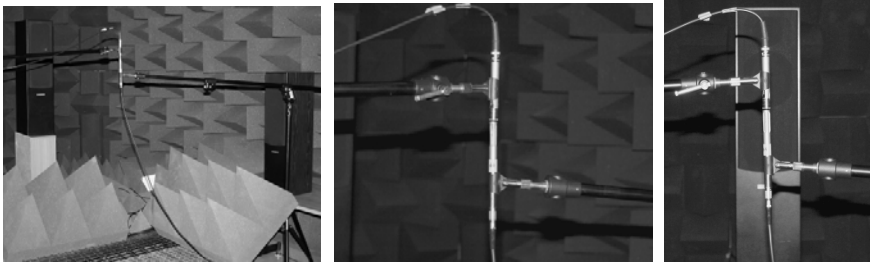
Fig. 5 Details of the measurements set-up

sound level was monotonously decreased by 1dB/s. Next, the same test signal was presented simultaneously with an additional disturbing pink noise. For both sessions the sound pressure level and angle value were noticed. Additionally, the sound level of the background noise for both sessions was determined. Such kind of data were used to properly compute the sensitivity of the radar expressed by the absolute sound pressure level and its accuracy in the disturbing conditions expressed by the Signal-To-Noise ratio as is in Eq. 3, for particular frequencies [10].

$$SNR_{dB} = SPL_{Signal\ dB} - SPL_{Noise\ dB} \tag{9.3}$$

To properly obtain the SNR_{dB} indicator, two sessions of measurements were required. During the first session, the $SPL_{Signal\ dB}$ was determined. In the next session, the background noise level was obtained ($SPL_{Noise\ dB}$). For that session the test signal was presented from one loudspeaker and the noise from another loudspeaker. The values of the angle of the sound source were determined for these conditions.

9.5   Measurement Results

The measurements were performed for different configurations of the AVS signal conditioning module. The frequency correction for the particle velocity channels could be switched off or on. The example measurement results were presented for both configurations of the conditioning module, but combined results were presented only for correction switched.

9.5.1   The Pure Tone Measurement Results

In Fig. 6 the example measurement results for 1kHz pure tone were presented. The disturbing noise sound source was off. The black line was obtained when the frequency correction was switched off. The greatest error values were obtained in such a case. When the frequency correction was switched on (blue line), the error level essentially decreased. The application of the recursive filtration algorithm for
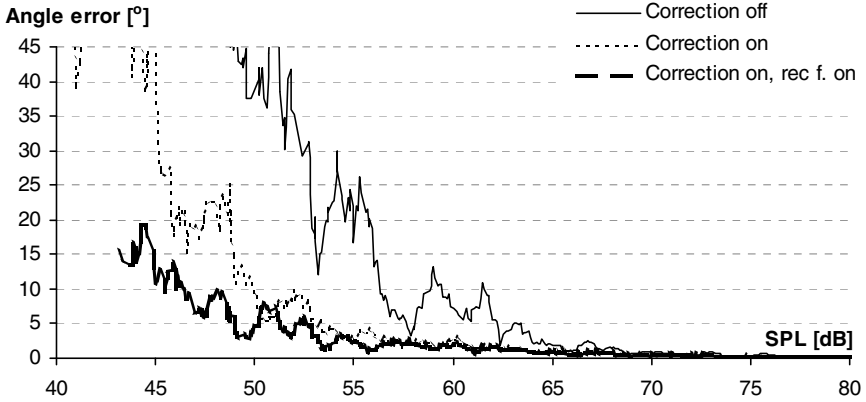
Fig. 6 Example measurement results for pure tone (1000 Hz), reference level: 20 μPa.
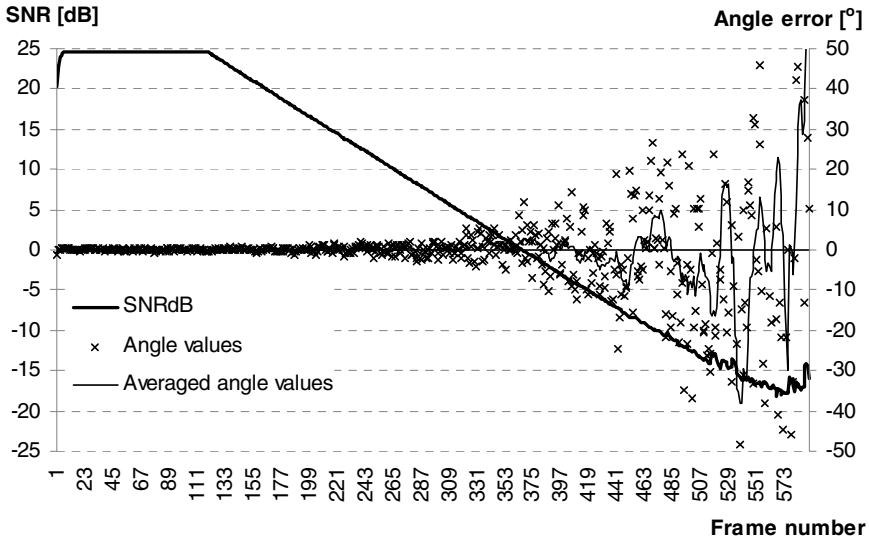
Fig. 7 Example angle error as a function of the *SNR* level for 2000 Hz

the given frequency additionally increased the accuracy of the sound source localization. The broadband background noise level for that measurement was equal to 45 dB SPL.

In Fig. 7 the example angle error as a function *SNR* level for 2000 Hz pure tone was presented. In that case, the disturbing noise source was on and its level was equal to 62 dB SPL. The recursive filtration was applied. The red line presents the SNR_{dB} values. The black line was used for averaged angle values. Small blue
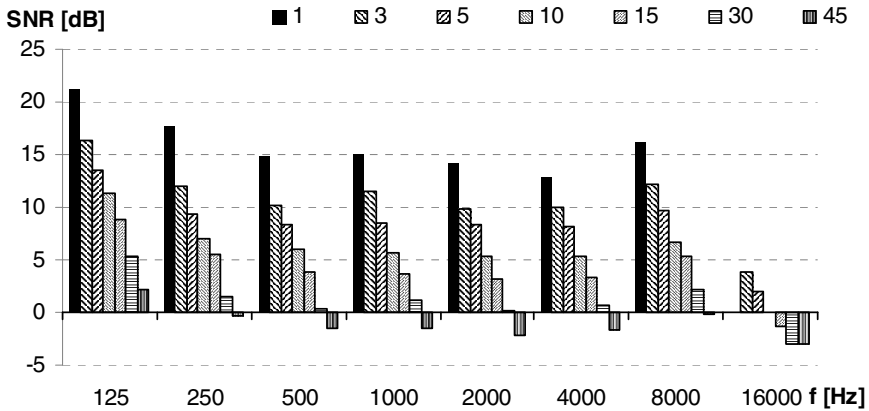
Fig. 8 Combined SNR_{dB} results for all examined frequencies. The recursive filtration was not applied
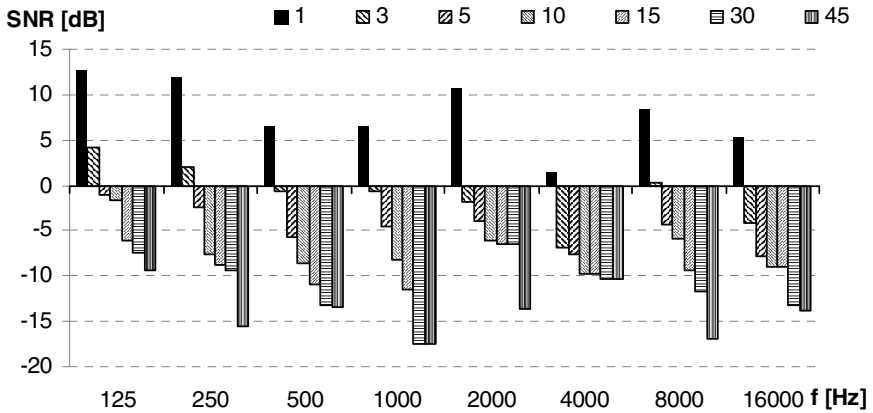
Fig. 9 Combined SNR_{dB} results for all examined frequencies. The recursive filtration was applied

crosses indicate the particular values of the angle error. It is important to emphasize that very high accuracy is obtained, even for negative values of the *SNR*. A rough estimate of the coming sound direction was obtained for extremely low *SNR* values.

The combined SNR_{dB} results for all examined frequencies were presented in Figs. 8 and 9 (the recursive filtration applied). The values were assigned for given accuracy levels: $\pm 1^{\circ}$, $\pm 3^{\circ}$, $\pm 5^{\circ}$, $\pm 10^{\circ}$, $\pm 15^{\circ}$, $\pm 30^{\circ}$ and $\pm 45^{\circ}$. Taking the obtained results into consideration, it was asserted that the developed algorithm of the passive acoustic radar have very good features in continuous pure tone tracking for all considered frequencies.

9.5.2   One-Third Octave Band Noise Measurement Results

For that kind of measurements, the noise signals limited to the one third octave band were used. The same numbers of test signals and presenting methodology were applied. In such case, the recursive filtration was not used. In Fig. 10 the example results for the one third octave band noise, centered at 1000 Hz, were shown. The black line was used to mark the averaged angle error for switched off frequency correction and the blue line was obtained for switched on frequency correction in the conditioning unit. The background noise level was equal to 45 dB SPL. The computed angle errors in comparison to pure tone examination are relatively similar. In Fig. 11 the combined SNR_{dB} results for all examined frequencies were presented.
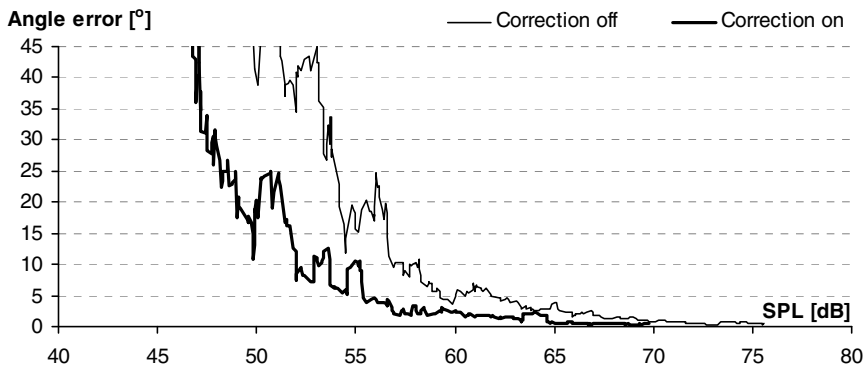
Fig. 10 Example results for one third octave band noise. Centre freq. equal to 1000 Hz
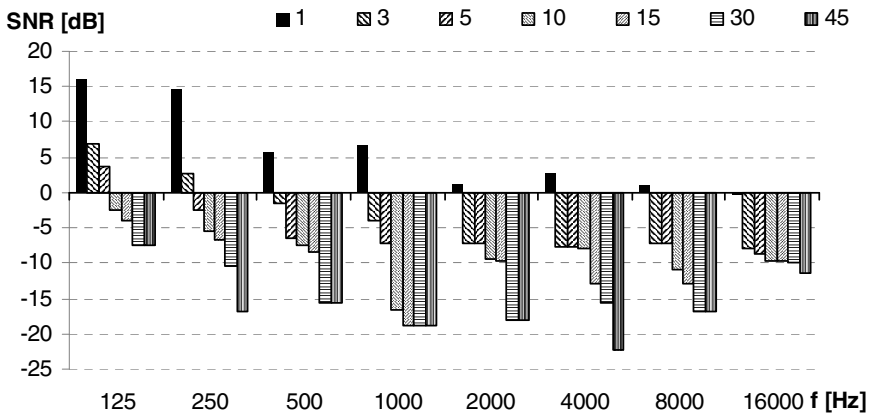
Fig. 11 The combined SNR_{dB} results for all examined frequencies

9.5.3 The Impulsive Sounds Measurement Results

In acoustic radar examinations, several kinds of impulsive sounds were used. During the first session the noise-like burst for different time lengths were used. The impulse length was equal to respectively: 32, 64, 128, 256, 512, 1024, 2048, 4096, 8192, 16384, 32768 samples. It corresponds to time periods from 0.0007s to 0.6827s. The next signal was based on a 4000 Hz tone burst with the same sample lengths. The level of that signal was constant and was 30 dB greater than the background noise (45 dB SPL). The results obtained for the particular kind of tests were presented in Fig. 9.12. For the tonal burst the recursive filtration was also applied (blue line). Another kind of impulsive test signal employed 4000 Hz.
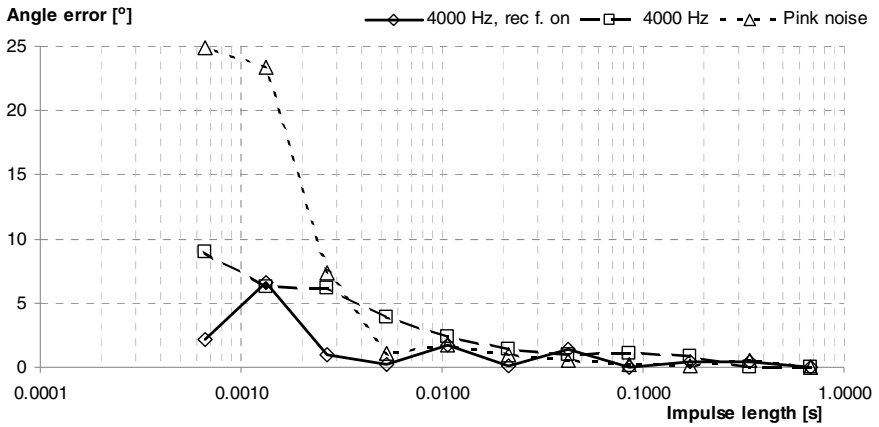
Fig. 12 An example of results obtained for different types of impulsive sounds. Noise and tonal burst with different lengths were used
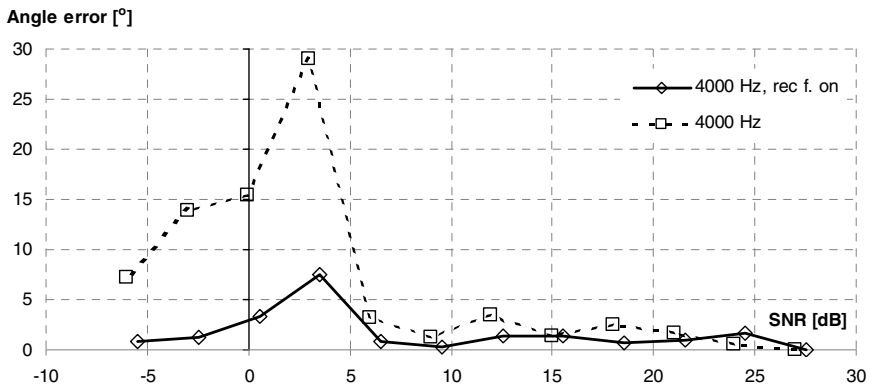
Fig. 13 Results obtained for impulse with different SPL level

It has a 4096 sample length and its amplification was decreased by 3 dB in 12 steps from SNR_{dB} = 27 to SNR_{dB} = -6. The obtained results were presented in Fig. 13. The recursive filtration was also applied (blue line).

9.5.4 PTZ Camera Control

Based on information about the direction of the given sound event occurrence, the PTZ digital camera was used to capture the motion images of the field, where that event took place. Even the angle error estimation of around 30° could be enough to effectively perceive the entire analyzed situation and can help to assess it correctly.

9.6 Conclusions

A concept and testing results of the passive acoustic radar were presented in the paper. Diverse types of test signals were used. Taking the obtained results of the realized experiments into consideration, it was ascertained that even an inconsiderable value of the signal to noise ratio was sufficient to localize sound source suitably (SNR_{dB} near to 0 dB). The application of recursive filtration can significantly improve sensitivity and accuracy of the acoustic radar (SNR_{dB} below -10 dB for tonal components). Such a kind of filtration can be used to discriminate between multiple sources. Examinations using impulsive sound were indicated which show that even extremely short and relatively quiet impulses could by properly detected and localized. Moreover, the automatic and continuous tracking of the selected sound source movement in real time is also possible. Additional procedures such as: sound source classification module or automatic control of the digital PTZ camera can be used to extend the usefulness of the presented device. Authors believe that the proposed device potentially can significantly improve the functionality of traditional surveillance monitoring systems.

References

1. Microflown Technologies – Home, http://www.microflown.com
2. Multimedia Systems Department Gdańsk University of Technology, http://sound.eti.pg.gda.pl
3. Hawkes, M., Nehorai, A.: Wideband Source Localization Using a Distributed Acoustic Vector-Sensor Array. IEEE Transactions on Signal Processing 51(6), 1479–1491 (2003)
4. de Bree, H.-E.: The Microflown: An acoustic particle velocity sensor. Acoustics Australia 31, 91–94 (2003)
5. de Bree, H.-E.: The Microflown, E-book: http://www.microflown.com/r&d_books_Ebook_Microflown.htm

6. de Vries, J., de Bree, H.-E.: Scan & Listen: a simple and fast method to find sources, SAE Brazil (2008)
7. Quinn, B.G.: Estimating frequency by interpolation using Fourier coefficients. IEEE Transactions on Signal Processing 42(5), 1264–1268 (1994)
8. Smith, S.W.: The Scientist and Engineer's Guide to Digital Signal Processing. California Technical Publishing (1997)
9. Basten, T., de Bree, H.-E., Tijs, E.: Localization and tracking of aircraft with ground based 3D sound probes. In: European Rotorcraft Forum, Kazan, Russia, vol. 33 (2007)
10. Lagrange, M., Marchand, S.: Estimating the Instantaneous Frequency of Sinusoidal Components Using Phase-Based Methods. J. Audio Eng. Soc. 55(5), 385–399 (2007)

Chapter 10
Superresolution Algorithm to Video Surveillance System

Tomasz Merta and Andrzej Czyżewski

Abstract. An application of a multiframe SR (superresolution) algorithm applied to video monitoring is described. The video signal generated by various types of video cameras with different parameters and signal distortions which may be very problematic for superresolution algorithms. The paper focuses on disadvantages in video signal which occur in video surveillance systems. Especially motion estimation and its influence on superresolution effectiveness is analyzed. In proposed initial solution a proper frame shift estimation is shown. Tests of the proposed algorithm performed video frames from real surveillance system in which many described difficulties were found. Result image examples show image resolution enhancement with plate numbers. The improvement of image quality is discussed in reference to further plate recognition.

10.1   Introduction

The general idea of superresolution algorithm is to enlarge the resolution and quality of original LR images. SR originally was used by spy satellites, in military, intelligence service. Presently SR is commonly used in medical diagnostics. Quality of ultrasonic, magnetic resonance pictures and videos are improved for better diseases detection. Also in video industry commercial SR are used to improve not only one frame but generally full video stream. Demonstration of such system is described in [11]. Generally SR can be based on one frame with some additional features of the image as well as on many frames. Typical example of basic methods which enhance single LR image is cubic filter with sharpening algorithms, eg. Wiener deconvolution [9].

This paper focuses on the multiframe superresolution algorithm which create a single high resolution frame using many low resolution frames. Pixels form many

---

Tomasz Merta and Andrzej Czyżewski
Gdansk University of Technology, Multimedia Systems Departament, Gdansk, Poland
e-mail: {tommerta,andcz}@sound.eti.pg.gda.pl

low resolution (LR) images are moved appropriately into high resolution (HR) grid [1]. A model of a multiframe SR algorithm is described in Section 2. Next sections discusses SR application in a surveillance system including system features which affect the quality of the output HR image. Finally an initial SR method for surveillance system and further image processing is proposed and tested with low quality images.

10.2   Multiframe Superresolution Algorithm

A model of SR algorithm assumes the original HR scene comes from CCD sensor. Original image is modified by motion of a camera. Additional blur occurs due to atmospheric turbulence, camera lens, dynamic movement of object. Resolution of the HR blurred image is decreased by decimation which results in LR image [1]. The model also assumes the presence of disturbing noise. Modifications of a single HR frame can be written with formula:

$$Y_k = DHF_k X + n \,, \tag{10.1}$$

where D is a downsampling operator which changes HR image into LR image, H is a blur operator caused by camera lens and atmospheric turbulence, F is a geometric motion operator which is determined by motion between original and actual LR frame, X is the original HR image, n is a present noise.

Operation of decimation decreases image grid by a given factor. Usually the operation simply copies every k-th pixel to a new grid, where k is the decimation factor. Blur operator is executed by convolving image with Point Spread Function (PSF). The real (apriori) blur caused both by camera lens and atmospheric factor is unknown. An exact value of PSF must be approximated for example with Gaussian blur. Apriori information about geometric motion is also unknown but contrary to blur effect motion parameters can be estimated [7].

Discussed model shows how to determine many LR images from HR image. Multiframe superresolution algorithm has an opposite function so similar inverse model and inverse operations should be used. A sequence of N frames is processed according to equation (2):

$$X = \sum_{l=1}^{N} D^{-1} H^{-1} F_l^{-1} Y_l \tag{10.2}$$

In the first operation LR image is shifted with F_l^{-1} operator to eliminate motion. Properly registered LR image is deblurred with operator H^{-1}. Due to lack of information about PSF deblurring may be applied with Wiener filter. Present noise was removed from equation (2) because pure denoising is not considered in this paper. Operator D^{-1} upsamples LR frame to HR grid. In enlarged image only one pixel for k^2-1 pixels has a value taken from processed LR image. Other pixel values are taken from next frames. This interpolation (with decimation) might be done in two ways:

- interpolation is done separately after frame registration
- frame registration and interpolation are implemented in one process.

In the first case SR methods [4] are focused on the process of interpolation. Missing pixels values are computed from other images with specified weights. Another method uses Delaunay triangulation [6, 8] to make irregular grid and then recalculate the value of a pixel in normal HR grid.

In the second case the whole procedure is considered as error minimalization and maximal likelihood problem. This simultaneously implemented approach uses L_1 and L_2 regularization functions as well as a cost function [3]. Best achieved results as far as denoising and deblurring is considered has total variation regularization method discussed in [2, 10].

All described methods are vulnerable to motion estimation which may cause inappropriate results. Output image may be even less readable than single LR image. Recalled situations may occur especially in a monitoring systems.

10.3 Superresolution Challenges in Surveillance System

Video signal in a surveillance system are very differentiated. Streams differ in coding format, compression, bitrate, etc. Moreover, existing systems still contain older analog video cameras. In first monitoring systems there was no need to ensure high quality video. A guard should have only watched or recorded camera view without special video processing. The main aim of a watchman was to notice moving objects and check observed area.

Present systems are based on more standardized digital equipment with resolution order of magnitude 1 Mpixel. The standard of a video coding is differentiated as well as CCD parameters.

Differentiated technologies, quality of signal cause many inconveniences in proper working of a SR algorithm. In case of older systems where analog video signal is split and converted to digital form additional distortions may occur and some modification cannot be described by static blurring effect.

Motion estimation described in previous section is the most problematic aspect especially in monitoring system. Images of small objects or fragments of objects often do not contain distinct differences, sharp edges which are used for estimation. Moreover large zoom with simultaneous large movement of objects causes large shift between next frames. Displacement, greater than analysis window in case of many estimation algorithms (e.g. block matching) causes improper results. Level of inaccuracy in such situation is measured not in fragments of pixel but in tens of pixels what brings about SR failure. An example with bad results of motion estimation is shown in Fig. 10.1. An example video was taken from [11]. Image in Fig. 10.1(a) is the original LR frame, 10.1(b) shows SR image with bilateral shift & add algorithm. The SR image quality has slightly improved. In this example there are no visible distinct errors. Image in Fig. 10.1(c) is an SR image with great shift estimation error. As a consequence object in an image is not recognizable. Improper motion estimation of even a single frame may result in serious damages of SR image.

Motion estimation algorithms based on image pyramid are more effective for large image shifts. However when shift between two images is very large a significant part of an image is not visible in referenced frame. Unknown area of an image is also a source of estimation error.

Fig. 10.1 (a) LR source image, (b) SR image, (c) SR image with motion estimation errors

Another issue is connected with object free movement. Observed objects may move in 3 dimensions so from camera's perspective not only affine transformation but also other geometric distortions are visible. Such transformation is not only difficult to detect in real video data but its compensation is not effective. After correction of a geometric distortion the pixels describing the object are highly modified. When one part of an object is adjusted often another part may be distorted what prevents to determine proper HR image.

In more specific situation some parts of an object may move in different way than the object itself. SR algorithm will only improve resolution of static parts of a moving object. Nevertheless changing fragment of an object may bring about motion estimation errors.

Of course effectiveness of SR algorithm is highly dependent on the number of frames. The more frames are used, the more blurred HR image may be due to dynamic changes of an object. A rational number of frames is about 25 which corresponds to time period 1 sec and 1.7 sec of a video respectively for 25 fps and 15 fps video stream. This amount of time is long enough to record dynamic changes of an object which disrupt both motion estimation and interpolation process.

Considering described problems proper motion estimation in a surveillance system should work properly with large image shifts. Because of possible estimation errors results of estimation should be statistically processed. During about 1 sec video image the motion can be observed and traced. In successive video frames a trend in movement of an object is visible. This movement is continuous what gives information about possible estimation errors. If there are rapid changes in movements trend an estimation error must occurred.

10.4 Algorithm

According to features and inconveniences related to SR algorithm experiments were focused on motion estimator. Implementation of this method was done in Visual Studio C++ with OpenCV 2.0 and it is based on optical flow using iterative Lukas-Kanade method with pyramids.

Because of estimation errors which occur in processing very small images, original LR images were resized using simple cubic filtration. The whole process is described in block diagram in Fig. 10.2.
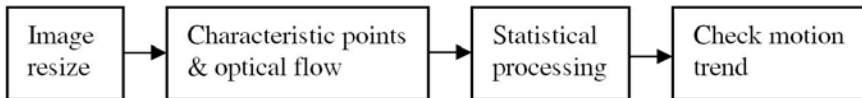
Fig. 10.2 Block diagram of motion estimation algorithm

In resized frames algorithm searches for strong corners in every image. All most prominent corners are located with subpixel precision. The number of corners depends on sharpness and resolution of an image. If too small number of corners is detected image enlarging factor is increased. Extracted points are used in Lukas-Kanade optical flow to calculate picture shift. Determined picture shift from all corners is processed in histogram which discard most infrequent results. In this attempt experimentally 50% less frequently appearing results are taken into account, so 50% of false prominent corners are rejected. Final shift value is calculated as a mean of determined results. Such scenario is done separately for vertical and horizontal shift. After estimating shift for every frame a trend is checked. If large non-subpixel shift between two frames is distinctly different from existing trend frame is not taken into final interpolation.

Such attempt is not robust against object rotation and all changes which cannot be modeled with affine transformations. In this implementation it was assumed that such geometric distortions will not be considered. Despite right estimation and compensation the existence of such distortions downgrade the quality of interpolated HR frame. Moreover in this attempt relevant object rotation recorded by a well installed video camera was not observed.

10.5 Experiment

In the experiment the analyzed frames concerned mainly vehicle number plate. Video stream was generated by Axis Q7406 video encoder which coded analog stream from existing surveillance system. The quality of video is considerably low because system is not new and original analog video signal is sent both to Axis coder and guards displays. Received signal resolution is 704x576 video stream compressed with MJPEG codec.

Fig. 10.3 Original color successive frames

Fig. 10.4 Processed frames before optical flow processing
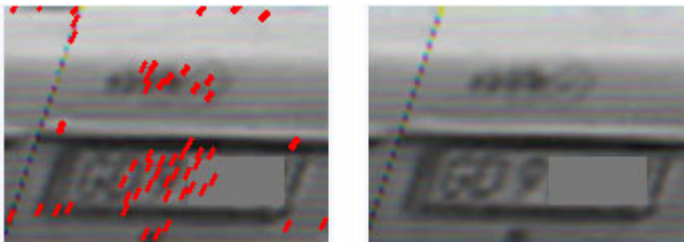
Fig. 10.5 (a) frame with marked optical flow, (b) second frame after correction shift

Resolution of cropped frame with registration number is only 81x59 pixels. Part of the registration number was hidden for privacy reasons. This modification was applied only to output frames and had no influence during processing.

Fig. 3. shows two original successive video frames received from Axis coder. Image shift between frames is clearly visible. Also some artifacts caused by compression are perceptible. In Fig. 10.4 the same successive frames are shown before optical flow process. Shift estimation bases precisely on those two images. The results of optical flow and shift estimation are visible respectively in Fig. 10.5(a) and 10.5(b).

In experiment shifted frames are processed with interpolation based on Delanuay triangulation accessible in Matlab in function *griddata* . Size of calculated HR frame is increased two times comparing to LR frame. Subjective visible

Fig. 10.6 SR based on Delaunay triangulation

increase of image resolution is smaller than increase of image size. The example result of proposed SR algorithm based on 15 frames is shown in Fig. 10.6.

In resulted frame improvement is not clearly visible. Edges are not sharpen but overall quality is improved what is visible while zooming the image. Final image was not additionally sharpened to see the plain effect after SR interpolation. The reason why quality of an image has not improved significantly is the low quality of LR image. Medium compression and analog distortions most likely cause considerable difficulties for SR algorithm.

10.6   Conclusion

Creating a robust superresolution algorithm in a surveillance system is not an easy task. Described experiment showed that in proposed SR algorithm, especially shift estimation, can work in video monitoring but the quality of received image may not be highly improved. Discussed attempt is a basic solution which does not allow rotation and other more complex geometric distortions.

In the future this method should be improved including interpolation algorithm. There is also a need for measuring what are the necessary demandings stated to cameras and the overall surveillance system to enable highly improved SR images. In further research the influence of atmospheric conditions and changes of the light should be explored to increase surveillance system effectiveness.

References

1. Elad, M., Feuer, A.: Restoration of a Single Superresolution Image from Several Blurred Noisy and Undersampled Measured Images. IEEE Trans. on Image Proc. 6(12) (1997)
2. Farsiu, S., Elad, M., Milanfar, P.: Multi-frame Demosaicing and Super-resolution of Color Images. IEEE Trans. on Image Proc. 15(1), 141–159 (2006)
3. Farsiu, S., Robinson, M.D.: Fast and Robust Multiframe Super Resolution. IEEE Trans. on Image Proc. 13(10) (2004)
4. Gilman, A., Bailey, D.G., Marshal, S.R.: Interpolation Models for Super-resolution. In: 4th IEEE Int. Symposium on Electronic Design Test & Applications, DELTA (2008)
5. Krylov, A.S., Lukin, A.S., Nasonov, A.V.: Edge-preserving nonlinear iterative image resampling method. In: 16th IEEE Int. Conf. on Image Proc. (2009)
6. Lertrattanapanichand, S., Bose, N.K.: High Resolution Image Formation From Low Resolution Frames Using Delaunay Triangulation. IEEE Trans. on Image Process 11(12) (2002)
7. Park, S.C., Park, M.K., Kang, M.G.: Super-Resolution Image Reconstruction: A Technical Overview. IEEE Signal Proc. Mag. 20, 21–36 (2003)
8. Sánchez-Beatoand, A., Pajares, G.: Noniterative Interpolation-Based Super-Resolution Minimizing Aliasing in the Reconstructed Image. IEEE Trans. on Image Proc. 17(10) (2008)
9. Sroubek, F., Cristobal, G., Flusser, J.: Simultaneous super-resolution and blind Deconvolution. In: 4th AIP Int. Conf. and 1st Congress of IPIA J. of Phys. Conf. Ser., vol. 124, 012048 (2008)
10. Zomet, A., Rav-Acha, A., Peleg, S.: Robust Super-resolution. In: Proc of 2001 IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition, vol. 1(1), pp. 645–650 (2001)
11. Topaz video enhancement video technology, `http://www.topazlabs.com`

Chapter 11
Social Network Analysis in Corporate Management

Sebastian Palus and Przemysław Kazienko

Abstract. The chapter provides an overview of essential analyses and comparisons helpful in corporate human resources management based on social network approach. Several ideas, measurements, interpretations and evaluation methods are presented and discussed, in particular group detection, centrality degree, dynamic analysis, social concept networks.

11.1   Introduction

Over the past few years, corporations have evolved from sets of individual units to collaborating social beings. Recent companies are implementing various ideas to help their employees to get known and co-operate with each other and therefore improve performance of their work. Some of them are company integration events, trips and more fresh as well as less expensive intranet social websites.

Hence, people get into various relationships due to their different job activities. Based on these relationships, a typical social network describing organizational connections can be created. These social connections between employees can be extracted from the data about pure communication like email exchange, phone calls or teleconferences. This chapter describes a general social network approach to help analyzing the knowledge flow in the organization [12] and therefore supporting corporate management.

Each company or organization can be compared to a living organism [5]. Like in the nature, each unit is dependent on others and only altogether they really form a complete system. Nevertheless, the essential part of a human body is the nervous system which steers and supervises all other processes. The similar role plays the knowledge flow for the corporate lifecycle. Thus, analysis and optimization of communication efficiency within organization is very important. Such analysis

---

Sebastian Palus and Przemysław Kazienko
Wroclaw University of Technology, Institute of Informatics
Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland
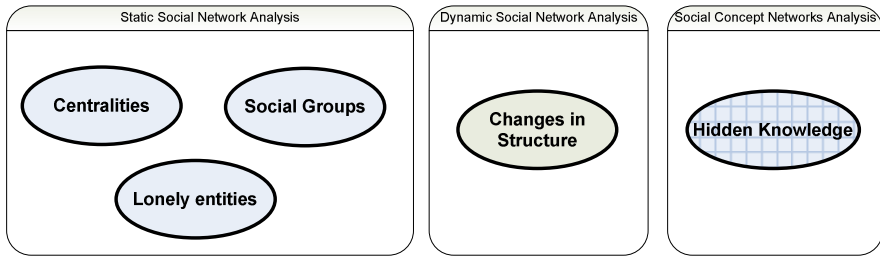
Fig. 11.1 Social Network Analysis in Organization

can detect invisible anomalies and suggest some improvements in managing corporate policies, hierarchy structure and social approach to employees.

11.2 Social Network Approach to Corporate Assessment

11.2.1 Social Network Extraction

A corporate social network can be extracted from various IT systems utilized in the organization, in particular from internal communication like email logs, phone billings and from common activities, e.g. events, meetings, projects, etc. Some other sources for the social network extraction may be intranet community forums and physical location of workplaces, i.e. fact of sitting in the same office room.
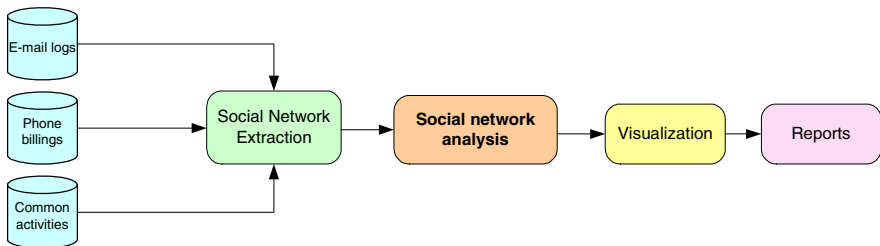
Fig. 11.2 System architecture overview

A process of generating social network requires to determine the objects connecting people – concentrating humans activities. These can be an email message, for which two roles of users can be distinguished: email sender and email recipient. Some additional recipients extracted from fields 'To', 'Cc', 'Bcc' can, in turn, effect relationship strengths. A phone call object can be treated similarly with roles of caller and receiver. These two object types (email and calls) are an example of direct relationships where actors through their mutual communication directly know about their connections.

Some other objects are container type. People are connected indirectly through them by being a part of an activity and it is not certain that they communicated directly at all. These objects are common events/meetings/projects in which some employees can participate in. For example, if two humans are both the team members in the same project, they are thought to be in common social relationship. The same approach can apply to Active Directory group memberships, forum discussions and even office room co-workers.

11.2.2 Comparison with Corporate Hierarchy

In most corporate information system structures, directory services are used to reflect internal hierarchy [3]. These directory services allow the administrator to build a hierarchy tree of organization units, departments, teams and individuals; leafs in this tree are single employees. The most common implementation of directory services is Microsoft Active Directory [6]. Assuming that information contained within directory structure is correct and up-to-date, the organizational hierarchy derived from Microsoft Active Directory can be directly used in structural comparisons, Fig. 11.3. This comparison confronts the known and visible organizational structure (extracted e.g. from Microsoft Active Directory) against the real but invisible social relationships computed from communication and common activities (structure of the social network).
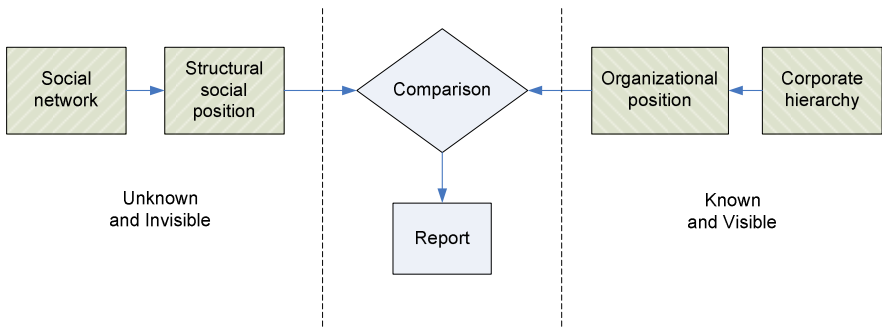
Fig. 11.3 Social Network vs. Corporate Hierarchy

Having both the formal structure of the organization and the structure of the social network we can estimate their equivalence. To be more accurate, we would like to know whether the positions and roles of actors in the network structure correspond to their official roles and positions. Summing up, all needed data can be divided into two groups: a) the data necessary to build the social network, b) the additional data for comparative analyses. Note that both source types can be automatically processed using data available in appropriate IT systems.

11.3   Static Analysis of Social Networks

11.3.1   Centralities

There are several structural measures which can be applied to static social network analysis (SNA). Primary benefit from applying graph theory analysis to social networks is the identification of most important actors [8,13,17]. In one measure, a central node of the graph is the one with the greatest number of connections to other nodes. Therefore, a central person in the social network is the most popular person in the certain community (local centrality) or in the whole network (global centrality) [9,16]. However, the main goal of the concept presented in this chapter is not to detect such actors, but to point out differences between the invisible social position (extracted from the social network in the analytical way) and the official, visible position in the organizational structure. It is suspected that the most important people in the corporation should probably be the department directors or team leaders. What happens if there are some other people with the higher centrality degree than them? Conclusions can be ambiguous. It could mean that there are some "hidden natural leaders" who may not fulfill their potential and their role in the official hierarchy may be too low [1]. However, it may also mean the opposite; the real leader position is too high for his abilities. It depends on the scale of difference between centrality degree of the real leader and the potential one. If this difference is small, there is probably no reason to apply any changes into organizational structure. For example, there can be a secretary who manages most of the business cases for the leader. Nevertheless, if the difference is high, it is highly probable that some serious changes in the hierarchy structure need to be performed to optimize team's performance.

11.3.2   Social Groups

A social network can be divided into smaller groups (subgraphs, communities). The equivalent structures in the corporate hierarchy are subtrees of teams and departments. As a result, corporate analysis of groups includes comparison of social network groups and hierarchy subtrees to detect differences. One of the possible scenario in such case is when a person is more connected to the other team members in the social network than it comes from the organizational hierarchy. As in the example in section 3.1, the possible set of actions for the management depends on the difference between link strengths to the own group in opposite to strengths of the links to other groups. If this difference is significant, then the system may suggest to move this employee to another department/team in order to improve his or her efficiency.

Other scenario is linked to the idea of developing swarm intelligence in the form of Collaborative Innovation Networks [4,5]. The groups, which are not present in the corporate hierarchy can be recognized as independent collaborative initiatives and, if given a free hand and the friendly environment, may lead to creation of new, fresh ideas. A perfect example of such structure is the Linux

kernel developers community, where members of different corporations in the whole world are working together on new ideas and taking knowledge development to the higher level based on their interests [11].

11.3.3 Lonely Entities

In each society, there are "outliers" - people who are not fitting well into their group. Such actors can be easily detected in the social network. Because of their weak ties to other people, they usually far from the centers of the network [14]. According to psychological studies, it is obvious that teamwork is not performing well because of them. Unless they are real geniuses, it is required to find out reasons of this behavior by doing internal investigation and take steps to deal with it, e.g. move them to another team or dismiss.

11.4 Dynamic Social Network Analysis

Studying dynamics of changes in social networks over time is currently one of the most interesting research topics [17]. Even in stable environments, social networks evolve. People establish new friendships as well as break others. Anyway, this chapter focuses on changes caused by HR management. Moving employees to other teams/departments, hiring new workers, dismissing others, promoting, relegating – all of these actions have a huge direct impact on social network structure. It usually requires some time to gather the appropriate data but after it happens, managers have a very powerful knowledge of rules linking HR management with corporate social structure. The analysis of dynamics in the social network, which is extracted from the long term data about user activities and communication, provides the answers to some tough questions. Which employees should be promoted? Which ones should be relegated? Which people should be fired to strengthen the others and their social ties? Which workers should be assigned to a new project? The more and diverse data is available for analysis the better prediction accuracy.

11.5 Social Concept Networks (SCN)

Concept maps are structures showing which terms are connected with each other by co-occurrence in the same object such as email message [2, 15]. Social concept networks is an idea of joining concept maps with social networks where the relationship between actors is based on email messages and common activities. Relationship strength is computed from the usage frequency of given terms/phrases in the linking objects. Actually, all SN-specific analysis can be applied to social concept networks. Interpretation of them is slightly different, though. For example, in this kind of social network, centrality degree identifies actors with the highest knowledge on given topic, experts.

Once keywords specific for a given project are declared, the social concept network immediately shows people with the expertise relevant to it. After comparing them to the group of people officially assigned to the project, it is possible to reveal actors with a "hidden knowledge", i.e. people who are not formally part of the project but are socially considered to be helpful. It can be a sign for the management to add these experts to this project and the future ones. Same analysis will also expose project members who are, in fact, not involved in discussions on project-specific topics.

11.6   Discussion

11.6.1   Profile of Relationships

Throughout the process of building and analyzing the social network only the existence of mutual communication of common activities has been considered so far. However, the fact that email message sent from person A to person B does not determine itself the nature of the relationship between A and B, e.g. whether it is rather positive or negative. The emotional character of the single message should also effect relationship strength. Nevertheless, sophisticated and powerful text recognition tools would be required to examine the profile of the messages, not mentioning the advanced forms of expression like irony, sarcasm or even the meaning of the attached images or videos. There are no effective methods for these purposes yet [10,19]. This is just to point out that even a strong relationship between two people not necessarily means that they like each other. We should always be aware that connections in the artificially extracted social network do not completely reflect the complex nature of human relationships.

11.6.2   Decision Making

All presented analysis methods should focus on two general evaluation rates: similarity of the structural position and the role both in the social network and in the official organizational hierarchy.

Having the differences between the social network and hierarchy recognized, the management of the organization can undertake appropriate decisions to decrease these differences. For example, the position in the structure can be affected by moving employees to other teams/departments. Changes of roles in the hierarchy are achieved by promoting/relegating.

Moreover, new positions and roles can be discovered in the social network which do not exist in the corporate hierarchy structure. As a result, the managers can create new positions in the organization.

11.7   Conclusions and Future Work

Social network approach to manage problems is capable to significantly improve the Human Resources efficiency by either detection of hidden anomalies in the

corporate hierarchy or making communication between employees more effective and easier. By analyzing semantics of email messages exchanged within the corporate network we are able to identify individuals with "the hidden knowledge".

Overall, social network approach to the problem of corporate management appears to be very helpful, however, all analysis need to be well interpreted to improve performance and social health of the company. This is only a tool. Still, human resources have to be managed by humans.

Future research will focus on interpretation of some measures not mentioned in this chapter, e.g. betweenness, prestige, density, etc. as well as on development of new reliable metrics for quantitative comparison of social network structures with corporate hierarchies.

References

1. Balkundi, P., Kilduff, M.: The ties that lead: A social network approach to leadership. Leadership Quarterly 16, 941–961 (2005)
2. Cañas, A.J., Carff, R., Hill, G., Carvalho, M., Arguedas, M., Eskridge, T.C., et al.: Concept maps: Integrating knowledge and information visualization. In: Tergan, S.-O., Keller, T. (eds.) Knowledge and Information Visualization. LNCS, vol. 3426, pp. 205–219. Springer, Heidelberg (2005)
3. Carter, G.: LDAP System Administration. O'Reilly Media, Sebastopol (2003)
4. Gloor, P.: Swarm Creativity, Competitive advantage through collaborative innovation networks. Oxford University Press, Oxford (2006)
5. Gloor, P.: Net Creators. Unlocking the Swarm Creativity of Cyberteams through Collaborative Innovation Networks (2004),
   `http://www.swarmcreativity.net/html/book_swarmcrea.htm`
6. Iseminger, D.: Active Directory Service for Microsoft Windows 2000 Technical Reference. Microsoft Press, Redmond (2000)
7. Kascarone, R., Paauwe, J., Zupan, N.: HR practices, interpersonal relations, and intrafirm knowledge transfer in knowledge-intensive firms: a social network perspective. Human Resource Management 48(4), 615–639 (2009)
8. Kazienko, P., Musiał, K., Zgrzywa, A.: Evaluation of Node Position Based on Email Communication. Control and Cybernetics 38(1), 67–86 (2009)
9. Klein, K.J., Lim, B., Saltz, J.L., Mayer, D.M.: How do they get there? An examination of the antecedents of centrality in team networks. Academy of Management Journal 47, 952–963 (2004)
10. Krippendorff, K.: Content Analysis: An Introduction to Its Methodology, 2nd edn. Sage, Thousand Oaks (2004)
11. Lee, G.K., Cole, R.E.: The Linux Kernel Development as a Model of Knowledge Development. Working Paper, October 25, 2000, Haas School of Business, University of California, Berkeley (2000)
12. Musiał, K., Juszczyszyn, K.: A method for evaluating organizational structure on the basis of social network analysis. Foundations of Control and Management Sciences 9, 97–108 (2008)

13. Musiał, K., Kazienko, P., Bródka, P.: User Position Measures in Social Networks. In: The third SNA-KDD Workshop on Social Network Mining and Analysis held in conjunction with The 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2009, Paris France, June 28, Article no. 6. ACM Press, New York (2009)

14. Musial, K., Juszczyszyn, K.: Properties of Bridge Nodes in Social Networks. In: Nguyen, N.T., Kowalczyk, R., Chen, S.-M. (eds.) ICCCI 2009. LNCS (LNAI), vol. 5796, pp. 357–364. Springer, Heidelberg (2009)

15. Novak, J.D., Cañas, A.J.: The Theory Underlying Concept Maps and How to Construct Them, Technical Report IHMC CmapTools 2006-01 Rev 01-2008, Florida Institute for Human and Machine Cognition (2008)

16. Scott, J.: Social network analysis: A handbook, 2nd edn. Sage, London (2000)

17. Wasserman, S., Faust, K.: Social network analysis: Methods and applications. Cambridge University Press, New York (1994)

18. White, D.R.: Network Analysis and Social Dynamics. Cybernetics and Systems 35, 173–192 (2004)

19. Wiebe, J., Wilson, T., Cardie, C.: Annotating Expressions of Opinions and Emotions in Language. Language Resources and Evaluation 1(2), 0–0 (2005)

Chapter 12
AAM Toolkit: A System for Visual Object Appearance Modeling⋆

Maciej Smiatacz and Damian Sikora

Abstract. The approach based on Active Appearance Models (AAM) can be used as a sophisticated technique of multimedia information analysis providing means for localization and recognition of objects in images or video sequences. Despite the large number of publications on AAMs it is still a challenging task to move from theoretical concepts to working implementation. In this paper we describe the software suite that allows the user to create an appearance model of any visual object. The relations between algorithmic issues and application architecture are emphasized. Preliminary experiments performed with *AAM Toolkit* are also presented.

12.1 Introduction

Efficient processing of multimedia information has become one of the most important and challenging tasks of computer science nowadays. It is closely related to artificial intelligence since the capabilities offered by human vision system are still out of reach for today's machines. Whether it comes to surveillance, law enforcement, information retrieval or entertainment applications, certain level of *image understanding* is essential. The first step towards it is to perform proper *detection* and *localization* of objects in the image. If the machine is able to name the objects, specify their position, size and orientation it may try to recognize the meaning of the scene. However, this task appears exceptionally difficult because objects vary in size and shape, and changes of illumination, perspective or background make the situation even more complicated.

Generally, this problem can be addressed in two ways: by applying the brute force and intense machine learning that involves huge number of training samples [4] or

Maciej Smiatacz and Damian Sikora
Gdansk University of Technology, Narutowicza 11/12, 80-233 Gdansk, Poland
e-mail: slowhand@eti.pg.gda.pl,Damian.Sikora@zak.eti.pg.gda.pl

by introducing some prior knowledge represented by the *model* of the object that we are looking for. The very important feature of such model is the *flexibility* that allows it to adapt to the changes in object's appearance. The model can describe the shape of the object only or may also include the information about its visual contents characterized by intensity values of pixels. Recently numerous methods of advanced modeling have been proposed and the whole family of so-called Active Appearance Models (AAM) emerged [3].

Currently we are working on the biometric security system for mobile devices, in which automatic face verification will be used extensively. We decided that the preliminary face detection will be performed by means of Viola & Jones algorithm [4], and the Active Appearance Model (as proposed in [1]) will provide final localization of features such as eyes, mouth, nose and so on. Additionally, the parameters of model matching the image can be used to support the verification process.

Consequently, we decided to create software package called *AAM Toolkit* offering the possibility of defining the configuration of the model, extracting statistical information from sample images and learning the structure that will guide the search process. The system is described in the following sections. In addition we present the results of simple experiments involving models created by our software. In some of the tests we also used our PG-AAM program [2].

It is important to notice that the localization technique based on Active Appearance Models is versatile and can be applied in different fields of computer vision, for example in medical image processing. Therefore we hope that the use of our system will not be limited to the development of face verification software only.

12.2   The Architecture of AAM Toolkit

Working on the *AAM Toolkit* architecture we decided to divide the model building process into three main blocks: creation of the training set, constructing the actual model by extracting the statistical information from training images, and learning the dependencies (represented by the so-called **A** matrix) between current model instance and the real image. In order to guarantee the clarity and reliability of the code three stand-alone applications were prepared: *Training Set Creator*, *Model Creator* and *Regression Matrix Creator*. The detailed description of Active Appearance Models can be found in [1]. We will briefly present only some theoretical concepts important from the point of view of our software architecture.

12.2.1   Training Set Creator

Active Shape Model (ASM), the predecessor of Active Appearance Models, is a structure containing information about the average shape of the object of some type and the data concerning the most characteristic shape deviations observed in the training set. AAM is an extension of the shape model including additional information about mean gray level pattern representing the object, and the description of the most important modes of object's appearance variation.

The ASM or AAM construction starts with the definition of the template of the object i.e. a set of labeled points that represent the shape. This is called the Point Distribution Model – PDM. The landmark points must be placed at equivalent locations on each of the M training examples. This way the shape of each sample object can be described as a vector containing coordinates of the landmark points. Although there were attempts to perform automatic generation of the landmark points, the safest way to create the set of training shapes is to do it manually.

At the beginning of the new project the *Training Set Creator* allows the user to select the folder with training images, load the first picture and add and position the landmark points along the border of the object. Some of the points can be linked to create a convenient visualization of the object's shape. This way we can define the Γ_i vectors that describe n points of each shape:

$$\Gamma_i = [x_{i0}, y_{i0}, x_{i1}, y_{i1}, ..., x_{ik}, y_{ik}, ..., x_{i(n-1)}, y_{i(n-1)}] \quad , \quad i = 1, ..., M \qquad (12.1)$$

To create the AAM we have to extract the intensity values of pixels located within the framework defined by Γ_i. Moreover, the image surrounded by Γ_i mask must be divided into triangles to make the warping possible at the next stage. Therefore the *Training Set Creator* performs the Delaunay triangulation and presents the result for verification. This is necessary because in some cases the triangles could be defined in such a way that some of them will overlap after changing the landmark positions on subsequent training images. If this risk becomes evident the user can rollback the triangulation and add some landmark points to make the framework more dense.

Some versions of the algorithms that fit the model to the image use the information about grey level profiles extracted along the lines perpendicular to the framework (normals) drawn from the landmark points. The length of the normals is an important parameter of the model as it defines the region that will be searched around each point during localization.

The screenshots of *Training Set Creator* are shown on Fig. 1. The triangulation results as well as normals and landmark points are visible.

Fig. 12.1 Screenshots of *Training Set Creator*.

As soon as the triangulation is accepted, the framework is locked so the new points cannot be added but their positions should be adjusted to match the objects on the subsequent images. The outcome of the program is the folder in which each training image is described by the two files: 1) the .xml file containing the coordinates of landmarks, definitions of triangles and grey level profiles extracted along the normals, 2) the .bmp file in which the red channel contains the intensity of each pixel while green and blue channels decode the identifier of the triangle to which it belongs.

12.2.2 Model Creator

In order to build the statistical model of visual appearance principal components analysis (PCA) must be applied several times. Using the $\boldsymbol{\Gamma}_i$ vectors stored in .xml files prepared by the *Training Set Creator* we can compute the mean shape

$$\bar{\boldsymbol{\Gamma}} = \frac{1}{M} \sum_{i=1}^{M} \boldsymbol{\Gamma}_i \tag{12.2}$$

and deviation from the mean for each element of the training set, i.e. $\boldsymbol{\Phi}_i = \boldsymbol{\Gamma}_i - \bar{\boldsymbol{\Gamma}}$. Now we must construct the appropriate covariance matrix

$$\mathbf{C}_s = \frac{1}{M} \sum_{i=1}^{M} \boldsymbol{\Phi}_i \boldsymbol{\Phi}_i^{\mathrm{T}} \tag{12.3}$$

and calculate its eigenvectors $\mathbf{u}_i$. A new instance of the model (i.e. the new shape of the object) can be generated by using the mean shape and the weighted sum of the most characteristic shape deviations represented by the eigenvectors $\mathbf{u}_i$, i.e.:

$$\mathbf{X} = \bar{\boldsymbol{\Gamma}} + \mathbf{P}_s \mathbf{b}_s \tag{12.4}$$

where $\mathbf{P}_s = (\mathbf{u}_1 \mathbf{u}_2 ... \mathbf{u}_K)$ is the matrix containing the first K eigenvectors of $\mathbf{C}_s$ and $\mathbf{b}_s = [b_1 b_2 ... b_K]^{\mathrm{T}}$ is the vector of their weights. Many shape modifications can be obtained by changing the appropriate $\mathbf{b}_s$ elements. However, before PCA can be applied to $\boldsymbol{\Gamma}_i$ vectors the information concerning scale and position must be removed from them by the aligning procedure so that only the shape description remains. Finally, the shape model can be stored in the .xml file containing the mean shape and eigenvectors $\mathbf{u}_k$. This way the first of the intermediate models (model S) is built.

Principal components analysis must also be applied to the data stored in the special .bmp files obtained during the creation of the training set. This time we want to prepare another intermediate model (model I), representing the distribution of pixel intensities within the object's framework. In this case two kinds of normalization are necessary. First, the bitmaps must be shape-normalized by carrying out the piecewise affine warping that transforms each of the training images into coordinates of the mean shape $\bar{\boldsymbol{\Gamma}}$. After this transformation all objects (for example all faces from the training set) have the same shape but their grey level appearances (or "textures"),

stored in $\mathbf{g}_i$ vectors, differ. The warping is performed by looking for corresponding pixels in the related triangles from the mean framework and the given shape. Next, the effect of global illumination variations must be minimized. To achieve this we try to calculate the scaling factor α and translation β so that after the following operation:

$$\mathbf{g}'_i = (\mathbf{g}_i - \beta\mathbf{1})/\alpha \tag{12.5}$$

$\mathbf{g}'_i$ is as similar to the mean "contents" of the object $\bar{\mathbf{g}}$ as possible. This normalization must be performed recursively: one of the training samples can be used as an initial estimate of $\bar{\mathbf{g}}$ and then for each image α and β parameters can be calculated as follows in subsequent iterations:

$$\alpha = \mathbf{g}_i \cdot \bar{\mathbf{g}} \quad , \quad \beta = (\mathbf{g}_i \cdot \mathbf{1})/n \tag{12.6}$$

where n is the number of pixels within the mean framework $\bar{\mathbf{\Gamma}}$. By applying the PCA to the normalized training samples $\mathbf{g}'_i$ we obtain the linear model of intensity distribution:

$$\mathbf{g} = \bar{\mathbf{g}} + \mathbf{P}_g\mathbf{b}_g \tag{12.7}$$

where $\mathbf{g}$ is the new instance of object's "contents", $\mathbf{P}_g$ denotes the set of orthogonal modes of grey-level variation (eigenvectors of $\mathbf{C}_g$ matrix) and $\mathbf{b}_g$ is the vector of model parameters. The .xml file containing $\bar{\mathbf{g}}$ and $\mathbf{P}_g$ describes intermediate I model. Fig. 2 shows the block diagram of the module responsible for intermediate models creation. One can notice that this part of the *AAM Toolkit* generates also the set of models described as H(i). Each of them is created by analyzing the image intensity variations around i-th landmark point in all training images. Thus, H(i) is the model of grey level distribution along the normal drawn from i-th landmark point and it contains mean intensity profile as well as the most important modes of intensity variations. H(i) models can appear useful in further experiments, for example to guide the process of fitting the shape model S to an image.

The *Model Creator* provides the visualization of shape and intensity models. It is possible to view the mean framework $\bar{\mathbf{\Gamma}}$, the mean bitmap $\bar{\mathbf{g}}$ and the modifications introduced by changing appropriate parameters. The influence of those parameters on the final results is illustrated on Fig. 3.

As we can see, the shape and the texture of each of the training objects can be described by the two vectors of parameters: $\mathbf{b}_s$ and $\mathbf{b}_g$. Because shapes can be correlated with intensity levels of the pixels within the objects, we use principal components analysis once again to obtain new vectors of parameters, $\mathbf{c}$, controlling the whole appearance – shape and texture at the same time. First, for each of the training samples the vectors $\mathbf{b}_s$ and $\mathbf{b}_g$ are concatenated

$$\mathbf{b} = \begin{pmatrix} \mathbf{W}_s\mathbf{b}_s \\ \mathbf{b}_g \end{pmatrix} \tag{12.8}$$

where $\mathbf{W}_s$ is the diagonal matrix of weights that compensate the differences in units between shape and grey level parameters. The simplest way to obtain $\mathbf{W}_s$ is to calculate the ratio r of sums of eigenvalues λ_{sk} and λ_{gk} of $\mathbf{C}_s$ and $\mathbf{C}_g$ matrices

$$r = \frac{\sum\limits_{k} \lambda_{gk}}{\sum\limits_{k} \lambda_{sk}} \qquad (12.9)$$

and to use r as the value of the diagonal elements of $\mathbf{W}_s$.

   If we construct the covariance matrix from all $\mathbf{b}_i$ vectors ($i = 1, ..., M$) describing training objects and create matrix $\mathbf{Q}$ from its eigenvectors we will get the following model:

$$\mathbf{b} = \mathbf{Qc} \qquad (12.10)$$

where $\mathbf{c}$ is the vector of values that encode the *appearance* of the object. Therefore, by changing the small number of appearance parameters $\mathbf{c}$ we are able to generate the larger set of values that describes the shape of the object (vector $\mathbf{b}_s$) and its texture ($\mathbf{b}_g$). Next we can prepare the "shapeless" bitmap (i.e. the vector of pixel intensities $\mathbf{g}$ within the mean framework $\bar{\mathbf{\Gamma}}$) and then warp the image according to shape parameters. This way the visualization of the new instance of the modeled object (e.g. a new face) will appear. The specialized module of *Model Creator* prepares final AAM (i.e. the .xml file containing matrix $\mathbf{Q}$ and its eigenvalues) from intermediate models of shape (S) and intensity (I).
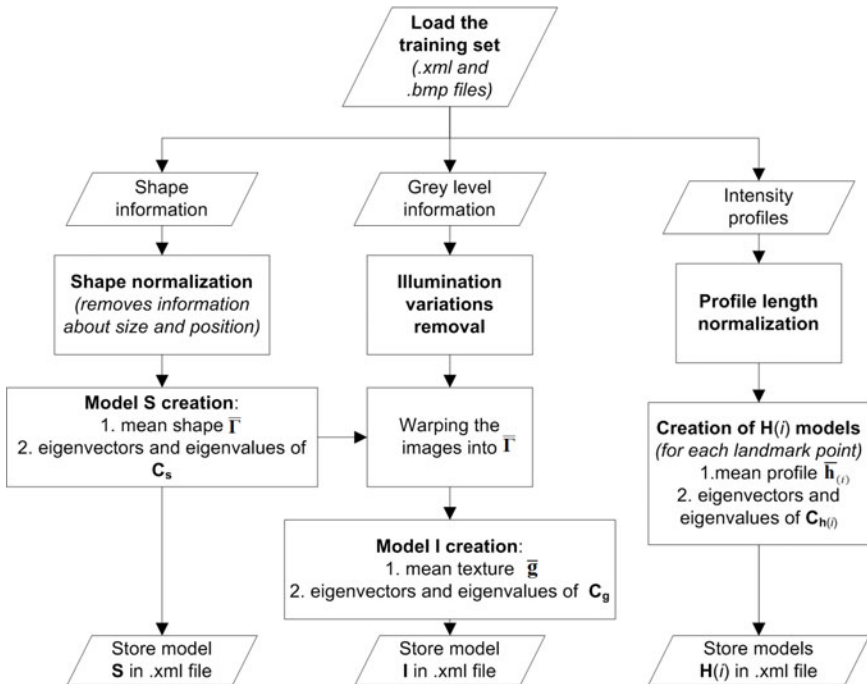
Fig. 12.2 The block diagram of the module responsible for intermediate models creation.
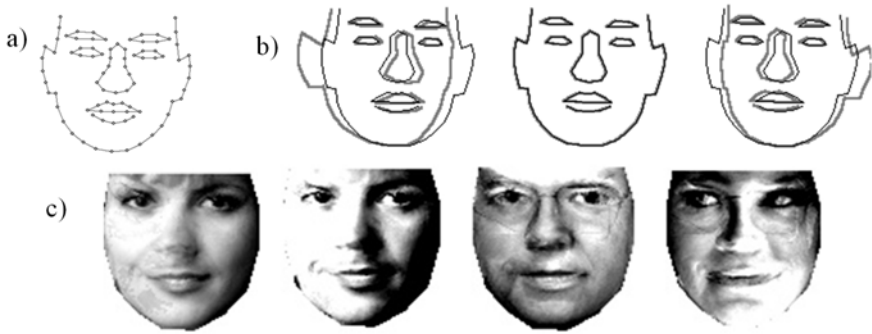
Fig. 12.3 a) PDM of a human face, b) shape modifications introduced by changing one of $\mathbf{b}_s$ elements, c) texture variations obtained by changing $\mathbf{b}_g$.

12.2.3   Regression Matrix Creator

Fitting the Active Appearance Model to the image can be treated as an optimization problem in which we want to find the vector of model parameters $\mathbf{c}$ that will minimize the difference $\Delta\mathbf{I}$ of two images:

$$\Delta\mathbf{I} = \mathbf{I} - \mathbf{I}_m \tag{12.11}$$

where $\mathbf{I}$ denotes the vector of grey-level values of the actual image and $\mathbf{I}_m$ is the corresponding vector generated by the model. We have to remember, however, that during the search the model will not only change its appearance but also the size, position and orientation. Therefore it is reasonable to extend the vector of model parameters with four values representing translation (t_x and t_x), rotation (θ) and scaling (s) of the model. This new, extended set of parameters will form the vector denoted as $\mathbf{v}$.

In general, finding $\Delta\mathbf{v}$ that minimizes $\Delta\mathbf{I}$ is a complex optimization task. In order to simplify it we can make two assumptions: 1) when the displacement is small there is a linear relation between $\Delta\mathbf{v}$ and $\Delta\mathbf{I}$, 2) since all attempts to match the model to the image are somewhat similar the relation can be learned from examples. Because we know the values of $\mathbf{c}$ (and $\mathbf{v}$) parameters for the training samples it is possible to introduce some displacements ($\Delta\mathbf{v}$) to them and to record the corresponding changes of model textures, $\Delta\mathbf{g}$. Treating $\Delta\mathbf{v}$ vectors as columns of $\mathbf{V}$ matrix and $\Delta\mathbf{g}$ vectors as columns of $\mathbf{G}$ matrix the following linear regression task can be formulated:

$$\mathbf{V} = \mathbf{A}\mathbf{G} \tag{12.12}$$

This problem can be solved by standard procedures. The *Regression Matrix Creator* from *AAM Toolkit* prepares an .xml file containing the resultant $\mathbf{A}$ matrix, using the .xml definition of appearance model as input. Number of samples that will form the $\mathbf{V}$ and $\mathbf{G}$ matrices as well as the ranges of parameter displacements can be selected with the help of graphical interface.

12.3 Experiments

The *AAM Toolkit* offers the possibility to create the active appearance model of any visual object and to calculate the **A** matrix that tells us how to change the parameters of the model to make it more similar to the real image in particular situation, described by the vector of image differences ($\Delta\mathbf{I}$) (11). The matrix could be used to guide the localization algorithm. However, matching the AAM to image is a complex issue and the discussion of different fitting strategies is beyond the scope of this work. So far we haven't tested them extensively and we plan to perform a comparative study in the nearest future. Nevertheless, we used active appearance models to carry out some experiments related to image approximation, which can be a basis of an object detection method.

During the experiments we examined the absolute difference of grey levels between the actual image **I** and its closest approximation generated by the model. The difference, denoted as *R*, was averaged over all pixels. Collection of face images was used to create the active model and to run the tests. We considered four cases: 1) the image being approximated (i.e. the testing image) was included in the training set, 2) the image was not included in the training set but other photograph of the same person was used during training, 3) the image showed a person whose face was not included in the training set, 4) the image was from outside of the training set and it presented an object not resembling a face (e.g. a flower). The experiments were repeated for training sets of different sizes and several image resolutions; the number of eigenvectors included in the model (matrix **Q**) also varied. Sample results are illustrated on Fig. 4.
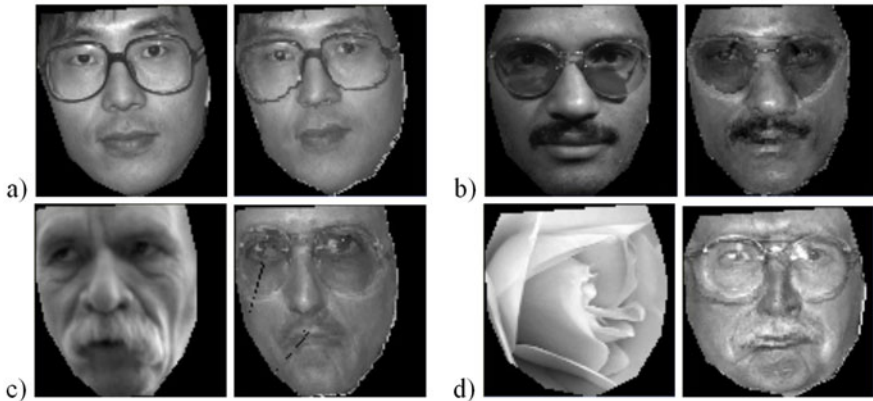
Fig. 12.4 Selected testing images and their approximations generated by the face AAM: a) a face from the training set, $R = 17.4$; b) an image from outside of the training set but representing a person whose face (shown on other photograph) was included in the training set, $R = 29.1$; c) a person not included in the training set, $R = 40.4$; d) an image that does not represent a face, $R = 59.3$.

The experiments showed that in all cases the parameter R had distinctively higher value when an image did not represent a face but some other object. This means that by comparing R with suitable threshold a straightforward face detection could be performed. Obviously, our goal is to achieve fine localization of the face features, so that the recognition module could operate on precisely cropped regions of interest. However, further research is necessary to find the optimal model fitting algorithm. So far, with the above experiments we have managed to prove that AAM Toolkit contains proper implementation of Active Appearance Model creation algorithms.

12.4   Conclusions

Active Appearance Models offer numerous possibilities of visual data processing, interpretation or recognition. It is a sophisticated method of computer vision but its practical usefulness hasn't been properly verified yet, mostly due to the complicated implementation and a large number of possible versions suggested in literature. By creating the *AAM Toolkit* we made the first step towards real-life applications of Active Appearance Models.

Although we plan to apply the technique of AAMs in our mobile face recognition system it is important to notice that the software package that we have created is quite versatile and can be used to build an active appearance model of any object that can be represented on images forming the training set.

References

1. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active Appearance Models. IEEE Trans. Patt. Analysis Mach. Intell. 23(6), 681–685 (2001)
2. Kaliszewski, F.: Active Appearance Models for Automatic Face Localization. MSc thesis (in Polish), Gdansk University of Technology (2008)
3. Matthews, I., Baker, S.: Active Appearance Models Revisited. Int. J. Comp. Vision 60(2), 135–164 (2004)
4. Viola, P., Jones, M.J.: Robust Real-Time Face Detection. Int. J. Comp. Vision 57(2), 137–154 (2004)

Chapter 13
Service Discovery Approach Based on Rough Sets for SOA Systems

Krzysztof Brzostowski, Jakub M. Tomczak, Witold Rekuć,
and Janusz Sobecki

Abstract. In the chapter an approach to solve the problem of service matchmaking and discovery is discussed. By a service we mean a independent component which has specified inputs and outputs and some functional and non-functional features. In our approach to describe services ontologies are used. It helps us to include some semantics to the services. Process of service discovery is started when user's request is translated into SLA contract. The problem of matchmaking and discovery can be defined as a problem of finding service which fulfil user's requirements as much as possible. In this work we present a system of service discovery and then our contribution to this domain – rough set-based approach to solve considered problem. Some simple example which illustrates proposed approach is shown.

13.1   Introduction

The problem of services discovery is the crucial concern in system which based on service such as Service Oriented Architecture (SOA) [2, 5]. It is very important to ensure access to the tools which help users to find desired service. Service Oriented Architecture systems offer tools to solve this problem.

In SOA systems some mechanisms and methods which help to meet expectation both service providers and service users are implemented. Generally we can denote these operations as *description*, *publication*, *discovery* and *selection*. Crucial part of SOA systems, which helps implementing mentioned operation, is ontology. Application of ontologies enables to convey the meaning of the formal terminologies to describe some real-world concepts [5]. In the service representation ontologies are

Krzysztof Brzostowski, Jakub M. Tomczak, Witold Rekuć, and Janusz Sobecki
Institute of Informatics, Wrocław University of Technology, Poland
e-mail: {Krzysztof.Brzostowski,Jakub.Tomczak}@pwr.wroc.pl
        {Witold.Rekuc,Janusz.Sobecki}@pwr.wroc.pl

used to provide the terminology utilized by each element of service to describe the relevant aspects of the certain domains.

The main purpose of this work is to present the problem of service discovery described by ontologies using rough set-based approach.

13.2 Service Discovery Problem

13.2.1 Problem Statement

In the service-oriented systems one of the most important issues is *service discovery* [1, 14]. By the term of service discovery we understand finding such services which fulfil user demand or requirement.

Let us assume that there is a set of available complex services:

$$CS = \{cs_1, cs_2, ..., cs_L\}, \qquad (13.2.1)$$

where L means the number of available services. Each complex service is described by a set of its functionalities:

$$\varphi_{cs} = \{\varphi_{cs,1}, \varphi_{cs,2}, ..., \varphi_{cs,N}\} \subset 2^{\mathsf{T}}, \qquad (13.2.2)$$

where N – number of terms describing functionalities, and T is the set of all possible terms.

Moreover, user request (demand) is represented by a vector:

$$\mathbf{d} = \{d_1 \ d_2 \ ... \ d_R\} \subset 2^{\mathsf{T}}, \qquad (13.2.3)$$

where R states for number of user requested terms. The user's request is specified due to the SLA contract (*Service Level Agreement*) which is obtained after contract negotiation and translation. Further, let us assume that there exists *the best* service cs^* which fulfils user request $\mathbf{d}$. Term *the best* is defined by following criterion

$$Q_d(cs) = q(cs, \mathbf{d}), \qquad (13.2.4)$$

where $q(\cdot, \cdot)$ is the error between service and user request. Now, having (13.2.3) and (13.2.4) we can formulate the problem of service discovery as an optimization problem,

$$Q_d(cs^*) = \max_{cs \in CS} Q_d(cs). \qquad (13.2.5)$$

To solve this problem we can propose following algorithm

$$U = \Psi(CS, \mathbf{d}). \qquad (13.2.6)$$

where $U = \{u \in CS : Q(u) = \max_{cs \in CS} Q(cs)\}$. In particular U is a singleton (one solution) or an empty set.

13.2.2   Service Discovery in SOA Systems

In order to solve the problem of service discovery presented in previous section we consider system presented in Fig. 13.2.1. Decision making in the SOA system consists of following steps.

First, user's demand (request) is transformed due to negotiation into SLA contract. It is worth mentioning that negotiation and translation use knowledge base for: 1) checking if user is registered to the system; 2) demand translation using domain ontologies into format readable for the system. This step is crucial for further proper service matching. More details are presented in Section 13.2.3.

Second, SLA contract is used in service discovery. Algorithms for service matching seek for services from given service repository which are the best due to chosen criterion. More details are presented in Section 13.2.4.
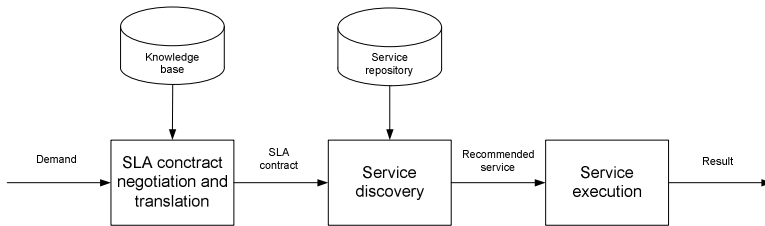
Fig. 13.2.1 Service discovery system for SOA

Third, system executes recommended service and result is presented to the user. This task is not considered in this chapter. However, very interesting results could be found in [6].

In the next section some general approach for formulating algorithms (13.2.6) is described. This is the clue matter in the service matchmaking and discovery.

13.2.3   SLA Contract Negotiation and Translation Using Ontologies

In ontology-based information retrieval systems [13] many different relationships are used, such as: *kind-of, is-a, part-of*, *has-part*. For example, in the ontology for vehicle repairs, the concept *engine* is in the relation *part-of* the concept *car*, and also the concept *wheel* is in the relation *has-part* the concept *tyre*.

Moreover, in ontology-based systems the user query is usually expanded by the related terms. So let us assume that we have the set of terms T connected by

relations in form of ontology O, and we have defined the relation r such that $r \in \{part-of, is-a, has-part, kind-of\}$ and $r \subset 2^T$. We can consider the following relations between elements from the set of request concepts and the service functionality description concepts we can distinguish the following relations:

− they are **identical** (it is called **exact match**);
− there is so called **plug-in** match between concepts, what is defined as follows:
  $(d_i, \varphi_j) \in \{part-of, is-a\}$ or there exist a sequence of concepts
  $c_1, c_2, ..., c_n$, $n = 1, 2, ...$, $(d_i, c_1), (c_1, c_2), ...(c_n, \varphi_j) \in \{part-of, is-a\}$;
− there is so called **subsumption** match between concepts, what is defined in the following way: $(d_i, \varphi_j) \in \{kind-of, has-part\}$ or there exist a sequence of concepts $c_1, c_2, ..., c_n$, $n = 1, 2, ...$ such that:
  $$(d_i, c_1), (c_1, c_2), ...(c_n, \varphi_j) \in \{kind-of, has-part\};$$
− there is so called **intersection** match, which holds if non of the above match occurs and there exist a sequence of concepts $c_1, c_2, ..., c_n$, $n = 1, 2, ...$ such that:
  $$(d_i, c_1), (c_1, c_2), ...(c_n, \varphi_j) \in \{part-of, is-a, kind-of, has-part\}$$
− if none of the above holds then there is **no match**.

Depending on the relation we expand the request set, the service description set or both in the following way:

− in the case of plug-in match we extend the request set with $c_1, c_2, ..., c_n$ and φ_j;
− in the case of subsumption match we extend the service description set with $c_1, c_2, ..., c_n$ and d_j;
− in the case of intersection match we extend the request or service description set depending on the type of relation, we extend the former with c_k, if there relation *part-of* or *is-a* holds, and the later if relation *has-part* or *kind-of* holds.

13.2.4 Performance Index for Matchmaking

The algorithm (13.2.6) could have different forms depending on various approaches:

• Information filtering [7, 12];
• Knowledge-based approach (e.g. [14]);
• Artificial Intelligence methods (e.g. [11]);

However, in the literature [2, 14] algorithm (2.6), which is used to find the best possible matching of user's request **d** with accessible scenarios collected in *CS* (13.2.1), is usually based on ideas presented in previous section.

Each type of matchmaking gives a set of services (in particular – singleton) which fulfil exact or partially user request. Up to the type of service-oriented system and situation, user could obtain different answers. For instance, *gold client* can get not only services given by exact match but also services with some other functionalities (subsume match). On the other hand, to a *silver client* only services given by plug-in or intersection match could be presented.

Nevertheless, types of matchmaking does not give final solution. After finding candidate services due to exact, plug-in, subsume, and intersection matching, set U is determined according to proposed quality criterion. In the paper [4] following criterion was proposed:

$$Q(u) = q(u, \mathbf{d}) = \max\left\{\left(\frac{N_u}{N_d} - \frac{N_a}{N_r}\right); 0\right\}, \qquad (13.2.7)$$

where N_u is the number of functionalities matched with user's demand, N_d is the number of all functionalities demanded by user, N_a is the number of additional functionalities which are provided by the service but were not requested by the user, N_r is the number of all functionalities which were not demanded by the user. According to the type of match, we can also expand user request or service description and then look for the maximum values. The example is given in the following section.

It is easy to notice that this criterion is a similarity measure [4]. Besides, it represents the approach that additional functionalities punished as lack of functionalities. Not always such approach is valid but from the system point of view it is more expensive to provide more functionalities than it was demanded.

13.3 Rough Set-Based Approach

13.3.1 *Rough Set Theory – Fundamental Definitions*

One way of representing information on objects described by the same set of attributes are so called *information systems* [8, 9, 10, 11].

Definition 13.3.1
Information system is a quadruple $IS = \langle X, A, V, f \rangle$ where X is a set of objects, A is a set of attributes, $V = \bigcup_{a \in A} V_a$ is a set of all possible attributes' values, $f : X \times A \rightarrow V$ is an informative function.

If $f: X \times A \to \wp(V) - \{\emptyset\}$, where $\wp(V)$ is a family of values' sets, then the information system is called *multivalued information system* [8]. In our further considerations we focus our attention on such kind of information systems.

Now let us assume that for given attributes from the set $B \subset A$ there is a set of chosen values $V_B \subseteq V$. Thus, we can define following sets according to formulations presented in [8], page 106.

Definition 13.3.2

The set of objects that takes values which consist at least V_B,

$$C(V_B) = \bigcap_{b \in B} C(b, V_B),$$

where $C(b, V_B) = \{x \in X : f(x,b) \supseteq V_B\}$.

Definition 13.3.3

The set of objects that take values which are subset of V_B,

$$\underline{C}(V_B) = \bigcap_{b \in B} \underline{C}(b, V_B),$$

where $\underline{C}(b, V_B) = \{x \in X : f(x,b) \subseteq V_B\}$. This set is called *lower approximation*.

Definition 13.3.4

The set of objects that take values which intersect with V_B,

$$\overline{C}(V_B) = \bigcap_{b \in B} \overline{C}(b, V_B),$$

where $\overline{C}(b, V_B) = \{x \in X : f(x,b) \cap V_B \neq \emptyset\}$. This set is called *upper approximation*.

Due to introduced definitions we can interpret the service-oriented system as the multi-valued information system in which objects are services ($X = CS$), there is one attribute which represents available functionalities $(A = \{\text{functionality}\})$, values of attributes describe functionalities ($V = \varphi \cup \{0\}$ where 0 means that service provides no functionality).

Now let us denote by $V_d \subseteq V$ a set of functionalities which are requested by the user. This set is obtained as a result of analysis of the vector **d**. We are interested in finding such services which are the best for the user. However, belonging

to such set could be ambiguous therefore we defined lower and upper approxima-
tion of which are crucial concepts in rough set-based reasoning [8, 9, 10, 11]

In the next section matchmaking methods are defined in terms of rough set
theory.

13.3.2 Rough Set-Based Approach

In [14] rough set theory was used in service matching and ranking. Nevertheless,
this approach was used to reduce irrelevant properties and matching which had a
different meaning as in Section 13.2.2. There were only some general information
on service discovery problem presented.

Nevertheless, having definitions 13.3.2, 13.3.3 and 13.3.4 following expres-
sions of matchmaking types can be introduced. Thus,

Exact match: $P = \underline{C}(V_d) \cap C(V_d)$.

Plug-in: $P = \underline{C}(V_d) - C(V_d)$.

Subsume: $P = \overline{C}(V_d) \cap \left(C(V_d) - \underline{C}(V_d)\right)$.

Intersection: $P = \overline{C}(V_d) - \left(C(V_d) \cup \underline{C}(V_d)\right)$.

No match: $CS - P = X - \overline{C}(V_d)$.

It is easy to note that above propositions fulfil definitions of matchmaking pre-
sented in [2] but in terms of set theory. Explanations and simple proofs could be
found in [4].

13.4 Application

13.4.1 Ontology of Vehicle Services

An analysis of the vehicle services offered on the market has shown the complex-
ity of this domain. Vehicle services are offered by service providers, which attrib-
utes have an impact on a quality of services. Services themselves have their own
nature and can be classified independently of their relation to other entities. In the
considered domain we identify three main concepts: 'serviced_object', 'service',
and 'service_provider' (see Fig. 13.4.1).

A serviced object is understood as a vehicle as a whole or any component of it.
In the elaborating of the vehicle services ontology one could need to acquire de-
tailed knowledge about the construction of vehicles. The construction of vehicles
should then have its own ontology. The work [1] is devoted to this issue and the
authors stress the real complexity of the vehicle structure as an ontology domain
(they say about nearly hundreds of thousands names of vehicle parts). It seems
that for the service finding purposes the ontology of vehicle construction could be
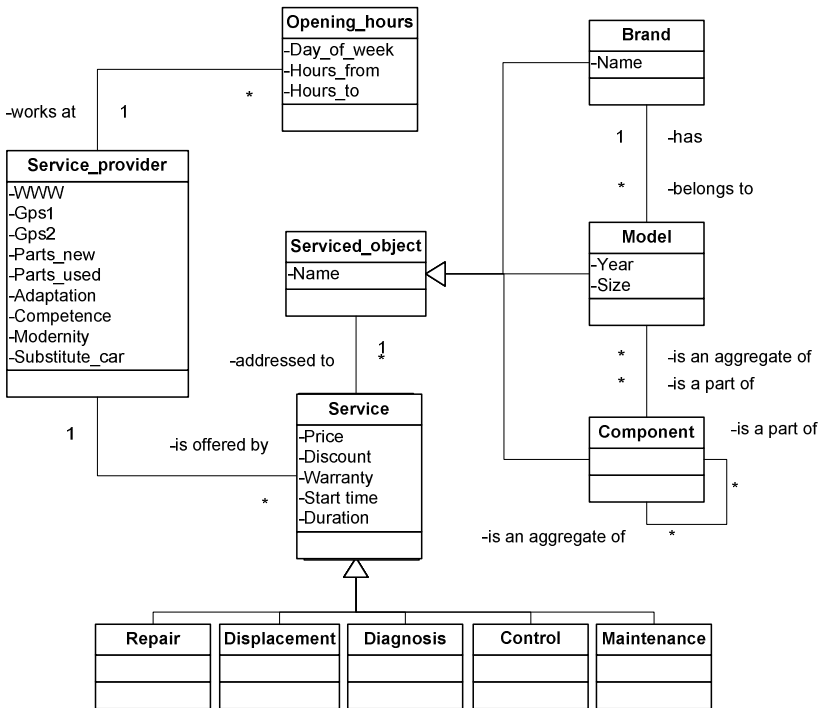
Fig. 13.4.1 A fragment of the vehicle service ontology expressed in the UML notation

simplified by introducing the most frequently used terms like vehicle systems and vehicle components. On Fig 13.4.1 we show that a vehicle (a model of it) can be composed of components which in turn can have their own components and so on. Many offers are addressed to vehicles by pointing out countries/continents of their origin, a brand, models and so on. Therefore they seem to be important concepts connected with the concept of serviced object. All of them have a concrete extensions (sets of individuals) in the domain considered. In our ontology we could use the concept called vehicle, but it seems that the pair of concepts: brand and model are more appropriate, because they represent a class of vehicles of the same structure and characteristics. But one can also use the concept vehicle when reference to the concrete individual vehicle is needed.

We divide the vehicle services into subclasses in accordance with the classification of service activities, classification of vehicles and their internal structure. We distinguish repair, replacement, diagnosis, control and maintenance as subclasses of the *Service* concept. The repair is understood as the activity aimed to restore the state of the serviced object in which it works properly. The replacement consist in the dismantling of the part and installment in place of it another one. The diagnosis should answer the question, where and what defect is the reason of improper working of the object? The control has the meaning of the activity targeted to the answer the question: Does the object work properly or is there any threat of

improper functioning an object? And the maintenance is understood as the activity of setting the properly working object to the better state.

In the vehicle service domain we have services offered by service providers that have their own characteristics, like location (GPS), offered parts (new, used or both), competence, modernity of equipment etc.

A vehicle service has its own characteristics, but also those adopted from concepts that service relates to, for example, location, possible time of the service provision, price, level of competence, modernity - from a service provider, class of a serviced object - from a serviced object.

Service classification is also strongly connected with the classification of vehicle components as well as of a vehicle classification itself. Classifying services on component types we obtain very well known sub-concepts like varnish, retread, tinsmith, mechanical, electrical electronic services. A review of services offered by their providers discloses also another way of services classification resulting from splitting the vehicles class into cars, trucks, buses, coaches, etc.

13.4.2 Illustrative Example

As an illustration of presented rough set-based approach for service discovery let us consider following example:

1. $A = \{functionality\}$;
2. $V = \{none,\ wheel\ changing,\ wheel\ balancing,\ painting,\ tinsmithing,\ retreading\}$;
3. $CS = \{cs_1, cs_2, cs_3, cs_5, cs_6, cs_7\}$ – all services are taken from the web page with companies browser: www.34y.pl;
4. Quality criterion as (13.2.7);
5. The information system is as follows:

Table 1 Service-oriented system as a multivalued information system

Service	functionality
cs_1 - EUROTRANSPORT	wheel balancing
cs_2 – ARDA	wheel changing, wheel balancing
cs_3 – ZWOLAK	wheel changing, retreading
cs_4 – DOBICKI MOTORS	painting, tinsmithing
cs_5 – AUTO-KOMPLEKS	wheel changing, wheel balancing, painting, tinsmithing, retreading
cs_6 – CAR AUDIO	none
cs_7 – CROSS sp.z o.o.	wheel changing, wheel balancing, painting

Now let us consider following request from an user: *I am interested in changing and balancing tyre.* Such request, using WordNet (semantics similiarity between *tyre* and *wheel* equals 0.12), is translated into following form: $V_d = \{$ *wheel changing, wheel balancing* $\}$.

Then we have following sets:

- $\underline{C}(V_d) = \{\text{EUROTRANSPORT, ARDA}\}$,
- $\overline{C}(V_d) = \{\text{EUROTRANSPORT, ARDA, ZWOLAK, AUTO-KOMPLEKS, CROSS}\}$,
- $C(V_d) = \{\text{ARDA, AUTO-KOMPLEKS, CROSS}\}$.

Thus we can advice services due to definitions introduced in Section 13.3.2:

- **Exact match**: $\underline{C}(V_d) \cap C(V_d) = \{\text{ARDA}\}$,
$$Q(\text{ARDA}) = \max\{1-0;\ 0\} = 1.$$
- **Plug-in match**: $\underline{C}(V_d) - C(V_d) = \{\text{EUROTRANSPORT}\}$,
$$Q(\text{EUROTRANSPORT}) = \max\{0.5-0;\ 0\} = 0.5.$$
- **Subsume match**: $\overline{C}(V_d) \cap \left(C(V_d) - \underline{C}(V_d)\right) = \{\text{AUTO-KOMPLEKS, CROSS}\}$,
$$Q(\text{AUTO-KOMPLEKS}) = \max\{1-1;\ 0\} = 0,$$
$$Q(\text{CROSS}) = \max\left\{1-\frac{1}{3};\ 0\right\} = \frac{2}{3}.$$
- **Intersection match**:
$$\overline{C}(V_d) - \left(C(V_d) \cup \underline{C}(V_d)\right) = \{\text{ZWOLAK}\},$$
$$Q(\text{ZWOLAK}) = \max\left\{\frac{1}{2} - \frac{1}{3};\ 0\right\} = \frac{1}{6}.$$

Hence, $U = \{\text{ARDA}\}$ and result is the service provided by company ARDA.

13.5 Discussion and Future Works

In this chapter a new approach for service discovery was presented. The novelty in proposed methodology is application of rough set-based algorithm to service matchmaking which are described by ontologies. Considered method was compared with existed approaches which are rested on descriptive logic.

To illustrate proposed approach an example of service discovery in vehicle services web search engine was discussed.

Further works are included extension of proposed approach to take into account not only functional but non-functional features as well. The second main concern

is utilization of services describe by formal languages such as OWL-S or WSMO in rough set-based algorithm to service discovery.

References

1. Angele, J., Erdmann, M., Wenke, D.: Ontology-based knowledge management in automotive engineering scenarios. In: Hepp, M., Leenheer, P., Moor, A., Sure, Y. (eds.) Ontology Management. Semantic Web, Semantic Web Services, and Business Applications. Series: Semantic Web and Beyond, vol. 7 (2008)
2. Bianchini, D., De Aantonellis, V., Melchiori, M.: Flexible Semantic-Based Service Matchmaking and Discovery. World Wide Web 11, 227–251 (2008)
3. Brzostowski, K., Drapała, J., Świątek, P., Tomczak, J.M.: Tools for automatic processing of users requirements in SOA architecture. In: Grzech, A. (ed.) Information systems architecture and technology: Service oriented distributed systems. Concepts and infrastructure, Oficyna PWr, Wrocław (2009)
4. Brzostowski, K., Tomczak, J.M.: Rough set-based approach for service discovery in service-oriented systems. In: Int. Conf. of Systems (2010) (to be published)
5. Fensel, D., et al.: Enabling Semantic Web Services. In: The Web Service Modeling Ontology. Springer, Berlin (2007)
6. Grzech, A., Rygielski, P., Świątek, P.: QoS-aware infrastructure resources allocation in systems based on service-oriented architecture paradigm. In: Proc. of Performance Modelling and Evaluation of Heterogeneous Networks, HET-NETs2010, Zakopane, Poland (2010)
7. Montaner, M., Lopez, B., de la Rosa, J.P.: A Taxonomy of Recommender Agents on the Internet. Artificial Intelligence Review 19, 285–330 (2003)
8. Pawlak, Z.: Information Systems. Theoretical foundation. WNT, Warszawa (1983) (in polish)
9. Pawlak, Z.: Rough Sets: Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, Dordrecht (1993)
10. Pawlak, Z.: Rough set theory and its applications. Journal of Telecommunications and Information Technology 3, 7–10 (2003)
11. Rutkowski, L.: Methods and Techniques in Artificial Intelligence. PWN, Warszawa (in polish)
12. Sobecki, J., Tomczak, J.M.: Student courses recommendation using Ant Colony Optimization. In: Nguyen, N.T. (ed.) ACIIDS 2010, Part I. LNCS (LNAI), vol. 5990. Springer, Heidelberg (2009)
13. Wang, H., Liu, S., Chia, L.T.: Does Ontology Help in Image Retrieval? — A Comparison between Keyword, Text Ontology and Multi-Modality Ontology Approaches. In: Proc. of MM 2006, Santa Barbara, California, USA (2006)
14. Yu, B., Li, M.: RSSM: A Rough sets based service matchmaking algorithm. In: Proc. of UK e-Science, AHM 2006, Nottingham, England (2006)

Chapter 14
Towards Self-defending Mechanisms Using Data Mining in the EFIPSANS Framework

Krzysztof Cabaj, Krzysztof Szczypiorski, and Sheila Becker

Abstract. In currently used networks there are no self-protection or autonomous defending mechanisms. This situation leads to the spread of self-propagating malware, which causes even more dangerous, and significant threats i.e. Botnets. In the EFIPSANS project a new architecture that includes self-* functionalities is introduced. Self-defending functionality, using data mining approach detects and reacts to some of network threats.

14.1 Introduction

In this paper, we present self-defending mechanisms for autonomic networks. This work is in relation to the European project EFIPSANS – Exposing the Features in IP version Six protocols that can be exploited/extended for the purposes of designing/building Autonomic Networks and Services. The objective of EFIPSANS project is to optimize communication in several dimensions, e.g. in wired or wireless networks, and to provide services of high quality, and functionality, as well as ubiquitous access possibilities. Within this project, a Generic Autonomic Network Architecture (GANA) is proposed for modeling and specifying autonomic behavior. In this context, self-description, self-advertisement, self-healing, self-configuration are seen as autonomic behavior. GANA is composed of different entities: Decision-Making-Elements (DME/DE) and Managed Entities (ME). Managed Entities are controlled and managed by Decision Making Elements, triggering actions as a result of reception of information from their information suppliers e.g. their associated Managed Entities. These actions represent autonomic behavior.

Krzysztof Cabaj and Krzysztof Szczypiorski
Warsaw University of Technology, Nowowiejska 15/19 00-665 Warsaw, Poland
e-mail: {kcabaj,kszczypi}@elka.pw.edu.pl

Sheila Becker
University of Luxembourg, 6 rue R. Coudenhove-Kalergie, L-1359 Luxembourg
e-mail: sheila.becker@uni.lu

Reliability and availability of this architecture are significant for its operating ability. One can never be certain that no malicious users/nodes are in the network and nowadays networks are lacking of self-protection and autonomic defending mechanisms. Therefore, it is important to guarantee reliability and availability by defining self-defending functionalities. These self-defending functionalities consist of detecting abnormal or malicious activities and of defining associated defending mechanisms. These functionalities are integrated in a Decision-Making-Element, for incorporating self-defending properties in GANA. This way, a node can protect and defend itself autonomously.

This paper is structured as follows: in Section 14.2 the proposed self-defending functionalities are investigated within a scenario. Section 14.3 describes current threats of an autonomic network that can be detected. The Detection mechanisms are explained in Section 14.4. Experimental results are shown in Section 14.5 and finally, we conclude our work and present possible future work in Section 14.6.

14.2 Self-defending Functionality

In currently used networks there are no self-protection and autonomous defending mechanisms. This situation leads to the spread of self-propagating malware, which causes even more dangerous, and significant threats i.e. Botnets [1]. Botnet is a group of infected computers (often called Zombies) that are controlled either by one person or an organization. These machines can do any malicious work ordered by controller, called botmaster. SPAM sending, hosting phishing sites and performing DDoS attack are examples of malicious activities.

In the future, self defending networks should detect such activities and protect other network users. The GANA architecture introduces autonomous behavior; one part of this behavior is called self-defending, it is responsible for security functions that become more and more important nowadays. Self-defending functionality is implemented in node-level security-management Decision-Making-Element (NODE_LEVEL_SEC_MNGT_DE). The threat detection and proper reaction to it could be performed in cooperation with other DEs, described in details below.

The proposed self-defending functionality can be divided into two phases. In the first phase, malicious or suspected activity is detected. In the second phase, according to a predefined policy, the appropriate reaction is performed.

In the first phase, suspicious activity must be detected. In this case, monitoring services provided by other GANA DEs are used, especially those provided by Monitoring Decision-Making-Element (MON_DE). For all interesting data NODE_LEVEL_SEC_MNGT_DE communicates with MON_DE. This communication is bidirectional: NODE_LEVEL_SEC_MNGT_DE sends requests for needed data and MON_DE provides acquired data. To ensure high performance some aggregations are performed by MON_DE. Acquired data is analyzed by NODE_LEVEL_SEC_MNGT_DE using data mining techniques. The class of attacks that can be detected is described in detail in the following section.

14.3   Current Threats to Be Detected

Some network activities cause generation of additional amount of traffic. Consequently, these threats may be detected by monitoring network traffic. The description of this kind of network threats and the characteristic of the traffic that is produced by them is outlined in the following. Of course not all of them are in fact attacks, for example, scanning or sudden rise of traffic destined to a particular machine. However, detection of this activity may be a first sign that something strange happens. If some of these activities are usual for certain machines, a properly configured policy can be used for disabling false positives.

The second phase is responsible for reaction to the previously discovered threats. The appropriate reaction steps are executed using the detected pattern, which describes a threat. Logs that inform about an observed event are generated and sent to the operator. It is possible to perform additional steps that can fulfill operators' needs. All parameters concerning this activity are placed in a policy. The policy connects protected machines' address with actions that should be performed in case determined threat is detected. Details of the policy will be defined in the course of the research. For example, according to the policy, a detected malicious machine can be removed from the network, malicious traffic can be completely removed or only maximally slowed down. In the policy, all parameters can be tuned, for example, a conducted reaction can depend on malicious machine IP address or type of detected threat. This flexibility of the policy tunings provides the ability to implement protection reaction according to all operators' needs. For example, an operator can immediately shut down an infected machine of a home client. On the other hand, traffic of the business client's critical server can be only slowed down. In this phase, services provided by Quality of Service Management DE (QoS_M_DE), Routing Management DE (RM_DE) and Forwarding Management DE(FWD_M_DE) are used.

14.4   Method of Detection

In the self-defending functionality, introduced in section 14.2, the detection of malicious or suspicious activity is performed using data mining techniques. There are many techniques that could be used for this purpose, e.g. neural networks, statistics or probabilistic methods. However, data mining approach was chosen due to simplicity and intelligibility of extracted knowledge. Detected data mining patterns are evidence of some malicious or suspected activity. Moreover to this information, they also carry description of detected activity. This description is in form, which can be easily understood by humans and easily converted by automatic systems for appropriate reaction.

Described in Section 14.3, threats generate significant amounts of packets that have similar characteristic. For example, scanning for vulnerable machine activity produces vast amounts of packets that have the same source IP address, protocol and port. In data mining techniques this kind of patterns are known as frequent patterns. In literature there are few such patterns, for example, frequent sets, sequential

patterns [2] or episodes [3]. During the research, a decision on which patterns can be used in the proposed solution was made. The main impact on the decision is associated with the ability to easily use extracted knowledge in other collaborating systems. Currently, there is a lack of IDS/IPS or firewall systems that can utilize complex relations between some networks related events. Due to this fact, sequences and episodes are omitted and pattern called frequent set is chosen for further analysis. Moreover, this decision is justified by the fact that there are many algorithms that can be used for detecting this pattern. It should be emphasized that some of these algorithms can detect this pattern in incremental fashion.

As a result of this finding, in the solution further described in this paragraph, we are interested only in the pattern called a *frequent set*. The *frequent set* idea was presented by Agrawal in 1993 [4]. Data that will be analyzed is treated as a collection of item sets. A subset of items that frequently appears in analyzed data is called the *frequent set*. The parameter, called *minimal support,* is used for deciding what is frequent. A support parameter is calculated as a number of item sets that contain a given subset of items. Minimal support parameter decides which item sets are detected.

Frequent set pattern can be detected in data in many ways. The most known algorithm called A-Priori is described by Agrawal et al. [5]. The other known solutions that are using tree structures are FP-Trees [6] and CATS trees [7]. The last two mentioned algorithms are interesting, because of the ability of incremental mining. This advantage causes that patterns can start to be discovered when data are acquired, and result is immediately known when last item set is delivered. In this place one remark must be made. All presented algorithms discover frequent sets in whole data sets. Due to fact that in our solution the detection of suspicious events should be almost in real time, acquired data is partitioned for periods of time, and analyzed in each partition independently using presented methods.

Choosing from a variety of available data mining techniques is not the most important part of setting up the functionality; choosing parameters that will be used during data mining is more crucial. During the research concerning currently observed threats, some of them are investigated, and a short description of this work is presented in Section 14.3 of the document. Based on this research, interesting features are proposed. For further analysis only the following packet fields are used:

- protocol, this attribute can assume one value from this set: ICMP, UDP, TCP and OTHER,
- source and destination IP address,
- for TCP and UDP packets, source and destination port,
- for TCP protocol, interesting flags: SYN, RST, FIN,
- for ICMP packets, ICMP code and type,
- for UDP and OTHER protocol, packet/datagram size.

A simple case study that describes the detection of suspicious activity using frequent sets is presented below. For first experiments simple, not incremental, algorithm extracting frequent sets is used.

First, data that has to be analyzed must be converted to a form appropriate for the above described algorithm. In our case, all interesting packets are converted to item sets. Each interesting packet's features previously described are converted to items. For example, TCP packet sent from IP address IP_A and source port 1234 to a machine with IP address IP_B and a destination port 80, with set SYN flag. Such packet is described as the following:

```
<TCP, SRC_IP_A, DST_IP_A, SRC_PORT_1234, DST_PORT_80, SYN>.
```

Above described item set is presented in human readable textual form. Obviously, during pattern extraction all items in item sets are described as integer numbers. Numbers are assigned in such way, that information, what feature is described by them, can be easily reviled. This feature is utilized when extracted frequent sets are analyzed. For example, the question, if in extracted frequent set information about source IP or destination port is present, can easily be answered. This information can be used for determine what threat is described by this frequent set.

In case, one machine starts scanning for vulnerable web servers during an analyzed period, many packets are send to the same port but various destination IP addresses. In this situation, a frequent set presented below will be discovered:

```
<TCP, SRC_IP_A, DST_PORT_80, SYN>
```

A detected pattern is easily understood and almost self describing; in analyzed period there are many packets using TCP protocol with SYN flag set sent from machine using IP address IP_A to port 80. Because in extracted pattern there is a lack of information about source port and destination address, it can be assumed that those features of packets are changing.

It seems that this approach is similar to a simple counter that counts packets, for example, to given ports. However, the advantage of this method is that any combination of analyzed features that appears in monitored packets can be detected. For example, using this method a DDoS attack that uses UDP packets sourced from port 3333 that has 533 bytes long can be detected. Using counters each combination of protocol, packet port and size should have its own counter.

For assurance of the high performance, in presented scenario only a minimal amount of data is analyzed. In normal network activity only high level information about protocols is analyzed. Where something suspicious is detected, additional information about IP address is requested from MON_DE. After detection of suspicious IP address, additional detailed information is requested only for them.

14.5 Experiments and Results

In this section, we present preliminary experiments that show that the proposed method can be utilized for protecting computer networks. During experiments, we injected some suspicious network activities into real network. The nature of malicious traffic is described below in more details. Captured network traffic was analyzed using proposed method. At the beginning, captured and interesting packets are preprocessed and in effect represented as item sets. Each item describes one

interesting feature, for example, protocol, used flag, packet size or destination address. Secondly, preprocessed data is divided into partitions each consisting of data captured during five minutes. In each partition, in data that has the form of item sets, frequent sets are discovered. During conducted experiments, for frequent sets discovery apriori algorithm was used. In the final solution incremental algorithm for discovery frequent sets will be used, probably one, which utilizes CATS tree.

We analyzed two different data sets. First data set contains TCP sessions performed from a laptop used by one of the authors. Second data set contains ICMP traffic captured from the laboratory network used by students for their master thesis experiments. Both performed experiments are described independently.

TCP traffic data was captured over the period of almost six and a half month (from 9th July 2009 till 19th January 2010). In this experiment, for each TCP connection source and destination IP address and port were recorded. As source of malicious TCP traffic nmap [8] network scanner is used. During normal user activity nmap scanner with various options was executed. As a result in network traffic many packets with similar characteristics appear. Because during experiments only one machine was scanned, pairs of source and destination address are the same, and this fact with additional details was detected. This can be treat as example of various malicious activity, where packets with the same characteristic appear, for example, DDoS or SPAM. In first attack simulation, nmap was used for performing scanning of all TCP ports. Two other simulations use only scanning of interesting ports with two scanning timings (T2 and T5). Various scanning timing influences number of packets that are send by nmap program. During analysis, periods of 5 minutes are used. For efficiency purpose only periods in which more than 500 events (TCP connections) were observed are analyzed using data mining. In all captured data only 6, five minute periods contains more than 500 events, respectively: 1066, 18474, 47686, 657, 1999, 739, 738. All periods except this containing 657 events are connected with scanning activity. During data mining analysis of scanning activities, frequent sets that have minimal support more than 500 are extracted (support in brackets):

```
tcp src IP 194.29.168.76 dest IP 194.29.168.6 (739)

src port 34209 tcp src IP 194.29.168.76 dest IP 194.29.168.6 (997)

src port 49040 src IP 194.29.168.76 dest IP 194.29.168.84 tcp (47451)
```

In last two frequent sets pattern associated with scanning is clearly viewed, because in extracted frequent set only source port is detected. This fact can be explained by nmap operation, which uses this same source port for performing scanning of multiple destination ports. In first case, when slowest scanning is executed if we decrease minimal support similar patterns can be discovered.

```
 src port 37658 tcp src IP 194.29.168.76 dest IP 194.29.168.6 (336)

 src port 37659 tcp src IP 194.29.168.76 dest IP 194.29.168.6 (336)
```

If we analyze this one period not associated with scanning activity, using frequent sets and minimal support set to 500 such items set can be discovered:

```
tcp dest port 80 src IP 192.168.1.10
```

This pattern can be associated with scanning for machines which opens port number 80. But if we decrease minimal support, such pattern can be discovered:

```
dest IP 80.67.20.51 tcp dest port 80 src IP 192.168.1.10 (142)
dest IP 216.137.61.254 tcp dest port 80 src IP 192.168.1.10(481)
```

Those patterns suggest that this activity is associated with two web servers that transfer each web resource in one connection. This pattern can be distinguished from the previous observation where the source port remains the same. (an in effect are not placed in extracted item sets) In contrast to the previous situation, the destination port remains fixed.

What is very important, obtained patterns can be easily converted into the signatures. Additionally, patterns can give information about what addresses are involved in suspicious activity and describe which address belongs to the attacker and which to the victim.

ICMP traffic was captured in one laboratory of Warsaw University of Technology. Analyzed data contains traffic recorder during 22 days from 11:48 at 6 January to 11:57 at 27 January 2010. For each ICMP packet source and destination IP, icmp type and code and datagram size are recorded. Similar to the previously described TCP experiments, activity is treated as malicious when in 5 minutes period more than 500 packets containing this same features appear. Additionally to captured data containing normal user activity, two malicious traces are recorded. Both contain traffic generated by pingtunnel, steganographic software, which hides user traffic using ICMP ping request and ping reply packets. In first malicious sample, ssh session is a tunnel via this software. Second malicious trace contains packet used for downloading of file using scp traffic hided in such manner.

First step of the experiment catches all those periods of time, where ICMP packet number exceeds 500. In analyzed data are five such periods present, which contain respectively 8744, 15101, 11090, 821 and 1229 ICMP packets. All other periods do not exceed 301 ICMP packets within five minutes period. The two last numbers concern the situation when intentionally malicious traffic using pingtunnel is executed. Extraction of frequent sets gives this same patterns:

```
icmp src IP 194.29.XX.XX dest IP 194.29.YY.YY type=0 code=0
```

respectively with support 799 and 971. For first sign, this activity can be treated as ICMP echo reply. What is suspected, and can give sign that this is malicious activity there is no echo request packets (for which echo reply is send). Additionally, because in extracted patterns size of packet is missing, this suggests that size changes. In normal ping usage all packets have the same size. Both of these information can be a sign, that detected traffic is malicious.

First three periods, with much higher numbers, was at first secret, and for first view suggest some DoS attack. What is interesting is that these periods are consequent, and what is worrying, is that this traffic was generated from a machine at the University, used for IP telephony experiments. Extracted pattern look like:

```
icmp type=3 code=3 size=32 srcIP 194.29.XX.YY destIP 85.222.XX.YY
```

and have support 15101. Analysis of extracted pattern suggests that this traffic is only a response to other traffic directed to this machine. Analysis of the volume of the traffic and main purpose of this server gives answer, what happens. In this machine a SIP server is running. Normal VoIP connection use small packets, carrying 20 ms of voice samples. This gives 3000 packets in each minute, and in effect 15000 in 5 minutes perthreiod. Due to some error in initial process of negotiating call parameters client send RTP traffic to a closed port, in effect generating such a number of icmp port unreachable packets.

14.6 Conclusions and Future Work

In this chapter, we presented self-defending mechanisms for autonomic networks, specifically for the Generic Autonomic Network Architecture (GANA) in relation to the EFISPANS project. We faced different threats in autonomic networks and we defined self-defending functionalities to face these threats. We propose the integration of these functionalities in Decision Making Elements, so that these functionalities are build as part of the network architecture. Furthermore, this paper explains what kind of detection methods we use and we show a proof of concept of this work with our experimental part.

Conducted experiments show that the data mining approach may be useful for detecting malicious activity by observing network traffic. The results presented in this paper require in some steps user activity. Future work should introduce automatic process of detecting and deciding which actions are really malicious. Additionally, future work should be directed to reaction mechanisms, which might be based on detection methods presented in this paper.

References

1. http://www.cisco.com/en/US/solutions/collateral/ns340/
   ns394/ns171/ns441/networking_solutions_whitepaper0900aecd80
   72a537.html
2. Agrawal, R., Srikant, R.: Mining Sequential Patterns. In: Proceedings of 1995 Int. Conf. Data Engineering (ICDE 1995), Taipei, Taiwan, pp. 3–14 (1995)
3. Mannila, H., Toivonen, H., Verkamo, A.I.: Discovering Frequent Episodes in Sequence. In: Proceedings of the First International Conference on Knowledge Discovery and Data Mining, Montreal, Quebec, pp. 144–155 (1995)
4. Agrawal, R., Imielinski, T., Swami, A.: Mining Association Rules Between Sets of Items in Large Databases. In: Proceedings of ACM SIGMOD Int. Conf. Management of Data (1993)

5. Agrawal, R., Srikant, R.: Fast algorithm for mining association rules. In: Bocca, J.B., Jarke, M., Zaniolo, C. (eds.) Proceedings 20th International Conference on Very Large Databases, pp. 487–499 (1994)
6. Han, J., Pei, J., Yin, Y.: Mining Frequent Patterns without Candidate Generation. In: Proceedings of the 2000 ACM SIGMOD international conference on Management of data, Dallas, Texas, United States (2000)
7. Cheung, W., Zaïane, O.: Incremental Mining of Frequent Patterns Without Candidate Generation or Support Constraint. In: 7th International Database Engineering and Applications Symposium (IDEAS 2003), Hong Kong, China. IEEE Computer Society, Los Alamitos (2003)
8. `http://nmap.org/book/man.html#man-description`

Part III
Information Systems Applications

Chapter 15
User Adaptivity Features of Secured Biomedical User Adaptive System

Dalibor Janckulik, Leona Motalova, Ondrej Krejcar, and Petr Czekaj

Abstract. User Adaptive Systems (UAS) are growing in past time along with expansion of intelligent ubiquitous systems to every piece of embedded device around us. These systems provide a many of interesting and useful services to support our day life needs in all areas include biomedical world of our bodies. Every such service need however to provide their results by some kind of interface. Therefore the need of a simple, intuitive and graphically attractive interface is much appreciated. Our chapter is focused to several areas of user interface design, user interface adaptivity and visualization. In all areas of developed system we implement secured ways.

15.1   Introduction

Biomedical Information Systems are widely known in medical and physics areas where some kind of biomedical data from patient is monitored through a whole measurement chain from mobile monitoring station by wireless connectivity to server and to any kind of computer interface for physicians which can access measured data. In many of commercially based products are not possible to get raw data from monitoring of patient. Such fact is however very sad for consequently processing of measured data in any way. Our developed biomedical user adaptive systems would like to solve mentioned problem and provide a way to access any kind of previously measured biomedical data or real time data monitoring by remotely connection (Fig. 15.1).

User adaptivity of the system is an important part of different platforms to work with biomedical data. GUI should be realigned according to various parameters

Dalibor Janckulik, Leona Motalova, Ondrej Krejcar, and Petr Czekaj
VSB Technical University of Ostrava, Center for Applied Cybernetics,
Department of Measurement and Control, Faculty of Electrical Engineering and Computer
Science, 17. Listopadu 15, 70833 Ostrava Poruba, Czech Republic
e-mail: `Dalibor.Janckulik@hotmail.com`, `Leona.Motalova@gmail.com`,
        `Ondrej.Krejcar@remoteworld.net`, `Petr.Czekaj@vsb.cz`
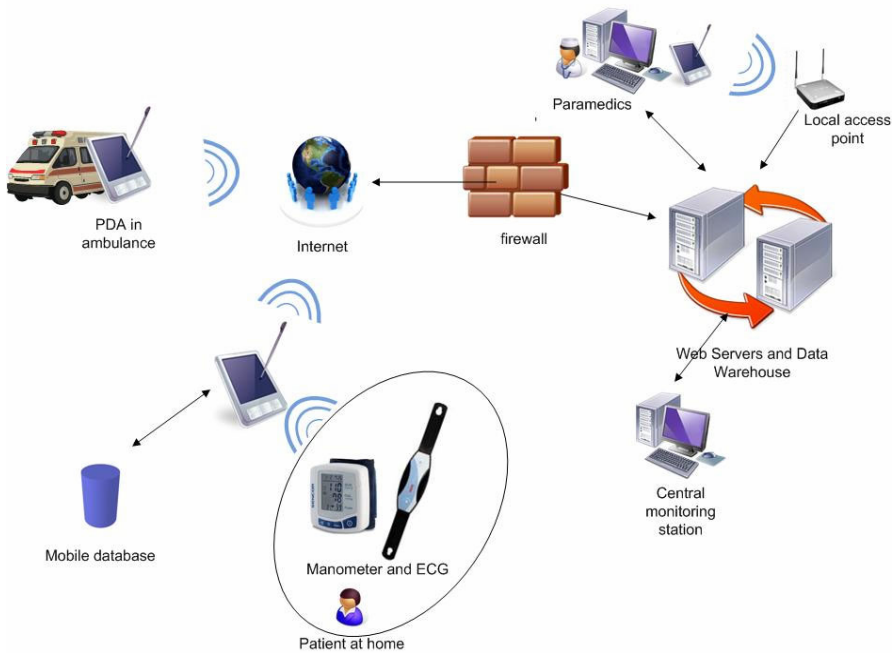
Fig. 15.1 Sample architecture of one option of system implementation

and sensed perceptions so that what at first glance most of describing the situation. Our system reacted too to stimuli from the hardware used by devices such as light sensors, accelerometers, GPS and GSM module. The main aim of the platform for patients' bio-parameters monitoring is to offer a solution providing services to help and make full health care more efficient. Physicians and other medical staff will not be forced to make difficult and manual work including unending paperwork, but they will be able to focus on the patients and their problems. All data will be accessible almost anytime anywhere through special applications designated for portable devices web browser or desktop clients and any changes will be made immediately at disposal to medical staff based on the security clearance.

For the physicians is important see the data directly and clearly on the maximum possible viewing area. This problem can be solved by dynamical programming, when we can load only important controls and functional code from database and via dynamicaly controls hiding in GUI on presentation layer. All this possibilities are described in the following sections.

From the database perspective and analysis of bio-signals, the data are stored and automatically analyzed by simply neuronal network.

15.2 The User Adaptivity of the System

The user adaptivity of the system can be divided into several areas:

- Interaction with the user is based on:
  - o used equipment
  - o application usage
  - o used environment
- Interaction with hardware devices based on the usage of applications

15.2.1   Logged User Context Adapting

The application responds as the most of applications on the resolution imaging devices - displays – on the placement of elements. For client applications, where user fills in or views his personal records, the application will behave in different ways while running on personal computer or running on mobile devices such as PDAs with QVGA, VGA or WVGA resolution. For the mobile platform many controls are not available, but we also need to adjust the controls for equipment or touch SmartPhone with a keyboard. Features in one desktop application form must be in the mobile application divided into logical groups. Elderly people may have difficulties with fine resolution and small font, the younger generation will not mind that and we can put in more information into the form.

15.2.2   UI Adapting by Interaction

A client application for an average user is responsible for, as already described above, showing data of the user. In the case of desktop, the monitor cannot be easily manipulated and rotated as we like it. In the case of the most wide-spread tablets text information can be shown in more natural form for reading (in height of the book). On the contrary, the graph is better to be displayed in width, so that the longest record gets on the screen. The same goes for using your mobile PDA. This functionality is available for devices with accelerometer, or even for devices without accelerometer, whose display can be rotated by the user's request button.

One of the novelties is a frequently used battery indicator color change. This is for the battery of mobile devices (tablets, PDA) as well as for the battery of device designed for collecting data (ECG).

Other features are associated with the popular monitoring of heart rate during exercise. The ECG, which we get we can determine the heart rate. Blue ECG device detects a pulse back. The frequencies of training rate are obtained from the users' data and each zone is indicated by color. So the user can simply by a glance get to know in which zone his heart works.

15.2.3   Smart Environment Adaptation

Here the user interface adaptation takes place according to the conditions in which it is used, at night in hospitals we do not want the monitor to light and be disruptive, but it must be legible. The solution is to switch the colors GUI so that the

contrast ratio of application isn't extremely disturbed. The influence is the back-light display device that makes it possible.

15.2.4 Hardware Adaptability

If the user views only historical records, it does not make any sense to have the Bluetooth on; data transfer is realized using WiFi or GPRS. Compression is related, or transfer of requested data. GPRS is slow, so we transfer only data that the user actually requested, for our WiFi connection bitrates primarily confined. Shutting down the unneeded hardware is today addressed for mobile devices, where device endurance is at one of the first places.

15.3 Architecture and Backgrounds of Biomedical Adaptive System

Complete proposition of solution and implementation of the patient's biotelemetry platform oriented for user adaptivity requires determination and teamwork. Every single part of the architecture has to be designed for easy application and connectivity without user extra effort, but user must be able to use given solution easily and effectively. Crucial parts of the whole architecture are network servers, database servers and client applications run-able from standard desktop operating system and client applications for Windows CE based mobile devices.

15.3.1 Server and Database Parts

Database background of solution is built on Microsoft SQL server. One server provides only data warehouse with stored procedures, which represent data interface for other application parts. There are stored all data of medical staff and patients. Data of patients include different records such as diagnosis, treatment progress or data which are results of measuring by small portable devices designated to home care. These data represent the greatest problem, because amount of these data rapidly increase with increasing amount of patients. Due to this fact database servers are very loaded. The stored procedures (programmed by Transact-SQL language) serve basic data parsing on weak mobile devices, which have too much problems with parsing and subsequently visualization of measured data. At this time our team is focused on implementation of analytical and reporting parts for more detailed analysis of measured/collected data. Business intelligence part of SQL server is a powerful tool which allows us to create reports faster than C# application on client side.

The only possibility of running our application on all platforms is an implementation of a view and controller layer as web application. We use two different technologies. ASP.NET is purely for web application (browser independent) and for web services. The second technology is Silverlight (only for Internet

Explorer and Mozilla Firefox). Silverlight application is possible to run in Out-of-Browser mode.

In order to run a web server, an operating system supporting IIS (Internet Information Services) is needed. IIS allow to users to connect to the web server by the HTTP protocol. The web service transfers data between the server and PDA/Embedded devices. Web service also read the data, sends acknowledgments, and stores the data in the database. The service is built upon ASP.NET 2.0 technology. The SOAP protocol is used for the transport of XML data.

Methods that devices communicating with the web service can use include:

- receiving measured data,
- receiving patient data,
- deleting a patient,
- patient data sending.
- RAW data parsing
- other ...

To observe measured data effectively, visualization is needed. A type of graph as used in professional solutions is an ideal solution. To achieve this in a server application, a .NET Chart Control can be used for ASP.NET 3.5. For data analysis, neural nets are a convenient solution. However, there are problems in the automatic detection of critical states. Every person has a specific ECG pattern. The Neural net has to learn to distinguish critical states of each patient separately.

15.3.2   Embedded and Desktop Part

The desktop client application is the main and the only part of the entire platform for patients' bio-parameters monitoring, which medical staff uses directly. It is obvious, that if Guardian should optimally replace classic paperwork, simplicity and trouble-free usage of client application are very important factors, which affect whether the doctors and medical staff accept this solution with enthusiasm and solution will be fully used or not. The options of desktop client application have to be easily upgraded. Therefore it is important to reliably design architecture which will allow that. Implementation of user functions is also important. Using the platform.NET in-build characteristics and open standards such as XML, XPath and other, is crucial. Because of that it is easy to configure or upgrade application. User interface is also easy to adjust to user request or clearance. Well designed architecture allows not only easier developing to software engineer, but brings also new and useful functions to the user. The design of appropriate architecture is crucial for the next development of implemented client application, which will be easily upgraded with new functions in the future without making any expensive and demanding changes in programmatic code.

- User Interface - represents the part of application, which is made by components of user interface.

- Command Manager - associates and administrates all existing objects "Command", which execute operations called mainly by user interface.
- UI Factories - dynamic assembling of some parts of user interface during the run of application or immediately after lunch.
- XML Navigators – reading of components of XML documents, which describe user interface.
- Configuration – reading and editing of XML files, which are designated for primary application configuration.
- Web Service Proxy - creates an impression of existence of local web services copies.
- XML - files, which are indispensable for running of entire application. These files control exact syntax, which enables their easy programmatic analysis.

XML represents great role in suggested solution. Options of this technology are used by dynamic assembling key components of graphic user interface, which enables its changing in dependence on roles or clearance of users. It is also used for easy application configuration.

15.3.3   Mobile Parts

The main part of the system is an Embedded or PDA device. The difference in applications for measurement units is the possibility to visualize the measured data in both Real-time Graph and Historical Trend Graph, which can be omitted on an embedded device. PDA is a much better choice for Personal Healthcare, where the patient is already healthy and needs to review his condition. Embedded devices can be designed for one user, with the option to use an external display used for settings or with the possibility of usage in extreme conditions.

The user adaptivity on mobile devices is control reorganization based on screen rotation provided by operating system. For automatical screen orientation change the mobile device must have accelerometer or G-sensor (HTC feature) implemented. The next step is color theme of application change. This feature is provided by measured level of actual lighting from built-in light-sensor. Devices which do not have this hardware parts can be set to defaults. Application reads default values from registry or from internal mobile database. Dynamically generated design for standard WindowsForm application is developed by dynamical programming technique, where small parts of application or controls source codes are inserted in database. When the application is running, the basic parts are loaded from database. Next choice for dynamically loaded design is XML generated design, described above in (section 15.3.2.).

15.4   Safety Features in Our System

Nowadays requirements on the safety on the internet, protection of the personal and private data from biomedical applications gain the importance. With new

possibilities of extending biomedical embedded systems of new interfaces which enable communication over local area networks or the internet, secured data transfer and communications regard even these, in earlier times this absolutely closed and from outside almost unattackable systems. Because of that, it is important to secure these systems properly and protect them from outside attacks and misuse of delicate data. A good way to secure biomedical embedded systems is use of PKI technologies. PKI technologies enable using different means to secure delicate data such as cryptography, sender and recipient authentication, digital signature auth. etc.

Actual secured data transfer between embedded devices is based on client-server model where server authenticates the client and vice versa, client checks the authenticity of the server which the client connecting to. Simplified communication is displayed in the picture.
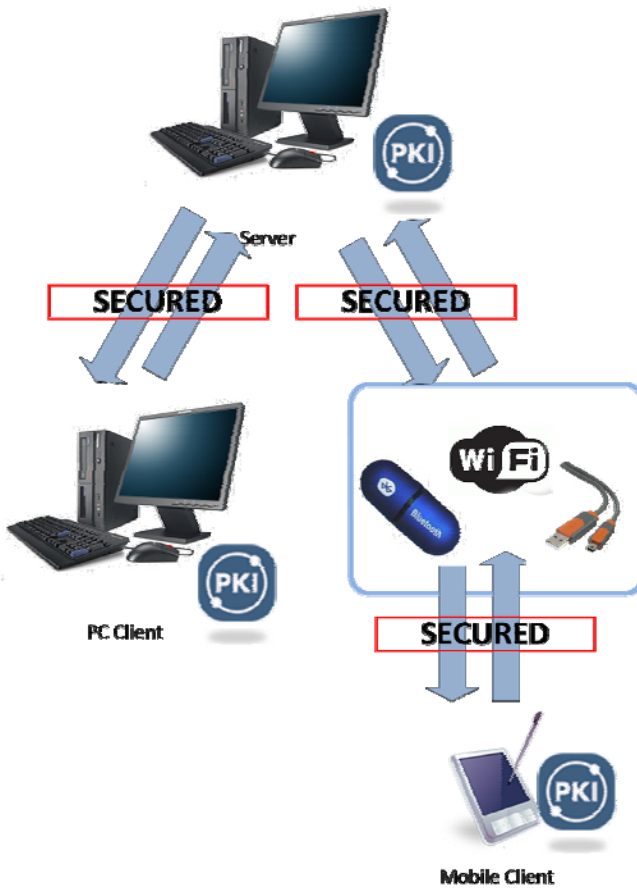
Fig. 15.2 Secured communication

One device runs an IIS web server or in general any secured web server (Apache, Tomcat). It provides among common services its own certificate authority. It can issue and manage, verify or even revoke its own digital certificates used during the secured SSL communication. It protects the access to the database and other network resources.

During each client's attempt of communication, server verifies the client's certificate thus it authenticates the client. Server can be any biomedical embedded device which is capable of running secured web server.

Server itself then presents to each client its own server certificate. This certificate is issued based on server IP address or server name. Server can sign the certificate itself (self signed certificate) or it can be signed by a commercial certification authority.

Server certificate is verified by a client before each communication. Client can be any other device such as microcontroller, PDA, cell phone, computer, in general any device which can communicate via LAN or Wi-Fi.

Message integrity is ensured using existing methods of symmetric and asymmetric cryptography.

15.5 User Interface Designing Adaptation

Proposal for this UI cyclically through several steps:

- Needs Analysis
- Design of UI
- Implementation of UI design and dynamics
- Testing

This is the so-called spiral development model, in which each spiral passage assess emerging requirements identified during testing of the current UI. At the very beginning of the development of UI, we have addressed a number of potential users. Questionnaires were sent to those under which we created the initial list of core requirements for the application and its UI. After the design and implementation of the application was sent back to users (in this case and inexperienced - viz. Testing UI). For testing, we obtained a list of comments on the application of its control and in some cases, suggestions of possible vylepšení.V this case, the tester user because our priority was to create such a UI, which is adapted to the user so that he worked well with him. Currently the application is again in the testing phase (4th course of the cycle) and in the near future we expect the results and any proposals for possible changes.

15.6 Visualization Adaptation

To make an ECG visualisation the measured data are needed at the beginning. The measurement is made on bipolar ECG corbel and 12 channels BlueECG.
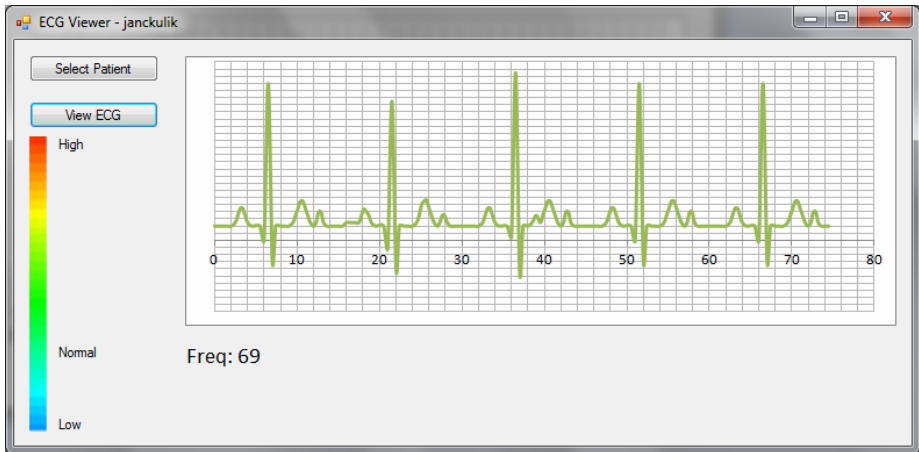
Fig. 15.3 UA highlighting via the ECG curve color (normal pulse)
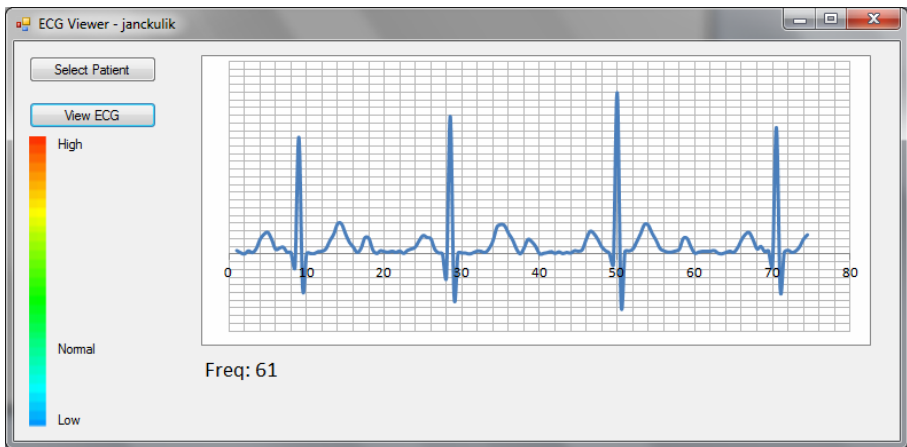
Fig. 15.4 UA highlighting via the ECG curve color (elevated pulse)

15.7 Conclusions

Described Biomedical Adaptive Systems along with his safety features provide a friendly user interface for all type of users who need to operate with. We try to find an easy to use, most comfortable and most intuitive interface to provide processed services with their data to user. In medical areas a time is most often the most important things for patients rescue. Therefore the need of the most intuitive and easy to use interfaces at every electronics device which the medical personal need to use is the most critical. While we suggest several suggestions to use them in adaptation, some of them are still be uncovered. The solution or the simplest

way for implementation friendly UI for embedded devices with Windows CE can be a Microsoft Silverlight technology which is possible startup on Windows CE devices. The visualization of the measured data was reached in case of WPF usage. As the final improvement in the future, the application would have some special algorithm, which could recognize any symptoms of the QRS curve, and make the job for the doctors much easier.

References

 1. Janckulik, D., Krejcar, O., Martinovic, J.: Personal Telemetric System – Guardian. In: Biodevices 2008, Insticc Setubal, Funchal,Portugal, pp. 170–173 (2008)
 2. Krejcar, O., Cernohorsky, J., Janckulik, D.: Portable devices in Architecture of Personal Biotelemetric Systems. In: 4th WSEAS International Conference on Cellular and Molecular Biology, Biophysics and Bioengineering, BIO 2008, Puerto De La Cruz, Canary Islands, Spain, December 15-17, pp. 60–64 (2008)
 3. Krejcar, O., Cernohorsky, J., Czekaj, P.: Secured Access to RT Database in Biotelemetric System. In: 4th WSEAS Int. Conference on Cellular and Molecular Biology, Biophysics and Bioengineering, BIO 2008, Puerto De La Cruz, Canary Islands, Spain, December 15-17, pp. 70–73 (2008)
 4. Krejcar, O., Cernohorsky, J., Janckulik, D.: Database Architecture for real-time accessing of Personal Biotelemetric Systems. In: 4th WSEAS Int. Conference on Cellular and Molecular Biology, Biophysics and Bioengineering, BIO 2008, Puerto De La Cruz, Canary Islands, Spain, December 15-17, pp. 85–89 (2008)
 5. Krejcar, O., Janckulik, D., Motalova, L., Kufel, J.: Mobile Monitoring Stations and Web Visualization of Biotelemetric System - Guardian II. In: Mehmood, R., et al. (eds.) EuropeComm 2009. LNICST, vol. 16, pp. 284–291. Springer, Heidelberg (2009)
 6. Krejcar, O., Janckulik, D., Motalova, L.: Complex Biomedical System with Mobile Clients. In: Dössel, O., Schlegel, W.C. (eds.) The World Congress on Medical Physics and Biomedical Engineering 2009, WC 2009 Proceedings, Munich, Germany, September 07-12. IFMBE, vol. 25(5), Springer, Heidelberg (2009)
 7. Krejcar, O., Janckulik, D., Motalova, L., Frischer, R.: Architecture of Mobile and Desktop Stations for Noninvasive Continuous Blood Pressure Measurement. In: Dössel, O., Schlegel, W.C. (eds.) The World Congress on Medical Physics and Biomedical Engineering 2009, WC 2009, Munich, Germany, September 07-12. IFMBE Proceedings, vol. 25(5). Springer, Heidelberg (2009)
 8. Penhaker, M., Cerny, M., Martinak, L., Spisak, J., Valkova, A.: HomeCare - Smart embedded biotelemetry system. In: World Congress on Medical Physics and Biomedical Engineering, Seoul, South Korea, August 27-September 01, vol. 14(PTS 1-6), pp. 711–714 (2006)
 9. Brida, P., Duha, J., Krasnovsky, M.: On the accuracy of weighted proximity based localization in wireless sensor networks. In: Personal Wireless Communications. IFIP, vol. 245, pp. 423–432 (2007)
10. Cerny, M., Penhaker, M.: Biotelemetry. In: 14th Nordic-Baltic Conference an Biomedical Engineering and Medical Physics, IFMBE Proceedings, Riga, Latvia, June 16-20, vol. 20, pp. 405–408 (2008)

Chapter 16
Exploration of Continuous Sequential Patterns Using the CPGrowth Algorithm

Marcin Gorawski, Pawel Jureczek, and Michal Gorawski

Abstract. In the following paper we present the UCP-Tree and a new algorithm called CPGrowth for continuous pattern mining. The UCP-Tree is an aggregation tree that stores common subsequences of input sequences in the same nodes. The characteristic feature of the CPGrowth algorithm is that it does not require transitional trees at the next recursion levels. Moreover, new sequences can be inserted into the UPC-Tree without rebuilding, which is a considerable advantage considering that Trajectory Data Warehouses store massive amounts of data. In this paper we compared the efficiency of the proposed index with one of the fastest continuous pattern mining algorithms.

16.1 Introduction

Frequent pattern mining in various applications enables to discover repeated situations. In the following paper we discuss the efficient mining of continuous patterns that helps maintaining the spatial continuity of regions of interest in a regular grid. The proposed continuous pattern mining algorithm, called CPGrowth, discovers frequent continuous routes that allow precise analysis of objects' behavior. The CPGrowth algorithm can also be utilized in other scientific fields. The base structure for this algorithm is the UCP-Tree (Updatable Continuous Pattern Tree) which is in fact an aggregation tree. This tree enables aggregation of similar sequences in the same nodes.

In the scientific literature we can find many algorithms that can be used directly or be adequately modified in order to obtain continuous patterns. In [5] the aggregation tree is presented, which is an index structure that stores sequences of the web sites visited by users. The usage of such tree enables aggregation of similar sequences in

Marcin Gorawski, Pawel Jureczek, and Michal Gorawski
Silesian University of Technology, Institute of Computer Science,
Akademicka 16, 44-100 Gliwice, Poland
e-mail: Marcin.Gorawski@polsl.pl, Pawel.Jureczek@polsl.pl, Michal.Gorawski@polsl.pl

the same nodes and thus reduces the size of analyzed data. The aggregation tree is also called the prefix tree. The authors of [3] extended the idea of the aggregation tree by an auxiliary structure (header table) that stores links to nodes containing the same labels. In the paper [4] authors present frequent pattern mining using the WAP-Tree index. This algorithm is quite respectable when mining sequences that contain the same repeated elements. However, comparing to our algorithm, WAP-Tree does not preserve continuity of input sequences. The authors of [8] based on WAP-Tree developed the SMAP-Mine algorithm along with the SMAP-Tree and CMAP-Tree indexes for mining sequential and continuous patterns associated with mobile services. The next interesting structure is the FLWAP-Tree shown in [6]. This structure stores in each item of the header table pointers to first occurrences of nodes with a given label in a prefix tree. In [7] a new method of mining is presented, which finds complete paths of mobile objects including recurrent, non-recurrent and backward patterns.

In comparison to above-mentioned research, our algorithm searches for patterns in a sequences set, in which a single sequence does not contain repetitions of elements. This assumption enables us to omit the creation of the prefix tree in a recursive call. Moreover, the presented index allows inserting new sequences into the prefix tree without the need of rebuilding the initial tree.

The rest of the paper is structured as follows. Section 2 introduces a definition of continuous patterns. Section 3 introduces a manner of continuous pattern mining followed by algorithm pseudocodes in Section 4. In Section 5 experimental results are shown and Section 6 summarizes the paper.

16.2   Continuous Sequential Patterns

To continue we have to make some assumptions:

1. Each element occurs only once in a sequence, which means that an element is unique for a certain sequence, e.g., the sequence A→ B→ A→ C is not valid, since the element A occurs twice. This assumption is not rigorous, because in the case of mobile object trajectories, after the mapping of these trajectories into region sequences, there are rarely sequences with repeated elements. For example, while planning a car route, we choose the shortest route or a navigation system does this for us.
2. Because we analyze frequent sequences of regions of interest for a regular grid, it is important to preserve the region continuity in sequences — this assumption concerns spatial continuity. This is due the fact that when we omit an element of a sequence we interrupt spatial continuity.

Taking into consideration the above-mentioned assumptions let us introduce the following definitions:

Definition 1. *Definition of including one continuous sequence in other sequence:*

The sequence $\langle b_1 b_2 \ldots b_m \rangle$ *is a continuous subsequence of the sequence* $\langle a_1 a_2 \ldots a_n \rangle$, *where* $n \geq m$, *when for certain integer* i $b_1 = a_i, b_2 = a_{i+1}, \ldots, b_m = a_{i+m-1}$, *and when for two elements* b_i *and* b_j $(i \neq j)$ *of any sequence we have* $b_i \neq b_j$.

Definition 2. *Definition of support:*
  For a sequence database S, *the support of pattern* s *is defined as:*
  $sup(s) = |\{s_i | s_i \in S \wedge s \subseteq s_i\}|/m$ *for* $1 \leq i \leq m$.

Definition 3. *Definition of a sequence frequency:*
  A pattern s *is called frequent if its support is no less than a threshold given by the user:*
  $sup(s) \geq minSup$.

16.3 The UCP-Tree Index and the CPGrowth Algorithm

The UCP-Tree index is created on a basis of input region sequences. While building the index, continuous sequences are inserted into a prefix tree and items are created in a header table. These items have pointers to elements of compressed sequences.

Example. In Table 1 input sequences are presented. These sequences will be used in the index building process.

Table 16.1 Input sequences

No.	Input sequences
1	A→ G→ F→ I→ B
2	A→ G→ F→ J→ K
3	C→ G→ Z→ B
4	B→ C→ J→ K
5	B→ G→ F
6	B→ C→ J

In Fig. 1 there is a prefix tree which was created after inserting the first three sequences from Table 1. The header table is created at the beginning of a tree building process and during insertion, new items are added along with pointers to appropriate tree nodes. The insertion of sequences into the prefix tree begins from its root, which is marked as *null*. Having sequence B→ C→ J→ K we check if the root has on the list of its children the node with the label B. Because such node does not exist, we have to create it, add it to the appropriate list in the header table and set its support to 1. Because the newly created node has no children, the described process is repeated for the last three elements of the sequence — C, J and K.

The final UCP-Tree is presented below in Fig. 2.

After creating the UCP-Tree, we traverse it for each frequent item of the header table at a given recursion level — the order of elements from this table does not influence generated patterns. The frequent pattern mining is realized by the CPGrowth algorithm.

In the CPGrowth algorithm we use the fact that at each recursion level it is sufficient to check only the number of occurrences of each element that occurs directly
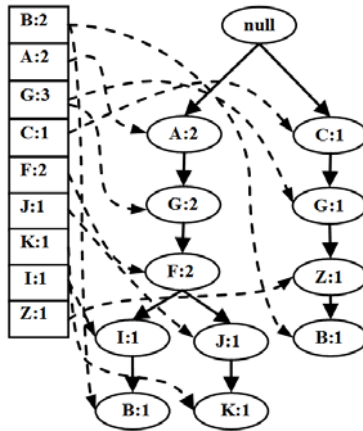
Fig. 16.1 Building the UCP-Tree
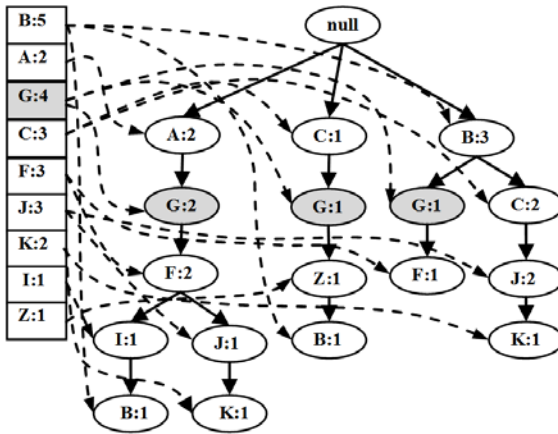
Fig. 16.2 Final UCP-Tree

after the element from the header table. Each item of the header table has a form *label*:*support*, where *label* denotes the label of an element and *support* informs in how many sequences this element occurred. Moreover, each item has a set of pointers which point to elements in a prefix tree.

Example. Let *minSup* be set to 2. When choosing the element G from the initial header table (Fig. 2), the CPGrowth algorithm will store it in a result set, will add this element to the end of a current prefix and will determine the supports for elements which occur in a prefix tree directly after G. In our case we have the elements F and Z, however, only the element F (support=3) is frequent. Because two elements are found, a new header table will be created, in which two elements F and Z will be placed and new pointer lists will be created — only for nodes that occurr directly
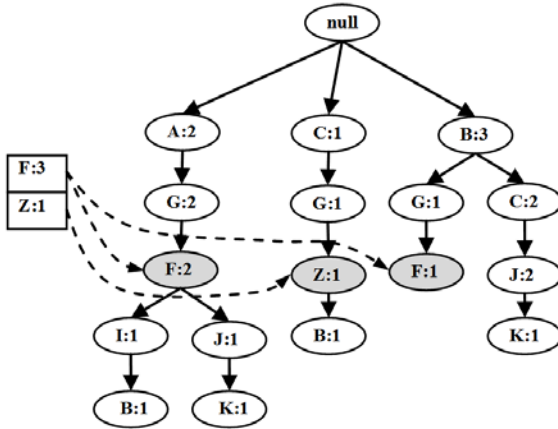
Fig. 16.3 CPGrowth algorithm

after G. In the next step the CPGrowth algorithm is called for the newly created header table and the prefix $\langle G \rangle$. The next iteration is shown in Fig. 3. Because there is no frequent element after F, the CPGrowth algorithm ends.

16.4   Pseudocodes

Below we present pseudocodes for the UCP-Tree index and the CPGrowth algorithm.

1. Scan a sequence database S once and build the UCP-Tree according to Algorithm 2.
2. Call the CPGrowth algorithm for the UCP-Tree.

During scanning the sequence database S, sequences are inserted into the UCP-Tree according to Algorithm 2. After the creation of the UCP-Tree, the CPGrowth algorithm is called.

1) Create the UCP-Tree root.
2) For each sequence s in S do:
a. Set the current node n to the UCP-Tree root and a pointer p to the first element in the sequence s.
b. Check if the node n has, among its children, a node with a label pointed by p, if so, increase its support by 1, set the current node n to this child and go to step c), if not, go to step d).
c. If the pointer p points to the last element of the sequence s, go to the next sequence in S — the next iteration of the loop 2). Otherwise set the pointer p to the next element of the sequence s and go to step b).
d. Create a new node w for the element pointed by p, set its support to 1 and set the current node n to the newly created node w. If the item e with a label pointed by

p does not exist in a header table, create it and set its support to 0. Add *w* to the list of the item *e* and increase its support by 1. Go to point c).

1) For each item *e* stored in the header table do:

a. If the item *e* has support greater or equal to *minSup*, add *e* to the end of the prefix. The pattern created in such manner is added to the result set.

b. Create a new header table based on the list of nodes *nl*, which are associated with the item *e* of the header table.

c. If a new header table is not empty, call the CPGrowth algorithm for this table and the prefix.

d. Remove the last element of the prefix.

The first iteration of the CPGrowth algorithm is for the empty prefix.

16.5 Experiments

In order to examine the efficiency of our index we generated trajectories using the Brinkhoff generator [1]. The efficiency of the CPGrowth algorithm was compared to the efficiency of the VGES algorithm which was proved to be one of the most efficient algorithms [2]. All algorithms were implemented in Java (JDK1.6) and experiments were conducted on a platform with: Intel Core Duo E6550 2.33GHz CPU and 4GB RAM. The operating system was Windows 7 with Oracle 10g database.

Fig. 4 shows the runtime comparison of the CPGrowth and VGES algorithms for different numbers of trajectories (10000, 20000, 30000, 40000 and 50000) and different values of the minimal support *minSup*.

Fig. 5 presents detailed results for the CPGrowth algorithm.

The charts in Fig. 5 prove that the CPGrowth algorithm is always more efficient than VGES. This is caused by the fact that the CPGrowth algorithm is called for the aggregation tree which enables simultaneous analysis of identical subsequences.
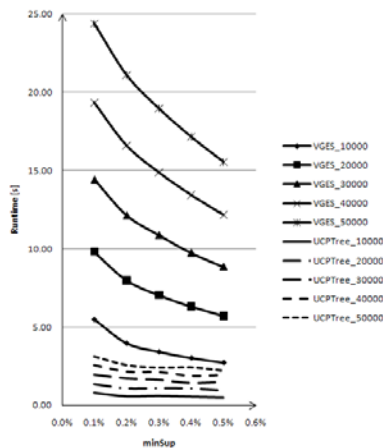
Fig. 16.4 CPGrowth vs. VGES
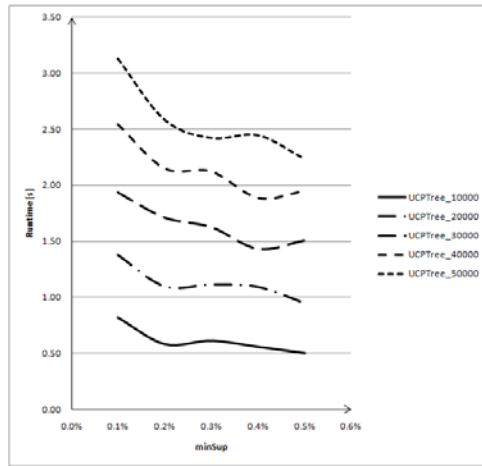
Fig. 16.5 Performance of CPGrowth

In the case of both algorithms an increase in the number of sequences increases runtimes. In the case of the CPGrowth algorithm this increase is partly caused by increase of sequences which have to be inserted into the UCP-Tree.

16.6   Conclusion

In the paper we present the CPGrowth algorithm for continuous pattern mining, which bases on the UCP-Tree. The UCP-Tree is an aggregation tree which enables aggregation of identical subsequences of input sequences in the same nodes. Moreover, the UCP-Tree allows adding new sequences without rebuilding the index. We also performed efficiency experiments of the CPGrowth algorithm and we compared it with the VGES algorithm.

References

1. Brinkhoff, T.A.: A Framework for Generating Network-Based Moving Objects. Geoinformatica, 153–180 (2002)
2. Gorawski, M., Jureczek, P.: A Proposal of Spatio-Temporal Pattern Queries. In: The 4th Int. Conf. on Complex, Intelligent and Software Intensive Systems, pp. 587–593 (2010)
3. Han, J., Pei, J., Yin, Y.: Mining Frequent Patterns without Candidate Generation. In: Proc. of the 2000 ACM SIGMOD Int. Conf. on Management of Data, pp. 1–12 (2000)
4. Pei, J., Han, J., Mortazavi-Asl, B., Zhu, H.: Mining Access Patterns Efficiently from Web Logs. In: Proc. of Pacific-Asia Conf. on Knowledge Discovery and Data Mining, pp. 396–407 (2000)
5. Spiliopoulou, M., Faulstich, L.C.: WUM: A tool for web utilization analysis. In: Atzeni, P., Mendelzon, A.O., Mecca, G. (eds.) WebDB 1998. LNCS, vol. 1590, pp. 184–203. Springer, Heidelberg (1999)

6. Tang, P., Turkia, M.P., Gallivan, K.A.: Mining web access patterns with first-occurrence linked WAP-trees. In: Proc. of the 16th Int. Conf. on Software Engineering and Data Engineering, pp. 247–252 (2007)
7. Tseng, S., Chan, W.C.: Mining complete user moving paths in a mobile environment. In: Proc. of the Int. Workshop on Databases and Software Engineering (2002)
8. Tseng, S., Lin, K.W.: Efficient mining and prediction of user behavior patterns in mobile web systems. Information and Software Technology, 357–369 (2006)

Chapter 17
Detecting New and Unknown Malwares Using Honeynet

Michał Szczepanik and Ireneusz Jóźwiak

Abstract. The importance of network security is rapidly increasing as more and more business is conducted via these systems. The proposed honeynet system can be used to detect bots or malware based on the evaluation of events occurring within a computer network. A honeypot is a trap set to detect, deflect or in some manner counteract attempts at unauthorized access to information systems. A honeynet (a network consisting of 2 or more honeypots) is used for surveillance of larger or more diverse networks for which one honeypot may not be sufficient. Honeynets are fast emerging as an indispensible forensic tool for the analysis of malicious network traffic. Honeypots can be considered to be traps for hackers and intruders and are generally deployed complimentary to Intrusion Detection Systems (IDS) and Intrusion Prevention Systems (IPS) in a network. Proposed system would be capable of providing cures for new fatal viruses which have not yet been discovered by security firms.

17.1   Introduction

A honeynet [1] is a network set up with intentional vulnerabilities. Its purpose is to invite attacks, so that an attacker's activities and methods can be studied and that information used to increase network security. A honeynet contains one or more honey pots, which are computer systems on the network set up expressly to attract and trap people who attempt to penetrate computer systems.

The primary purpose of a honeynet is to gather information about the attacker's methods and specific targets, but the operator can benefit from it in other ways too, e.g. through diverting attackers from a real network and its resources. The honeypot/honeynet usually has real applications and services so that it appears as a normal network and worthwhile target. However, since the honeynet doesn't actually serve any authorized users, any attempt to contact it from without is likely an

Michał Szczepanik and Ireneusz Jóźwiak
Wrocław University of Technology, Institute of Informatics, Poland
e-mail: {michal.szczepanik,ireneusz.jozwiak}@pwr.wroc.pl

illicit attempt to breach its security, and any outbound activity is likely evidence that the system has been compromised. For this reason, information about the suspect is much more apparent than it would be in an actual network, where it would have to be found amidst all the legitimate network data.

17.2 Honeypot's Network

Honeypots can be classified based on their deployment and based on their level of involvement [4]. Based on the deployment, honeypots may be classified as production honeypot and research honeypot. Production honeypots are easy to use, capture only limited information, and are used primarily by companies or corporations. Research honeypots are complex to deploy and maintain, capture extensive information, and are used primarily by research, military, or government organizations.
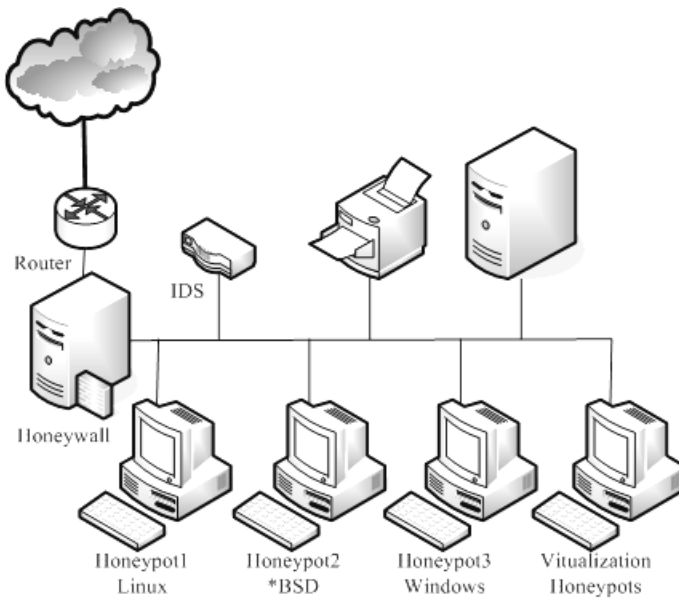
Fig. 1 Honeynet structure

Honeypots are not production systems, the honeynet itself has no production activity and no authorized services. As a result, any interaction with a honeynet implies malicious or unauthorized activity. Any inbound connection initiated to your honeynet is most likely a probe, scan, or attack. Any unauthorized outbound connections from your honeynet imply someone has compromised the system and ~~has~~ initiated outbound activity. This makes analyzing activity within your honeynet very simple. With traditional security technologies, such as firewall logs or IDS sensors, you have to sift gigabytes of data, or thousands of alerts.

Honeynets are usually implemented as parts of larger network intrusion-detection systems [2] with IDS (software), sniffer, logs server and PC's with with different OS to simulation all types of systems on the network.

A Honeywall is a device or set of devices configured to monitor, analyze and permit or deny all computer traffic between different domains based upon a set of rules and other criteria. A Honeywall is the first system to check traffic data between honeypots and Internet. Every kind of anomaly will be checked by IDS/IPS or a firewall in the honeywall.

17.3 Malware

The term malware, also known as malicious code and malicious software, refers to a program that is inserted into a system, usually covertly, with the intent of compromising the confidentiality, integrity, or availability of the victim's data, applications, or operating system. Malware is usually designed to perform these nefarious functions in such a way that users are unaware of them, at least initially. Today, hackers have developed all sorts of ways to invade computers and they new trends in malware are trojan horses, rootkits, and backdoors. In this chapter, we propose use agents in honeynet to detection this type of malware.

Organizations should strive to detect and validate malware incidents rapidly because infections can spread through an organization within a matter of minutes. Early detection can help minimize the number of infected systems, which will lessen the magnitude of the recovery effort and the amount of damage the organization sustains.

Since no indication is completely reliable and even antivirus software might misinterpret benign activity as a malicious incident, handlers need to analyze any suspicious activity and ascertain that malware truly is the underlying cause. In some cases, such as a massive, organization-wide infection, validation may be unnecessary because of the obvious nature of the incident. The goal is for incident handlers to be as certain as feasible that an incident is caused by malware and to have basic understanding of the type of malware threat responsible, such as a worm or a Trojan horse. If the source of the incident cannot be easily confirmed, it is often better to respond as if it were caused by malware and to alter response efforts if it is later determined that malware is not involved. Waiting for conclusive evidence of malware might have a serious negative impact on response efforts and significantly increase the damage sustained by the organization.

As part of the analysis and validation process, incident handlers typically identify characteristics of the malware activity by examining detection sources. Understanding the characteristics of an activity is very helpful in assigning appropriate priority to the incident response efforts and planning effective containment and eradication.

Detection tools weren't capable of recognizing or stopping malware when it was a new threat, but honeynets enable collecting extensive information on a variety of threats and recognizing them .In order To obtain this information systems have to allow malicious code access - potentially privileged access - to the honeynet. First, systems have to give the malware some degree of freedom. The more

activity we allow the malicious code to perform, the more we can potentially learn about it. However, the more freedom we give it, the greater the risk it will circumvent data control and harm other systems. In that case the system should deny access to the malicious code by, e.g. simply shutting down.

Containment of malware consists of two major steps: stopping their spread and preventing further damage Nearly every malware incident requires containment actions. In addressing an incident, it is important to decide which methods of containment to employ, early in the response. Containment of isolated incidents and incidents involving noninfectious forms of malware is generally straightforward, involving such actions as disconnecting the affected systems from networks or shutting them down. For more widespread malware incidents, a strategy should be devised to contain the incident in most systems as quickly as possible. The actions undertaken should keep the number of infected machines, the damage done and the time needed to complete data and service recovery to a minimum.

17.4 Multi-agents System in a Honeynet

Constant monitoring of a network by a human operator is very expensive and not effective. Using multi-agents system is faster and more effective, because they can detect every anomaly and check it in real time. It can be he best method to detect trojan horses, rootkits, and backdoors Analysis of honeypots is simpler than that of production networks, so offer a worthwhile option for detecting malware by agents. Each system anomaly arises only from the infected system.

The proposed system uses three types of agents. The first and second acts in Honeynet. The third analyze the production network. First agents work like an IDS and check system activity and look for typically compromised system anomalies [3] like: services crashes; users complaining of slow access to hosts on the internet; exhaustion of system resources; slow disk access or slow system boots; programs starting, running slowly or not running at all; unknown processes; unusual and unexpected port openings (typical for Trojan horses and backdoors); deletion, corruption or lack of access to files; port scans and failed connection attempts targeted at the vulnerable service; filenames with unusual characters; configuration changes disabling of security controls such as antivirus software and personal firewalls. This agent first analyzes a malware program in a controlled environment (honeynet) to build a model that characterizes its behavior and then sent information (model) to next agents.

Should something suspicious happen, a second agent should check what processes are run and when. Then scan them with antivirus software or compare with other systems without anomalies. This agent monitors and logs all of the threat's activities within the honeynet (Fig 17.2). It is this captured data that is subsequently analyzed to study the tactics of attackers. If found information like which versions of operating systems, devices, applications, etc., may be affected and how the malware infects the system (e.g., vulnerability, misconfiguration) connected with others agents in production system and inform them how the malware affects the infected system, including the names and locations of affected files, altered configuration settings, installed backdoor ports.
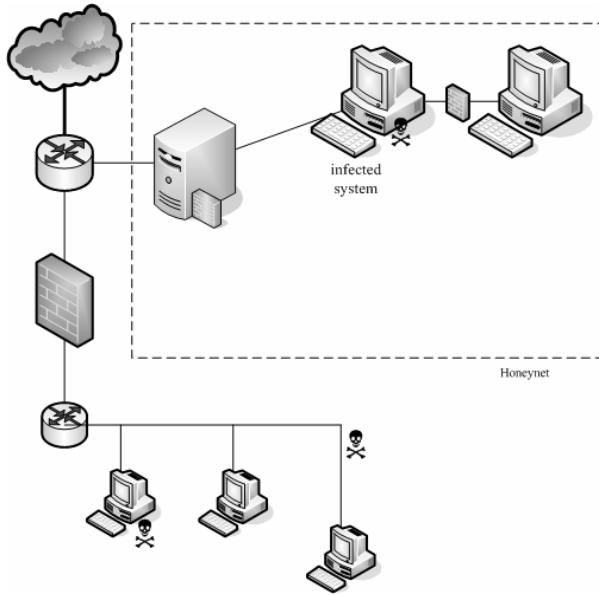
Fig. 17.2 Infected system. In this system, the first infected is vulnerable to attacks honeynet system. The protected network of production has not yet been attacked and the multiagent system can begin an analysis of the honeynet.
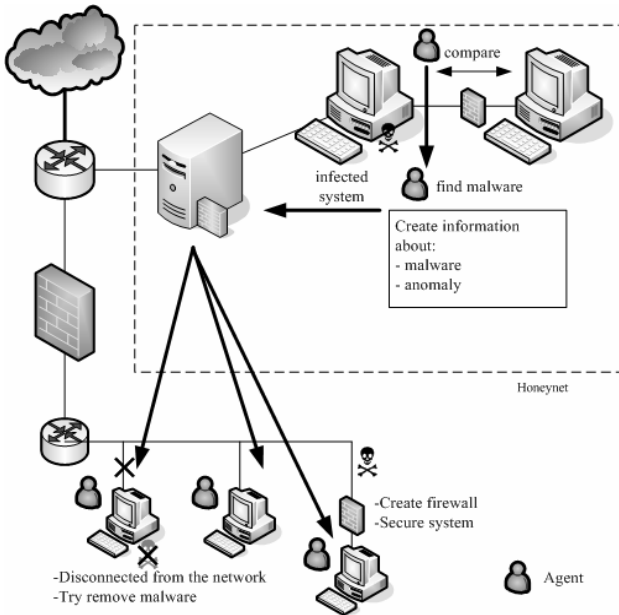
Fig. 17.3 Honeynet with agents on. Agents located in the production network to eliminate the threat. They are based on information gathered by HoneyNet.

Last agents using information form first and second agent identify hosts in the production network, which are infected by malware (Fig. 17.3). Once identified, infected hosts can undergo appropriate containment (for example disconnect them from networks), eradication, and recovery actions (run antivirus or update it).

The main task of the first and second agent is to collect information about malware and an attempt to identify them by using antivirus software. The last agent on the basis of collected information in their attempts to remove or restrict the operation of malware in the production network. Use HoneyNet greatly simplifies the collection of information, as the system is devoid of traffic and incidents occurring in the production network.

17.5 Detection of Malicious Traffic

Detection of new variants of malware is among the major research activities that are carried out by security firms in the IT industry. A majority of antivirus firms and research organizations use honeynets to capture variants of existing as well as new malware. They work upon the acquired data and binary samples to produce patches for defense against malware threats. The combination of such data from a variety of sources proves to be extremely useful in creating efficient security patches [5].Several technologies (ex. monitor IDS, IPS) would be involved in the creation of a mechanism for the detection of malware. We will tweak these existing technologies into our honeypot. This would ensure that our system produces efficient output to the maximum possible extent with a minimum number of false positives.

Apart from the ability to detect malware, our honeypots would in the future also be capable of generating cures in the form of signature of malware for every malware detected by it. The first phase of our system involves monitoring of all network traffic that passes through our honeynet and detection of code that possibly may be malicious in nature by compare this system with other secured system without this anomaly. Suspected pieces of malicious code will then be redirected to a virtual machine installed on the honeypot itself. Program monitoring tools in the virtual machine would then perform analysis on the code to ensure if it really is malicious in nature or not. If the code is indeed of malicious nature, then the virtual machine would further monitor its activities and create a log of the changes that the code makes to the operating system. Therefore, in order to reduce overheads on the honeypot as well as the virtual machine, it is important for the malware detection algorithms that will be used in this phase to be extremely robust.

Most of the existing malware detection algorithms rely on pre-defined malware signature databases to scan network traffic for traces of patterns which are known to them. However, the aim of our system is to protect a network against new and unknown malware. The use of pre-defined signature based detection algorithms would not produce satisfactory results in our case. Therefore, we need to device a new algorithm that would enable us to detect new and unknown patterns that may potentially be malicious in nature.

We propose a malware detector service which analyses the captured network traffic, extracts relevant network traffic patterns from it and then uses several

statistical methods to determine which patterns may be malicious in nature. It would help if the captured packets are converted into packet sets before analysis. We would be generate hash codes for patterns that are relevant to the malware detection process. The main advantage of using hashing to compare patterns over direct string comparison lies in the speed of comparison. Let us consider a pattern (C1 C2 … Cn) to be hashed. Let the lower limit of pattern length be K and the number of bits in Bloom filter be M.

$$H = c_1q^{k-1} + c_2q^{k-2} + \ldots + c_{k-1}q + c_k \ (Mod \ M)$$

The Bloom filter, conceived by Burton H. Bloom in 1970, is a space-efficient probabilistic data structure that is used to test whether an element is a member of a set. False positives are possible, but false negatives are not. Elements can be added to the set, but not removed (though this can be addressed with a counting filter). The more elements that are added to the set, the larger the probability of false positives.

Each packet set contains several packets captured in a sequence. This ensures that every byte that is transferred on a network is also stored in the captured packets. Our next task is to extract from these packet sets, all the patterns that are relevant to our detection process. Relevance may be judged on the basis of several parameters, the most common one being length. Extremely short patterns (those with length less than 10 bytes) have the tendency of occurring with high frequencies in malware as well as legitimate traffic rendering such patterns useless for our algorithm. In contrast, studies show that the length of a typical malware code does not exceed a certain amount to ensure its fast propagation over the Internet. This implies that we need to look for patterns with length that is optimal for malware identification by setting lower and upper thresholds.

After all the packet sets are searched for patterns of interest to us, our coincidence count tables get ready for statistical analysis. We lay more emphasis on those patterns that have a higher value of count of that pattern. Intuitively, we can set a threshold above which all the patterns need to be analyzed in detail to see the changes they make to an OS. However, proceeding in this fashion may lead to a high rate of false positives, something that we wish to reduce in order to improve the overall performance of our system. We therefore use a statistical technique called Inverse Distribution followed by Standard Deviation to further analyze value of relevant patterns. An efficient algorithm for performing inverse distribution analysis has been presented by [7]. We wish to use a similar technique to reduce the number of false positives generated by our detection scheme.

This bloom filter is use to analyze network traffic in both the production network and honeypots network to detect where and which part of network can be infected.

17.6 Conclusion

Although the primary goal of eradication is to remove malware from infected systems, eradication is typically more involved than that. If an infection was successful because of a system vulnerability or other security weakness, such as an unsecured

file share, then eradication includes the elimination or mitigation of that weakness, which should prevent the system from becoming reinfection or infection by another instance of malware or a different variant of the original threat Honeynets are a form of a high-interaction honeypots. They can help find new malware in networks and use mulit-agents systems to inform all hosts in production networks about infection.

References

1. Piotrowski, M.: The protection of computer networks through technology honeypot (2007) (in Polish)
2. Rush, M., Orebaugh, A., Clark, G., Pinkard, B., Babbin, J.: Intrusion Prevention and Active Response - Deploying Network and Host IPS (2005)
3. Erickson, J.: Hacking: The Art of Exploitation (2008)
4. Holz, T., Dornseif M.: Hands on Honeypot Technology (Black Hat 2005), `http://blackhat.com`
5. Auerbach, O.: AVIRA, Evolution from a Honeypot to a distributed Honeynet
6. Ahmadi, M., Wong, S.: A Cache Architecture for Counting Bloom Filters. In: 15th International Conference on Networks ICON 2007 (2007)
7. Karamcheti, V.: Detecting Malicious Network Traffic, Using Inverse Distributions of Packet Contents. In: SIGCOMM 2005 Workshops, Philadelphia, PA, USA, August 22-26 (2005)

Chapter 18
Average Prior Distribution of All Possible Probability Density Distributions

Andrzej Piegat and Marek Landowski

Abstract. Bayes' rule is universally applied in artificial intelligence and especially in Bayes reasoning, Bayes networks, in decision–making, in generating rules for probabilistic knowledge bases. However, its application requires knowledge about a priori distribution of probability or probability density that frequently is not given. Then, to find at least an approximate solution to a problem, the uniform a priori distribution is used. Do we always have to use this distribution? The paper shows that it is not true. The uniform prior should only be used if there is no knowledge about the real distribution. If however, we possess certain qualitative knowledge, e.g. that the real distribution is the unimodal one, or that its expected value is less than 0.5, then we can use this knowledge and apply a priori distribution being the average distribution of all possible unimodal distributions, instead of the uniform distribution. As a result we will usually get better approximation of the problem solution and will avoid large approximation errors. The paper explains the concept of average distributions and shows how they can be determined with a special method of granulation diminution of elementary events and probability.

18.1   Introduction

Bayes' rule is universally used in artificial intelligence in problems with uncertain data, with probabilistic variables, as Bayesian inference, Bayesian networks, automated reasoning and others [4, 6, 9, 12, 13]. Frequently probabilities of events are

Andrzej Piegat
West Pomeranian University of Technology, Faculty of Computer Science and Information Systems, Zolnierska 49, Szczecin 71–210, Poland
e-mail: apiegat@wi.zut.edu.pl

Marek Landowski
Maritime University of Szczecin, Quantitative Methods Institute,
Waly Chrobrego 1–2, Szczecin 70–500, Poland
e-mail: m.landowski@am.szczecin.pl

not known precisely (no measurements, no trials, no experiments are at disposal). Then their approximations can be determined by experts on the basis of their knowledge about the problem. Suppose an expert, e.g. an academic teacher gives evaluation "probability that student X will pass examination from Artificial Intelligence equals 0.2". This evaluation will not be fully credible in respect of its high precision, if the teacher does not possess the results of statistical investigations and bases his/her evaluation only on the general knowledge about students or about the student X and on his intuition. A more credible are teachers' evaluations as "the probability is about 0.2 or approximately 0.2" or "the probability lies between 0.1 and 0.3" etc. Such evaluations can be identified and described (modeled) by distributions of probability density [3]. In such situation a special version of Bayes' theorem for probability densities has to be used,

$$f_X(x|Y=y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{f_Y(y|X=x)f_X(x)}{f_Y(y)}$$

where: $f_{X,Y}(x,y)$ is the joint density function of X and Y, $f_X(x|Y=y)$ is the posterior probability density function of X given $Y=y$, $f_Y(y|X=x) = L(x|y)$ is (as a function of x) the likelihood function of X given $Y=y$, $f_X(x)$ and $f_Y(y)$ are the marginal probability density functions of X and Y respectively, where $f_X(x)$ is the prior probability density function of X.

As the practice of application of Bayes' networks shows prior distributions $f_X(x)$ frequently are not known [4, 9, 12, 13, 14] because their determining requires a lot of measurements, experiments, observations, etc. In this situation, according to the indifference principle [1, 2, 5, 8] the uniform a priori distribution can be used. In scientific literature, e.g. in [4, 10], the uniform distribution is sometimes called maximum entropy distribution. This, certainly, is the true feature of the uniform distribution. However, the uniform distribution also possesses other features. It can be shown that the uniform distribution is also the average distribution of an infinite number of all possible distributions of probability density that the real distribution can take, in the problem under consideration, assuming that probability of each distribution is equal. However, how could the average distribution of an infinite number of possible distributions be determined? It seems impossible! However, this task can be realized with the method of granulation diminution (decreasing) of elementary events which was conceived by Andrzej Piegat [11].

18.2 Safe a Priori Distribution of Probability Density in the Case of Complete Ignorance of the Real Distribution

Let us suppose we only know the interval of variable X: $x \in [x_{\min}, x_{\max}]$ and for simplicity let us assume that $x_{\min} = 0$ and $x_{\max} = 1$. Apart of the interval we know nothing about the real distribution, and in particular we do not know its modality (the number of maxima), whether it is symmetric or asymmetric (and if asymmetric
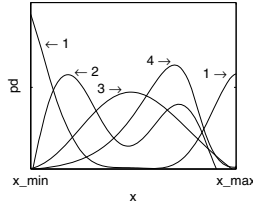
Fig. 18.1 4 example distributions of an infinite number of possible distributions of variable X (notice that distribution 1 is border distribution and distribution 2 is bimodal)

then whether it is positive or negative skew [7]) etc. Therefore we can not exclude any form of the real distribution, Fig. 18.1.

If we have no knowledge which of the possible distributions is more and which is less probable then, according to Laplace' indifference principle, equal probability of each distribution is assumed. However, if we would had a certain qualitative knowledge about the real distribution (unimodal, bimodal, symmetric, positive skew, etc), it could be used in determining average distributions as it will be shown in next chapters. It is important to correctly understand the sentence "we assume equal probability of each possible distribution". The assumption of equal probability is not a statement that each distribution has equal probability because there exists no evidence for such statement. The assumption of equal probability is only a working assumption that allows to avoid extremely large solution errors. There exists here a full analogy with a biased coin flipping [1, 2]. If we do not know which side of the coin has large probability then it is safely to assume, as start–hypothesis in the problem solving the equal probability 1/2 of both sides. In this assumption the maximal absolute error of the bias evaluation can be 1/2. If we assume the start–probability e.g. 0.95 for head then the maximal possible error can be 0.95. Thus, the indifference principle allows to generate safe assumptions on condition of full ignorance that protects us against error catastrophes. Understanding that indifference principle suggests not statements but only safe, working assumptions is a key to explanation of many "paradoxes" concerning this principle that can be found in literature [4, 14]. Determining the average distribution of an infinite number of all possible distributions is possible with the method of granulation diminution (GD–method) mentioned in the introduction.

In the first step of the method the interval $[0,1]$ of variable x is partitioned into small number of subintervals, e.g. in 3. Then 3–elementary events for variable x are possible:

$$E1: x \in [0,1/3), E2: x \in [1/3,2/3), E3: x \in [2/3,1].$$

The above means that granulation of variable x equals 1/3. Also granulation of event probability is assumed equal to 1/3, which means $p_{Ei} = p_i \in \{0,1/3,2/3,1\}$ and the probability will assume only one of 4 possible discrete values. For granulation 1/3, 10 different distributions of probability are possible, Fig. 18.2.
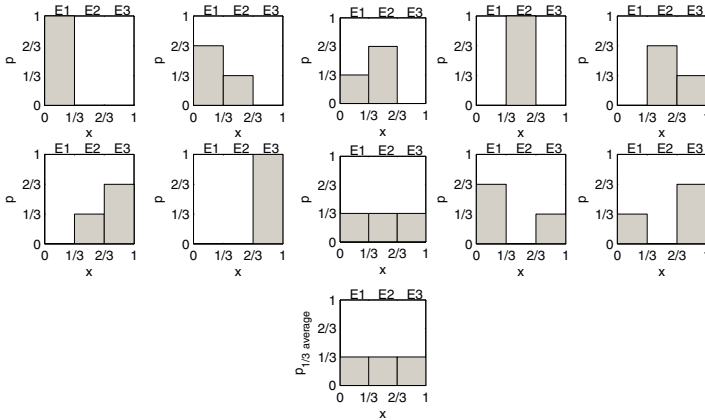
Fig. 18.2 10 distributions of probability possible for granulation 1/3 of x–events and of probability and their average distribution.

Table 18.1 10 distributions of probability possible for granulation 1/3 and their average square (MSE) and absolute (MAE) errors in relation to all 10 distributions ($i = 3$: uniform distribution).

Probability distribution	Interval of the variable x				
	$[0, 1/3)$	$[1/3, 2/3)$	$[2/3, 1]$	MSE	MAE
j	$p_{j,1}$	$p_{j,2}$	$p_{j,3}$		
1	3/3	0	0	0.3333	0.4444
2	2/3	1/3	0	0.1852	0.3333
3	1/3	2/3	0	0.1852	0.3333
4	0	3/3	0	0.3333	0.4444
5	0	2/3	1/3	0.1852	0.3333
6	0	1/3	2/3	0.1852	0.3333
7	0	0	3/3	0.3333	0.4444
8	1/3	1/3	1/3	0.1111	0.2667
9	2/3	0	1/3	0.1852	0.3333
10	1/3	0	2/3	0.1852	0.3333
$\sum_{j=1}^{10} p_{j,i}$	10/3	10/3	10/3		
p_{aver}	1/3	1/3	1/3	0.1111	0.2667

It should be noticed that 7 of 10 possible distributions are unimodal and 2 distributions are bimodal. Distributions from Fig. 18.2 can be presented in the form of Table 18.1.

As it is showd in Fig. 18.2 and Table 18.1 the average distribution of the 10 possible distributions for granulation 1/3 is uniform. The average distribution represents all 10 possible distributions. It would be interesting to determine the average absolute error (MAE) and the average square error (MSE) of the average distribution ($i = 3$) in relation to all 10 possible distributions and to determine the same errors for any other distribution $i \neq 3$.

$$MSE_i = \frac{1}{10} \sum_{j=1}^{10} (p_j - p_i)^2$$

$$MAE_i = \frac{1}{10} \sum_{j=1}^{10} |p_j - p_i|]$$

Values of the above errors are given in Table 18.1. As Table 18.1 shows the lowest average absolute error (MAE) and square error (MSE) have the uniform distribution ($i = 3$). Fig. 18.3 shows the average distribution of all 10 distributions possible for granulation $1/3$.
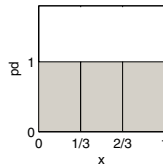
Fig. 18.3 The average distribution of all 10 distributions possible for granulation $1/3$ of x–events and probability – this distribution has min MSE– and MAE–error in relation to all distributions.

In the next step of GD–method the granulation is decreased to $1/4$. Thus 4 elementary events for variable x are possible:

$$E1 : x \in [0, 1/4), E2 : x \in [1/4, 2/4), E3 : x \in [2/4, 3/4), E4 : x \in [3/4, 1].$$

Similarly, for probability p the same granulation is assumed: $p \in \{0, 1/4, 2/4, 3/4, 1\}$. Table 18.2 presents all 35 distributions of probability possible for granulation $1/4$.

If granulation is further decreased then in each case the average distribution is the uniform one. This distribution is not only the average one (as the average distribution it has the minimal average square error MSE) but also the median distribution (as the median distribution it has the minimal absolute error MAE) of all possible distributions because the number of all possible positive skew and negative skew distributions is identical in relation to the uniform distribution. For an infinite small granulation of x–events a continuous distribution (distribution of probability density) is achieved.

In literature, e.g. in [10, 13, 14], the uniform distribution is referred to as the maximum entropy distribution or as a distribution representing maximal ignorance of possible events. However, these are only 2 features of the uniform distribution. As the above presented results have shown the uniform distribution has further important properties: it is the average distribution of all possible distributions and as such has the minimal square error (MSE) in relation to all distributions. It is also the median distribution and as such it has the minimal absolute error (MAE). The above is true on the assumption that probability of all possible distributions is equal.

Table 18.2 35 distributions possible for granulation $1/4$ of x–events and probability, the average distribution (p_{aver}), the average square (MSE) and absolute (MAE) errors of each distribution j in relation to all 35 distributions.

Probability distribution	Interval of the variable x				MSE	MAE
	$[0,1/4)$	$[1/4,2/4)$	$[2/4,3/4)$	$[3/4,1]$		
j	$p_{j,1}$	$p_{j,2}$	$p_{j,3}$	$p_{j,4}$		
1	4/4	0	0	0	0.2625	0.3750
2	3/4	1/4	0	0	0.1688	0.3071
3	2/4	2/4	0	0	0.1375	0.2857
4	2/4	1/4	1/4	0	0.1063	0.2500
5	1/4	1/4	1/4	1/4	0.0750	0.2143
6	0	4/4	0	0	0.2625	0.3750
7	0	3/4	1/4	0	0.1688	0.3071
8	1/4	3/4	0	0	0.1688	0.3071
9	0	2/4	2/4	0	0.1375	0.2857
10	0	2/4	1/4	1/4	0.1063	0.2500
11	1/4	2/4	1/4	0	0.1063	0.2500
12	0	0	4/4	0	0.2625	0.3750
13	0	0	3/4	1/4	0.1688	0.3071
14	0	1/4	3/4	0	0.1688	0.3071
15	1/4	1/4	2/4	0	0.1063	0.2500
16	0	0	2/4	2/4	0.1375	0.2857
17	0	1/4	2/4	1/4	0.1063	0.2500
18	0	0	0	4/4	0.2625	0.3750
19	0	0	1/4	3/4	0.1688	0.3071
20	0	1/4	1/4	2/4	0.1063	0.2500
21	3/4	0	1/4	0	0.1688	0.3071
22	3/4	0	0	1/4	0.1688	0.3071
23	2/4	0	2/4	0	0.1375	0.2857
24	2/4	0	0	2/4	0.1375	0.2857
25	2/4	0	1/4	1/4	0.1063	0.2500
26	2/4	1/4	0	1/4	0.1063	0.2500
27	0	3/4	0	1/4	0.1688	0.3071
28	0	2/4	0	2/4	0.1375	0.2857
29	1/4	2/4	0	1/4	0.1063	0.2500
30	1/4	0	3/4	0	0.1688	0.3071
31	1/4	0	2/4	1/4	0.1063	0.2500
32	1/4	0	0	3/4	0.1688	0.3071
33	0	1/4	0	3/4	0.1688	0.3071
34	1/4	0	1/4	2/4	0.1063	0.2500
35	1/4	1/4	0	2/4	0.1063	0.2500
$\sum_{j=1}^{35} p_{j,i}$	35/4	35/4	35/4	35/4		
p_{aver}	1/4	1/4	1/4	1/4	0.0750	0.2143

18.3 The Average Distribution in the Case when the Real, Unknown Distribution Is Unimodal

In the case of complete ignorance of the real distribution we can not exclude that it is a unimodal or bimodal, or n–modal. Therefore, for determining the average distribution, distributions of all modalities have to be found and used in calculations. However, in real problems an expert often knows the distribution modality in the specific problem under consideration. E.g. an expert may know that the distribution is unimodal (frequent case) though he/she may know nothing more about the distribution, and in particular: whether the distribution is symmetric or asymmetric, whether it is border or not border, etc. To determine the safe, average distribution
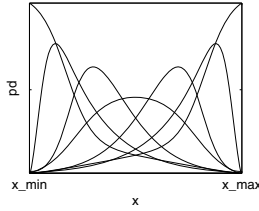
Fig. 18.4 Examples of unimodal distributions of various skewness with border–unimodal
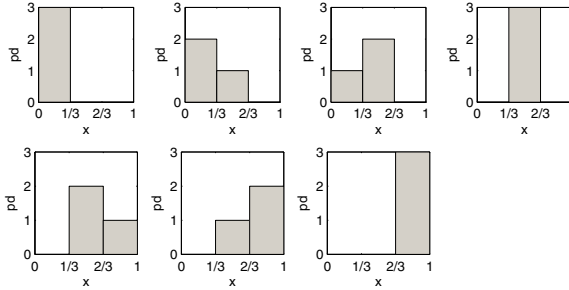
Fig. 18.5 7 unimodal distributions possible for granulation 1/3 of elementary events.

all possible unimodal distributions have to be found. It can be made with the GD–method. Fig. 18.4 shows examples of unimodal distributions of various asymmetry that were generated.

In the first step of the GD–method granulation 1/3 of x–events and of probability have been assumed and all 7 possible for this granulation, unimodal distributions were generated, Fig. 18.5.

It should be noticed that the number 7 of possible unimodal distributions is smaller than the number 10 of possible distributions in the case of full ignorance: compare Fig. 18.2 and Fig. 18.5. The average unimodal distribution of probability of all possible (7) distributions is shown in Fig. 18.6.
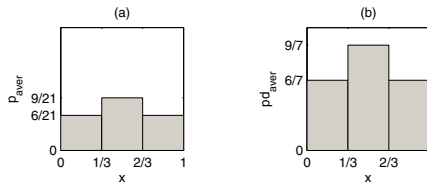
Fig. 18.6 The average distribution (histogram) of probability (p) achieved from 7 unimodal distributions possible for granulation 1/3 (a) and the corresponding average distribution of probability density (pd) (b).

The average, unimodal distribution for granulation 1/3 was calculated with formula as below, where j– number of the distribution and i– number of the sub–interval.

$$p_{aver}(i) = \frac{1}{7} \sum_{j=1}^{7} p_{ji}$$

The following results have been achieved: $p_{aver}(1) = 6/21$, $p_{aver}(2) = 9/21$, $p_{aver}(3) = 6/21$. This distribution, as the average, has the smallest square error (MSE) in relation to all 7 possible distributions.

In the next step of the GD–method the granulation has been decreased to 1/4. For this granulation 16 possible unimodal distributions were found. Because of the bigger number, the distributions were shown not in graphical but in table form, Table 18.3.

Table 18.3 16 unimodal distributions possible for granulation 1/4 and the resulting, average unimodal distribution.

Probability distribution	Interval of the variable x			
	$[0, 1/4)$	$[1/4, 2/4)$	$[2/4, 3/4)$	$[3/4, 1]$
j	$p_{j,1}$	$p_{j,2}$	$p_{j,3}$	$p_{j,4}$
1	4/4	0	0	0
2	3/4	1/4	0	0
3	2/4	1/4	1/4	0
4	0	4/4	0	0
5	0	3/4	1/4	0
6	1/4	3/4	0	0
7	0	2/4	1/4	1/4
8	1/4	2/4	1/4	0
9	0	0	4/4	0
10	0	0	3/4	1/4
11	0	1/4	3/4	0
12	1/4	1/4	2/4	0
13	0	1/4	2/4	1/4
14	0	0	0	4/4
15	0	0	1/4	3/4
16	0	1/4	1/4	2/4
$\sum_{j=1}^{20} p_{j,i}$	12/4	20/4	20/4	12/4
$p_{aver}(i)$	12/64	20/64	20/64	12/64

Fig. 18.7 presents the average unimodal distribution for granulation 1/4 in graphical form.

The comparison of Table 18.2 and Table 18.3 shows that in case of unimodal distributions their number 16 is smaller then number 35 of all possible distributions where all modalities of distributions were allowed. In the next steps of the GD–method granulation was still decreased and all possible unimodal distributions were found. The number of possible unimodal distributions with decreasing granulation of elementary events rapidly increase (increasing precision of the average distributions). For granulation 1/2 are 2 unimodal distributions, for 1/3 are 7 distr., for 1/4 16 distr., for 1/5 38 distr., for 1/6 76 distr., ... for granulation 1/25 are 812118 unimodal distributions. Fig. 18.8 shows graphs of the polynomial approximation of average distributions of probability density for various granulations from 1/4 to 1/25.
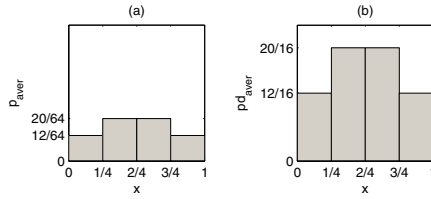
Fig. 18.7 The average distribution of 16 distributions (histograms) of probability (p) for granulation 1/4 (a) and the corresponding average distribution of probability density (pd) (b).
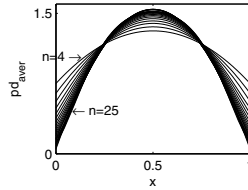
Fig. 18.8 The average, smoothed (polynomial approximation) distributions of probability density for decreasing granulations 1/4,1/5,...,1/25 that indicate existence of a limiting distribution for an infinitely small granulation.

It can be seen in Fig. 18.8 that differences between particular distributions are noticeable only for greater granulations (great value of $1/n$, were n is the number of x–subintervals). For small granulation the difference vanishes and for granulation 1/25 it is practically unnoticeable. Thus, it can be assumed that the average distribution for granulation 1/27 is in practice the limiting distribution of an infinitely small granulation $1/n$ where $n \to \infty$, Fig. 18.9.

The average unimodal distribution for granulation 1/27 was approximated with a polynomial of order 6:

$$pd_{aver}(x) = (96.5601x^6 - 289.6802x^5 + 329.861x^4 \\ -176.9218x^3 + 38.3903x^2 + 1.7905x + 0.0505)/0.9985$$

If variable x in a given problem is not limited to the interval $[0,1]$, the above formula can be transformed to the real interval by appropriate substitutions and then the achieved formula should be normalized so that the area equals 1.
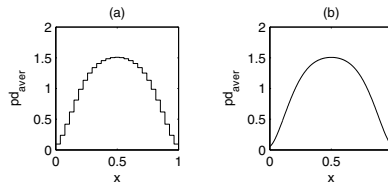
Fig. 18.9 The average distribution of 1.657.331 unimodal distributions possible for granulation 1/27 (a), which sufficiently precisely approximates the limiting distribution for infinitely small granulation and its smoothed version (polynomial approximation) (b).

Chapter 19
Interactive Visualization of a Product Search Space

Michał Ciesielczyk, Andrzej Szwabe, and Czesław Jędrzejek

Abstract. This chapter presents a system that uses visualization techniques of multidimensional vector spaces for product search, after processing a document corpus by means of LSI. Introduced algorithms are able to dynamically generate mappings of the most closely related items according to the 'thematic context' in the current session. A user navigation starts with required category of Internet shop product at the centre of the diagram, and then focuses on refining choices. Display capabilities include several novel adaptive features. Examples of the system applicability and future directions are also discussed.

19.1 Introduction

The goal of information visualization is to transform abstract information into a visual form that enhances cognition. Visualization is effective at providing insight about data for many tasks, mostly of an analytic nature, however, it has been somewhat less successful for search engines [1]. It is usually assumed that search engines enable retrieval of large quantities of information with a reasonable precision. However, the use of obtained results is, in most cases, limited to a serial inspection. Analyzing a long list of 'hits' can be overwhelming and, as a consequence, it is impossible to effectively view more than a small fraction of available information matching the specified query. That is why recent work in the area of information visualization has been intensively focused on visual analytics systems, which enable reasoning about contents of large collections of unstructured text, through a three-stage process. Initially, pre-processing methods are used, then statistical methods are employed, and finally, visualization techniques allow one to display results [2].

Michał Ciesielczyk, Andrzej Szwabe, and Czesław Jędrzejek
Institute of Control and Information Engineering, Poznan University of Technology, Poland
e-mail: {Michal.Ciesielczyk, Andrzej.Szwabe}@put.poznan.pl,
    Czeslaw.Jedrzejek@put.poznan.pl

In the process, methods for projecting multidimensional vectors onto a plane are frequently invoked to make sense out of large data sets. However, describing multiple features using only two values (e.g., values corresponding to an object position) is a complex task and in all practical cases some of the data is always neglected. Despite these obstacles, visualization technologies are being incorporated into an increasing number of general purpose information systems [2].

Visualization technologies are largely inspired by the fact that the human brain gains more information from images than from text [1]. To achieve this, two principles must be used: proximity - objects that are located near one another spatially belong to the same group, and similarity - objects sharing similar visual attributes belong to the same group.

This work concerns text visualization related to commercial product descriptions – each visualized object corresponding to a separate document. In our work, LSI (Latent Semantic Indexing) is used to distribute documents as points in n-dimensional space [3]. The major innovation of the proposed solution over previous works discussed in Section 19.2 is the methods used to contextually display objects placed in a two-dimensional vector space. Section 19.3 describes our requirements modelled after those proposed in literature for similar applications. The applied methodology is introduced in Section 19.4. Section 19.5 is devoted to presentation of the usage scenarios of the system. The main results of the work presented are discussed in Section 19.6.

19.2   Related Work

Existing visualization systems can be divided based on the following factors: an appearance of terms in one versus many documents, a vector-space analysis, a projection of multidimensional space into two-dimensional space, and a choice of visual attributes in two-dimensional space.

In most browsing structures, categories are usually manually defined and documents are assigned to those categories, either by hand or automatically. Here, we demonstrate operation of a system using a generic group of commercial products – 'lamps'. This category is subdivided into subcategories: standing lamps, desk lamps, halogen lamps or ceiling lamps. However, since aspects are not mutually exclusive, we do not, in general, use a faceted search.

Some of the most advanced systems, such as Starlight [4] and IN-SPIRE [5] are able to handle a massive amount of data and are capable of 2/3D visualization. These solutions are complete frameworks, with libraries using combinations of LSI (PCA), multidimensional scaling (MDS) [6] or a force-directed algorithm [7]. Other systems we investigated and made some comparisons with are WEBSOM [8], BellKor's Pragmatic Chaos team tool (TVland and MusicLand) [9], NIH Visual Browser [10] and the Titan Informatics Toolkit [11].

Out of systems that employ a starfield representation, the most notable is DocuBurst [12]. In this system a 'concept' (synset) is the root (centre) of the visualization. Occurrences of subconcepts of an "idea" appear as wedges in concentric circles. The size of the wedge indicates the number of times a given word from the Wordnet taxonomy appeared in the collection. Our approach is similar to

DocuBurst in that the concept of interest is the centre of visualization, but with a much different representation of visual attributes.

19.3 Requirements

To design details of the system, we adopted various literature suggestions [1, 13, 14, 15] for an effective, user oriented system. The results are presented below.

Table 19.1 Definitions of requirement features that have been incorporated into the system

Requirement	Definition
Adaptive (Ada.)	The display should be context-depended and provide a dynamic rearrangement of objects if needed.
Intuitive (Int.)	A visualization map should be easily understandable, supporting widely know navigation functions such as zooming and panning, and using a 'proximity means similarity' metaphor.
Adjustable (Adj.)	A user should be able to modify parameters of the visualization, show more details on demand and remove/filter unwanted objects.
Simple (Sim.)	Basic, but proven solutions should also be provided (e.g., text querying to find particular items).
Not overwhelming (NOv.)	Only currently relevant information should be presented on the screen.
Meaningful (Me.)	Visual variables such as a colour or a shape should be used to simultaneously present multiple features.
Web Available (WE.)	The application should be compatible with most popular browsers.

Table 19.2 Comparison of compliance of the systems referred to in Section 19.2 with regards to requirements presented in Table 19.1

	Ada.	Int.	Adj.	Sim.	NOv.	Me.	WE.
IN-SPIRE	-	+	+	+	+	+	-
WEBSOM	-	+	-	-	-	+/-	+
Starlight	+/-	+	+	+	+	+	+/-
TVLand	-	+	-	-	-	+/-	+
NIH Visual Browser	-	+	+/-	+	-	+	+
Titan	+	+	+	+	+	+	+/-

19.4 Methodology

Most of the techniques used for visual analytics of vector-space data consist of a few basic steps, shown in Fig. 19.1 [3, 16, 17].

Fig. 19.1 Visual analytics pipeline

The creation of multidimensional data from a text corpus requires conversion of each document into a vector, obtained by parsing text strings into tokens, which are then filtered and normalized. Filtering and normalization can include case regularization, spell checking, error correction, stop-words removal, stemming, lemmatization and multiword term grouping [16].

A vector-space model is an algebraic model for representing documents and terms from a text corpus [3]. This model provides conceptual simplicity – spatial proximity is used to represent semantic[1] proximity. Therefore, to compute similarity between objects, many different algorithms based on vector algebra may be used. In our work we use the cosine measure [3]:

$$\delta_{ab} = \frac{a \cdot b}{\|a\| \times \|b\|} \tag{19.1}$$

To address the issue of high number of dimensions several different methods of dimensionality reduction are employed. LSI is the most common technique used for this task [3]. In addition to LSI, several different alternatives to models of distributional semantics, which depend on the SVD, have been developed and can be found in the literature [18, 19]. We have used LSI because none of these other methods seems to give substantially better results.

A graphical layout is a critical aspect of all visualization applications, especially if they tend to present large multidimensional data sets in a meaningful way. However, even after dimensionality reduction operations, such as LSI, text vectors still have too many independent features to be properly visualized in the form of a 2–dimensional map. In order to enable humans to reasonably interpret similarity relationships, further dimensionality reduction is needed. The desired product is the most faithful graphical representation of conceptual similarities between documents, preserving only the most important characteristics.

We propose a taxonomy summarizing the most common visualization techniques shown in Fig. 19.2. A brief description of these methods may be found in [20].

---

[1] Within this chapter the term 'semantic' is understood in a way that follows widely-referenced statistical semantics studies [21] (focused on measuring the co-occurrence frequency of the context words near a given target word). Therefore, it is different from the way the word 'semantic' is understood by authors of ontology-oriented methods.

In our work, algorithms are usually invoked in a 'process loop':

1) Initially, the most relevant objects are selected for visualization purposes.
2) Then, coordinates of the localization of every object are computed.
3) Finally, user interactions are handled, animations are performed, and the session vector is updated.

A session vector (s) may be understood as the current 'view' on multidimensional space. It is computed dynamically, depending on user interactions with an object (v), according to:

$$\vec{s'} = \vec{s} + w(d - d')\vec{v}, w \in (0,1\rangle \qquad (19.2)$$

where, each 'drag&drop' action is represented by *(d-d')* and *w* is a weight-like parameter corresponding to system responsiveness. The most relevant vectors (along with the data associated with them) for the given session vector are chosen. The relevance may be evaluated using a cosine measure (19.1).
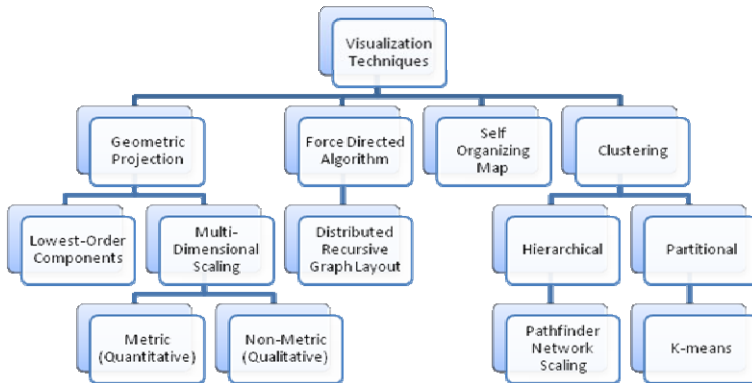
Fig. 19.2 Visualization techniques overview.

All of the introduced algorithms, implemented in the software design, place an object representing the vector of the session at the centre of the 2-dimensional (x,y) visualization map. The size of each object determines the length of its corresponding vector (which is dependent on the incidence of the object). The current relevance of each item or term (a) is represented by its distance from the centre (c), which is given by:

$$\forall_a : \delta_{ac} \approx d_{ac} = \sqrt{(x_c - x_a)^2 + (y_c - y_a)^2} \qquad (19.3)$$

After determining the distances mentioned above, angular distances (*ang*) are computed, so that, all objects are placed in the diagram to preserve (as close as possible) the similarities between them (19.4). A few algorithms were implemented to meet this problem. However, only two of them will be mentioned here:

the Energy Minimization Arrangement Algorithm and the Clustering Arrangement Algorithm (more details may be found in [20]).

$$\forall_{ab} : \delta_{ab} \approx d_{ab} = \sqrt{(x_b - x_a)^2 + (y_b - y_a)^2} \qquad (19.4)$$

The Energy Minimization Arrangement Algorithm is comparable to a force-directed algorithm.

1) Initially, a normalized similarity matrix is calculated.
2) New angular distance for several randomly chosen objects (a) are computed according to:

$$ang_a' = ang_a + \sum_i \delta_{ai} * d_{ai} \qquad (19.5)$$

3) Using an evaluation function, such as MSE (Mean Squared Error), the 'energy' of the whole system is measured. If the 'energy' is smaller than it was before, the objects are moved to their new locations.
4) Steps 2-3 are repeated until no system with lower energy level can be found.

The second algorithm presented here, the Clustering Arrangement Algorithm, is comparable to clustering techniques.

1) A tree containing all objects is constructed, using a bottom-up technique, so that the path between more similar objects is shorter.
2) An angular distance is assigned for each entity, in such a way that differences between angular distances of entities in the same cluster are the smallest.

Afterwards, objects may be placed in the diagram using simple mathematical conversions:

$$\forall_a : x_a = \cos(ang_a) * \delta_{ac}, y_a = \sin(ang_a) * \delta_{ac} \qquad (19.6)$$

Sometimes, after calculating objects' locations on the visualization, a few entities overlap and it is impossible to perceive objects that are placed below the others. As a solution, two dispersion methods are implemented, circular (19.7) and spiral (19.8).

$$x_a' = \cos(ang_a + cf) * d_{ac}, y_a' = \sin(ang_a + cf) * d_{ac} \qquad (19.7)$$

$$x_a' = x_a + \cos(sf) * sf, y_a' = y_a + \sin(sf) * sf \qquad (19.8)$$

where, *sf* and *cf* are determining the dispersion accuracy.

More details about implemented algorithms may be found in [20].

Finally, since most of the multidimensional data visualization systems have to deal with a huge number of objects, it is crucial to provide means for interactive and intuitive navigation. In our work, the display is interpretable as a map, and

functions such as zooming and panning are supported. Moreover, a dynamic rear-rangement of the objects to obtain more up-to-date information is also provided. Other analytically useful features include active searching, showing additional data and removing unwanted objects.

19.5   Presentation of the System Use

In this chapter three example use cases are introduced. The text data used in all of them has been extracted (with the full consent of the owner) from an e-shop selling office supplies. In our case, documents are product descriptions.

19.5.1   Use Case I – Searching an Item

The first use case shows an example of searching for an explicit item – a desk lamp. The user types in the name of the item he is looking for. In this case he writes in *lampa*[2] (a lamp). When the user enters the next letters of the search phrase, a list of suggestions appears automatically below the search box.

In the next phase, the vector of the centre entity is updated with the vector of the search phrase. New entities, related to *lampa* (lamp), are loaded in the client
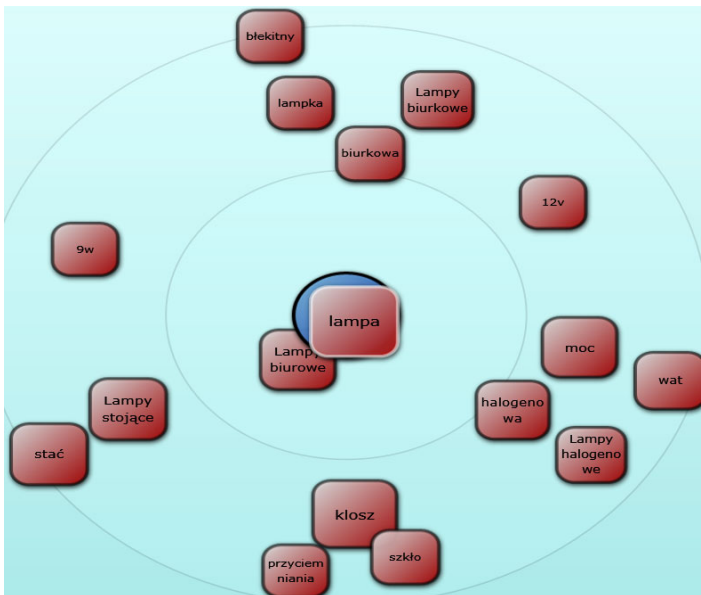
Fig. 19.3 Terms related to the term *lampa*.

[2] Since tests have been done using Polish e-shops, we present descriptions in figures in Polish (in the text, Polish terms are in italics).
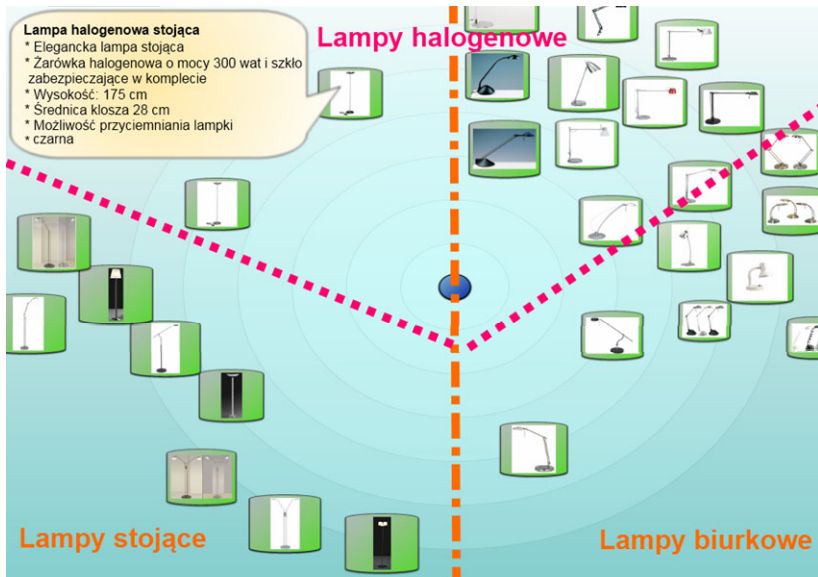
Fig. 19.4 Items related to the term *lampa*

application (Fig. 19.3). Similar terms are placed closer to each other. For example the noun *lampka* (small lamp), the descriptive adjective *biurkowa* (desk), and the term *Lampy biurkowe* (desk lamps) are closely related. Similarly, the verb *stać* (to stand) fits mostly the term *Lampy stojące* (floor lamps), and the adjective *halogenowa* (halogen) is closest to *Lampy halogenowe* (halogen lamps). Moreover, the most similar term to *lampa* (lamp) is *Lampy biurowe* (office lamps), because all lamps in the data set are intended for office use.

Next, a user wants to see items that match his/her search query. Fig. 19.4 presents the most similar items to the term *lampa*.

Subcategories are separated in such a way that 'floor lamps' (*Lampy stojące*) are on the left side of the diagram, while 'desk lamps' (*Lampy biurkowe*) are on the right side. Moreover, in the upper side of the display, all 'halogen lamps' (*Lampy halogenowe*) (both desk and floor) are located. Inset in Fig. 19.4 *Stojąca lampa halogenowa* (floor halogen lamp) is characterized by the following features:

- Elegant floor lamp
- Halogen bulb, 300 Watt, protection glass included
- Height 175 cm
- Lampshade diameter 20 cm
- Lamp dimmer
- Black.

This level of detail is typical of product information in an e-shop.

After analyzing some of the presented items, a user may choose one of them to find similar ones (by double clicking it). The user is also able to read a short description of each item, as shown above.

19.5.2   Use Case II – Navigating through a Vector Space

Another way to modify a session vector is to drag and drop proper elements over the visualization map. If the user is interested in items (or terms) similar to one of those on the current visualization map, he may move the vector closer to the centre. This will modify his point of view and may be found more elements that may be of interest.

Use case II introduces an example session showing how the user may discover interesting information, while interacting with the designed system.

Suppose a user moves an item, 'automatic pencil', to the centre of the diagram, in order to see more related items. Afterwards, new data are loaded, and as a consequence, new entities are added to the visualization map. As expected, newly added objects include more related items – e.g., three new automatic pencils, one package of pencils and a case with a pen and a automatic pencil. This technique may be also used to choose a set of desired attributes. The user is now able to focus on interesting parameters, or eliminate those which he does not want.

19.5.3   Use Case III – Comparing Data Entity Arrangement Algorithms

While discovering relations between objects in multidimensional space, applying various arrangement algorithms and parameters may help a user to understand the information provided in more detail. Apart from choosing an arrangement algorithm, the user is for example able to modify a number of visualized dimensions, turn on or off the dispersion algorithms or modify a vector similarity method.

Every implemented arrangement algorithm locates similar objects close to each other in a different manner. Nevertheless, every algorithm presented provides comprehensible results. In general, in our opinion, maps that are generated by the Clustering Arrangement Algorithm offer the most predictable outcome. Moreover, in contrast to the other algorithms, the method allows one to dynamically add new entities to the view without recalculating the whole visualization map. Additionally, adding new elements or modifying configuration parameters tends to maintain the current state of the visualization (preserves locations of all objects). This behaviour may be very helpful in some particular cases. Finally, applied additional dispersion algorithms help to detect an object which in case of overlap would be hard to see.

19.6   Conclusions and Future Work

Over the past several years a range of successful methods used to process and display multidimensional data sets have been developed and used in a number of

applications. However, the solutions proposed so far were oriented rather towards data analytics than visual search applications [1].

To the knowledge of the authors, the system presented in this chapter is the first visual search application in the area of e-commerce featuring The dynamic, contextual and interactive optimization with organization of the product information presented to a user. A user sees only the objects that are contextually relevant to his/her interactively specified point of view (represented by a session vector). The method is amenable to incorporation of faceted classification, which, however, has to be done on a level of documents by using controlled vocabularies.

One drawbacks of our method is difficulty of filtering with regards to rules, i.e. finding a product with a price below a certain level. To get around this problem one could define cheap, medium-priced or expensive products based on average product prices of given categories. However, for a user, it is more important whether he/she can afford a product, and this is difficult to describe by only taxonomic means.

The results need further verification. More user-testing is needed to understand user preferences. In particular, user preference for the Energy Minimization Arrangement Algorithm or the Clustering Arrangement Algorithm should be further studied. The Clustering Arrangement Algorithm enhances local similarity features, and tends to cause crowding, which necessitates measures to prevent object overlap in a display. Sometimes, a forced spread of crowded objects may distort their mutual relations. The Energy Minimization Arrangement Algorithm assigns equal weights to similarities, causing more uniform object distribution in a display.

A system should also support users in making their decisions faster. As a consequence, only those parameters that help to achieve this goal should be presented, instead of those which are only similar. This goal can be achieved by using different measures than the cosine (e.g., entropy-based) and by analysing users' behaviour from the new perspective of actions specific to a highly graphical and interactive user interface.

There is still no strong evidence that visualization significantly improves performance of search or product recommendation applications. This may be partly a result of the complexity of an effective, integrated processing of textual and behavioural data. Although the problem remains crucial in the context of an interactive and personalized product recommendation task, the proposed approach to multidimensional data visualisation provides some new options for supporting users in browsing rich product spaces. As a result, the interaction between an e-commerce application and a user may become more interactive, context-aware, and thus likely more intuitive than the interaction offered by currently used systems.

References

1. Hearst, M.A.: Search User Interfaces. Cambridge University Press, Cambridge (2009)
2. Risch, J., et al.: Text Visualization for Visual Text Analytics. In: Simoff, S.J., Böhlen, M.H., Mazeika, A. (eds.) Visual Data Mining. LNCS, vol. 4404, pp. 154–171. Springer, Heidelberg (2008)
3. Manning, C.D., Raghavan, P., Schütze, H.: An Introduction to Information Retrieval. Cambridge University Press, Cambridge (2008) ISBN:9780521865715
4. Pacific Northwest National Laboratory. PNNL: Starlight Information Visualization Technologies (May 2008), `http://starlight.pnl.gov/` (Cited: March 30, 2010)
5. Pacific Northwest National Laboratory. PNNL: InfoViz - Visualization (May 2008), `http://infoviz.pnl.gov/visualization.stm` (Cited: March 30, 2010)
6. Borg, I., Groenen, P.: Modern Multidimensional Scaling: theory and applications. Springer, New York (2005) ISBN:978-0-387-25150-9
7. Boyack, K.W., Wylie, B.N., Davidson, G.S.: Information Visualization, Human-Computer Interaction, and Cognitive Psychology: Domain Visualizations. In: Joint Conference on Digital Libraries (2001)
8. Lagus, K., Kaski, S., Kohonen, T.: Mining massive document collections by the WEBSOM method. Information Sciences: an International Journal 163, 135–156 (2004)
9. Gansner, E., et al.: Putting Recommendations on the Map – Visualizing. In: Proceedings of the third ACM conference on Recommender systems, pp. 345–348. ACM, New York (2009)
10. Herr II., B.W., et al.: The NIH Visual Browser: An Interactive Visualization of Biomedical Research. In: 13th International Conference Information Visualisation. IEEE, Los Alamitos (2009)
11. Wylie, B., Baumes, J.: A unified toolkit for information and scientific visualization. In: Proceedings of visualization and data analysis (2009)
12. Collins, C., Carpendale, S., Penn, G.: DocuBurst: Visualizing Document Content using Language Structure. In: Computer Graphics Forum, Proceedings of Eurographics/IEEE-VGTC Symposium on Visualization (EuroVis 2009), vol. 28(3), pp. 1039–1046 (2009)
13. Tintarev, N., Masthoff, J.: A Survey of Explanations in Recommender Systems. In: Workshop on Recommender Systems and Intelligent User Interfaces associated with ICDE 2007, pp. 801–810 (2007)
14. Zanker, M., Jessenitschnig, M.: Case-studies on exploiting explicit customer requirements in recommender systems. User Model. User-Adapt. Interact. 19(1-2), 133–166 (2009)
15. Thomas, J.J., Cook, K.A.: Illuminating the Path: The R&D Agenda for Visual Analytics. National Visualization and Analytics Center (2005)
16. Rajman, M., Vesely, M., Andrews, P.: Report of WG1. State of the Art, Evaluation and Recommendations regarding "Document Processing and Visualization Techniques". Lausanne: Work Package: 3 (Establishment and Operation of Working Groups), D-3.1 (2004)
17. Dunlavy, D.: CSRI Workshop on Combinatorial Algebraic Topology (CAT). Persistent homology for parameter sensitivity in large-scale text-analysis (informatics) graphs. s.l.: Sandia National Laboratories (2009)

18. Shi, Y., et al.: A Comparison of SVD, SVR, ADE and IRR for Latent Semantic Indexing. In: Communications in Computer and Information Science, vol. 35, pp. 266–274. Springer, Berlin (2009)
19. Sahlgren, M.: An Introduction to Random Indexing. In: Proceedings of the Methods and Applications of Semantic Indexing Workshop at the 7th International Conference on Terminology and Knowledge Engineering (2005)
20. Ciesielczyk, M.: Visualization of large multidimensional data sets. Master's thesis, Poznań (2010)
21. Turney, P.D.: Similarity of Semantic Relations. Computational Linguistics 32, 379–416 (2006)

Part IV
Web Systems and Network Technologies

Chapter 20
Adaptive User Profile in Web IR System with Heuristic-Based Acquisition of Significant Terms

Agnieszka Indyka-Piasecka

Abstract. This contribution presents a method of adaptive user profile creation, modification at the field of Web search systems by using query terms and weighted terms of retrieved documents. One of the essential parts of the method is heuristic-based significant terms selection from relevant documents, which relevance evaluation is assigned by the user. Created during retrieval cycles sub-profiles, represent user interests, and are used for user query modification. The experiments concerning adaptive user profile, as a personalization mechanism of Web search system, are presented and discussed.

20.1 Introduction

In today's World Wide Web reality, the common facts are: increasingly growing number of documents, high frequency of their modifications and, as consequence, the difficulty for users to find important and valuable information. These problems caused that much attention is paid to helping user in finding important information on nowadays Internet IR systems. Individual characteristic and user's needs are taken under consideration, what leads to system personalization. System personalization is usually achieved by introducing *user model* into information system. User model might include information about user's preferences and interests, attitudes and goals [3], knowledge and beliefs [6], personal characteristics [7], or history of user's interaction with a system [12]. User model is also called *user profile* in the domain of IR. The profile represents user's information needs, such as interests and preferences.

Agnieszka Indyka-Piasecka
Wrocław University of Technology, Institute of Informatics,
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
e-mail: `agnieszka.indyka@pwr.wroc.pl`

In the literature, few types of user profile applications can be distinguished.

- Profile can be used for ranking documents received from the IR system. Such ranking is usually created due to degree of similarity between user query and retrieved documents [8] [14].
- In Information Filtering (IF) system, profile became a query during the process of IF. Such profile represents user information need, relatively stable over a period of time [1].
- There are approaches where user profile was used for query expansion, based on explicit and implicit information obtained from the user [5].

The main issue, in the domain of user profile applied for IR, is a representation of user information needs and interests. Usually user interests are represented as a set of keywords or a n–dimensional vector of keywords, where every keyword's weight or position at the vector represents importance of keyword to representing user interests [8] [9]. The approaches with more sophisticated structures representing knowledge about user's preferences are also described:

- stereotypes – the set of characteristics of a prototype user of some class of users, sharing the same interests [4], or
- semantic net, which discriminates subject of user interests with underlining the main topic of interests [2].

The approaches to determine user profile can be also divided into few groups. The first group includes methods where user's interests are stayed explicitly by the user in specially prepared forms or during answering standard questions [4] [8], or by an example piece of text, written by the user [11]. The second group can be these approaches, where user profile is based on the analysis of terms frequency in user queries directed to IR system [8]. There is an assumption for these methods that the interest of the user, represented by a term, is higher as the term is more frequent in the user query. Analysis of the queries with the use of genetic algorithms [13], reinforcement learning [18] or semantic nets [2] are extensions to this approach. The third group of approaches includes methods, where the user evaluates retrieved documents. From documents assessed as interesting by the user, additional index terms, describing user interests, are added to the user profile [4] [8]. Most of research, concerning user modelling for IR, acquire user information needs expressed directly by the user of the IR system. The severe difficulties in expressing the real information needs by the user are frequently neglected. The fact that user usually does not know, which words he should use to formulate his interests to receive valuable documents from the IR system is ignored. We claim that user can express own preferences by retrieved documents relevance valuation.

20.2 Analysis of Web System Answer

Documents of the answer include documents returned by Web search engine. Not all documents found by the system pertains to the real user interest. We assume that when we ask a user to select among documents of the answer those which are

more interesting for him, we can come closer in this way to the picture of the do-main of his interest. In order to make the selection process realistic we assume that user is making only a binary choice: interesting vs non-interesting, i.e. to point out some documents without further assessment, e.g., of a kind of strength of how they pertain to his needs.

For the needs of IR on the Web, a domain of user interest is represented and identified by keywords, that originate and were extracted from the selected and relevant documents of the answer. An important element of the newly proposed representation of user interest, i.e. the user profile, is a analysis of answer docu-ments. The process is aimed to identify these terms which are representative for relevant documents of the answer and on the other hand are key terms from do-main of user interest. The ultimate goal of the answer documents analysis is iden-tification of vocabulary used in a certain knowledge domain which is the domain of user interest and the construction of the user interest representation on the basis of the relevant documents.

20.2.1 Relevant Document's Terms Weighting

Each term from a document indexed in a IR system is assigned the weight d_i ac-cording to the *tf–idf* schema [17]. The weights allow for identification of *index terms* which describe properly the document content. A term t_i can occur in more than one relevant document. Thus the weight wz_i' is influenced also by the weights d_i of this term in each of the selected relevant documents of an answer.

20.2.2 Significant Terms Selection from Relevant Documents

Key terms, which are important for the domain of user interest, are automatically extracted from the relevant documents selected by the user. These terms are stored in the appropriate subprofile and next used during modification of some of the fol-lowing questions asked by the user to the system.

Key terms chosen from the relevant documents selected are henceforth called *significant terms*. The proposed term selection method was inspired by the idea of discriminative terms [16] [19] and the cue validity factor [10]. Selected significant terms are introduced into a subprofile and next used to modify a user query.

The proposed method of significant terms selection is performed in several steps among which assignment of wz_i' weights to each term from relevant docu-ments is one of the steps of extracting a set of significant terms tz_i from the rele-vant documents. A joint application of two term selection criteria is important novelty of our approach. In this method, term weights from the relevant docu-ments together with the cue validity factor are combined to form a kind of two-step filter. A criterion obtain in this way is a *weighted sum*. As an effect of combining two discussed method of weighting terms from relevant documents, among all terms belonging to the relevant documents, only the terms pertaining to vocabulary used in the domain of user interest are selected. A weight of term–candidate for inclusion into the set of significant terms is calculated as following:

$$wz_i = \alpha \, wz_i{}' + \beta \, cv_i \qquad\qquad (20.1)$$

where values of factors α and β were chosen experimentally.

Selection of discriminative terms subset on the basis of a constant threshold is the technique most often reported in literature and applied in IR, e.g. it was used for the authoritative collection of documents. According to this technique only terms with weight above some constant, pre-defined threshold are added to the selected subgroup.

However collections of Web documents express substantially different properties. These collections are characterized by huge divergence of topics and large dynamics in time in relation to both: the number of terms and documents. In such collections, term significance as represented by term weight is changing along with modifications introduced into the collection, i.e. after adding new documents to the collection or altering already present ones. A typical method of discriminative terms identification, i.e. on the basis of the threshold expressed by the constant values, when applied to the Web collection would not produce an expected set of significant terms. That is why we propose thresholds in a form of multistage criterion for significant terms identification. Values of such thresholds are constant, but defined on the basis of functions taking into account the dynamic of term weights in the Web collection.

The process of significant terms selection is performed according to the following:

1. The user verifies the answer from Web search engine by selecting relevant documents (binary selection).
2. The weight d_i is calculated for each term occurring in the relevant document. The weights are calculated according to the *tf–idf* schema, where the number of documents in which a given term occurs is calculated in comparison to all documents in the collection (i.e. on the basis of a search engine index).
3. Each term t_i, which belongs to all relevant documents is assigned $wz_i{}'$ weight, which is the minimum over weights d_i of term t_i in the relevant documents (the way of calculating the weight $wz_i{}'$ was inspired by the research on grouping collection of documents made by Voorhees [19]). The term t_i is further considered as a potential significant term.
4. The *df* rule is next applied to above identified set of potential significant terms. Only these terms are considered for the further analysis which *df* value is between $df_{\min \text{ value}}$ and $df_{\max}$ value (the values were set experimentally).
5. The *cue validity* factor cv_i is calculated for each term t_i selected at the step 4.
6. The weight wz_i (1) is calculated for each t_i selected at the step 4.
7. The threshold τ, called *significance factor ι* is applied for further terms filtering and ranking. If the wz_i weight of the term t_i is higher than ι, we assume t_i to be a proper *significant term tz_i*.

Only terms occurring in all relevant documents can be included into a set of significant terms. The above criteria of terms selection have been aimed on finding only terms describing the domain of user interests and improving the number of relevant documents in the answer.

20.3 User Profile

The IR system is defined in this contribution by four elements: set of documents D, user profiles P, set of queries Q and set of terms in the dictionary T. There is retrieval function $\omega: Q \rightarrow 2^D$. Retrieval function returns the set of documents, which is the answer for the query q. The set $T = \{t_1, t_2, ..., t_n\}$ contains terms from documents, which have been indexed in a Web retrieval system (i.e. in a search engine). Set T is called dictionary. User profile is an object $p \in P$, where P is the set of all possible user's profiles. The profile p is described by the function π, which maps: user's query q, the set of retrieved relevant documents D_q and the previous user profile p_{m-1}, into a new user profile p_m. Thus, the profile is the following structure determined by the function π: $p_0 = \emptyset$, $p_m = \pi(q_m, D_q, p_{m-1})$. The function π is responsible for profile modifications. D_q is the set of relevant documents among the documents retrieved $D_q{}'$ for query q: $D_q{}' = \omega(q, D)$ and $D_q \subseteq D_q{}'$. For the user profile we define also the set of *user subprofiles SP* (see below).

The user profile is created on the basis of the information received from the user after *user's verification of the documents* retrieved by the system. During the verification the user points out these documents which he considers relevant for him.

The user query pattern s_j we call a Boolean statement, the same as the user query q: $s_j = r_1 \wedge r_2 \wedge r_3 \wedge \ldots \wedge r_n$, where r_i is a term: $r_i = t_i$, a negated term: $r_i = \neg t_i$ or logical one: $r_i = 1$ (for terms which does not appear at the question). The user query pattern s_j indicates the subprofile and is connected with only one subprofile.

The user subprofile $sp \in SP$ we call a n-dimensional vector of weights of terms from the relevant documents: $sp_j^{(k)} = (w_{j,1}^{(k)}, w_{j,2}^{(k)}, w_{j,3}^{(k)}, ..., w_{j,n}^{(k)})$, where SP is the set of subprofiles, n – the terms in the dictionary T: $n = |T|$, $w_{j,i}^{(k)}$ – the weight of the significant term tz_i in subprofile after the k–th subprofile modification.

The position of weight $w_{ji}^{(k)}$ in the subprofile (its co-ordinate in the vector of the subprofile) indicates the significant term $tz_i \in T$. The terms from dictionary T are an indexing terms at a Web search system, that index documents retrieved for the query q and these terms belong to these relevant documents.

The user profile $p \in P$ we define as the following set of pairs:

$$p = \left\{ \langle s_1, sp_1 \rangle, \atop \langle s_2, sp_2 \rangle, \atop ..., \atop \langle s_l, sp_l \rangle \right\} \tag{20.2}$$

where: s_j – a user query pattern, sp_j – a user subprofile (an user query pattern indicates one user subprofile univocally).

The weight of significant term tz_i in subprofile is calculated according to the formula:

$$w_{j,i}^{(k)} = \frac{1}{k}((k-1)w_{j,i}^{(k-1)} + wz_i^{(k)})$$ (20.3)

where: k – the number of retrievals of documents made so far for this subprofile, i – the index of the term in the dictionary T, j – the index of a subprofile, $w_{j,i}^{(k)}$ – the weight of the significant term tz_i in the subprofile after the k–th modification of the subprofile[1], which is indicated by the pattern s_j (i.e. after the k–th document retrieval with the use of this subprofile), $wz_i^{(k)}$ – the weight of the significant term tz_i in the k–th selection of these terms.

20.4 Modification of User Profile

The adaptive user profile expresses the translation between the terminology used by the user and the terminology accepted in some field of knowledge. This translation describes the meaning of the words used by the user in a context fixed by relevant documents and it is described by assigning to the user's query pattern s_j a subprofile ('translation') created during the process of significant terms tz_i selection from relevant documents of an answer. We assume the following designations: q – the user query, $D_q^{'}$ – the set of documents retrieved for the user query q, $D_q^{'} \subseteq D$, D_q – the set of documents pointed by the user as relevant documents among the documents retrieved for user query q, $D_q \subseteq D_q^{'}$.

As it was described above, the user profile p_m is the representation of the user query q, the set of relevant documents D_q and the previous (former) user profile p_{m-1}. After every retrieval and verification of documents made by the user, the profile is modified. The modification is performed according to the following procedure: $p_0 = \varnothing$, $p_m = \pi(q_m, D_q, p_{m-1})$ where p_0 – the initial profile, this profile is empty, p_m – the profile after m–times the user has asked different queries and after each retrieval the analysis of relevant documents was made.

Traditionally, a user profile is represented by one n–dimension vector of terms describing user interests. User interests change, and so should the profile. Usually changes of a profile are achieved by modifications of weights of the terms in the vector. After appearance of queries from various domains, modifications made for this profile can lead to an unpredictable state of the profile. By the unpredictable state we mean a disproportional increase of the weights of some terms in the vector representing the profile, that could not be connected with an increase of user interests in the domain represented by these terms. The weights of terms can grow, because of high frequency of these terms in the whole collection of documents, regardless of the domain of actual retrieval.

The representation of a profile as one vector could also cause ambiguity during the use of this profile for query modification. At certain moment a query refers only to one domain of user's interests. To use the profile mentioned above we need a mechanism of choosing from the vector of terms representing various user'

---

[1] The weight, called *cue validity*, is calculated according to the frequency of term tz_i in relevant documents retrieved by the system in k–th retrieval and the frequency of this term in whole documents of the collection.

interests only these terms that are related to the domain of a current query. To obtain this information, usually knowledge about relationship between terms from a query and a profile, and between terms in the profile is needed. In literature, this information is obtained from a co-occurrence matrix created for a collection of documents [15] or from a semantic net [9]. One of disadvantages of presented approaches is that the two structures, namely a user profile and a structure representing term dependencies, should be maintained and managed for each user and also that creating the structure representing term relationships is difficult for a so diverging and frequently changing environment as Internet.

There are no such problems for the user profile p created as described in this contribution. After each retrieval, only the weighs of the terms from the user subprofile identified by pattern s_j (identical to user's query) are modified, not weighs of all terms in the user profile. Similarly, when the profile is used to modify user's query, the direct translation between the current user query q and the significant terms from the domain associated with the query is used. In the user profile p, a kind of mapping existing between a single user query pattern s_j and a single subprofile sp_j represents this translation.

In the IR system the user profile is created during a period of time – during a sequence of retrievals. There could appear a problem how many subprofiles should be kept in the user profile. We have decided that only subprofiles that are frequently used for query modifications should not be deleted. If a subprofile is frequently used, it is important for representing user's interests.

The modification of the user subprofile sp is made always when from the set of relevant documents pointed out from retrieved documents by the user, some significant terms tz_i are determined. The $w_{j,i}^{(k)}$ weights are modified only for these terms and only in one appropriate subprofile identified by the user query pattern s_j. The term tz_i weight is calculated according to (3). During each retrieval cycle the modification takes place only in one subprofile and for all significant terms tz_i obtain during the k–th selection of these significant terms from the relevant documents retrieved for the query q, which was asked k–th time. If the modification took place for significant terms tz_i for every subprofile in whole user profile, it would cause disfigurement of significant terms importance for single question.

20.5 Experiments

The adaptive user profile was implemented as a part of a Web search engine - the *Profiler* system. The user profile is used as a mechanism of retrieval personalisation, i.e. by user query modifications. The modification of user query takes place as a result of user interaction with a search engine (i.e. a verification of documents). After verification, the system automatically asks the modified query to the search engine and presents the new answer to the user.

The experiments were forked into two directions. In the preliminary case, the aim was to establish all parameters (i.e. thresholds) for the *Profiler*. In the second case – to verify the usefulness of the adaptive profile. So the retrieval simulations were arranged in a testing environment.

The aim of second case was also to prove that for any field of knowledge the profile converges. It means that starting from any randomly generated query q, which consist of terms from dictionary T, the proposed analysis of relevant (to the user) documents, the selection method of important terms and methods of profile creation and the query modification will lead to the greater number of relevant documents retrieved in the next retrieval process. What is more, the search engine answer in the next retrieval process is less numerous. In the test environment testing sets of documents were established by the group of 13 persons - users of Web search engine. Relevant documents were identified by the above mentioned persons. Three types of relevant documents sets were used: *dense sets*, *loose* and *mixed sets*.

The set of 50 randomly generated queries, with terms from dictionary T, were created automatically. Each random query was asked to the search engine. The retrieval process was run. If in the answer there were relevant documents, the randomly generated query was modified – the significant terms replaced the preliminary query. The modified query was automatically asked to the search engine and the next relevant documents were found from testing sets of documents.

Each stage of above described cyclic process we called *iteration*. Iterations were repeated until all relevant documents from the set were found (for the dense sets of relevant documents) or no change in number of relevant documents was observed (for the loose and mixed sets of relevant documents). The experiments were repeated separately for each type of documents sets.

For every random query in experiment: the number of all retrieved documents D'_q, the percentage of the relevant retrieved documents *%DR*, and the precision $Dokl_m$ (for the first m=10, 20, 30 documents in the answer) at every iteration were calculated. These were retrieval improvement measures for the proposed method.

The retrievals made during experiments for dense sets of relevant documents confirmed that for most of the modified queries the retrieval results were better in comparison to the preliminary query. For more then 82% of preliminary queries, the *Dokl* and the *%DR* were increasing with each iteration of query modification (Table 20.1). The number of all retrieved documents diminishes with every iteration.

The retrievals made during experiments for loose sets of relevant documents showed that a method of adaptive profile creation, modification and application assure that all modified queries (from one starting query) always focus on the

Table 20.1 Improvement measures for retrievals made during experiments.

	Percent of modified queries			Relevant retrieved documents %DR			
	improvement of retrieval *Dokl*	no improvement of retrieval *Dokl*	partially loss of improvement of retrieval *Dokl*	75-100%	50-75%	25-50%	0-25 %
dense sets	82 %	12 %	6 %	54 %	10 %	18 %	18 %
loose sets	67 %	12 %	21 %	0 %	3 %	68 %	29 %
mixed sets	58 %	18 %	24 %	6 %	36 %	29 %	29 %

same field of user interests. Each next modified query will not move out to other field of knowledge. In experiments for loose sets of relevant documents, for more then 67% of preliminary queries, the above measures rose with each iteration of query modification. The number of all retrieved documents diminishes as well. For the rest of the questions at that part of experiments, the measured parameters were worse, because only one document was found by modified query. The reason was in structure of the loose sets (no similar documents were found).

20.6   Conclusions and Future Work

Adaptive user profile presented in this contribution is a new approach to the representation of user's interests and preferences. The profile includes both interests given explicitly by the user as a query and also preferences expressed by the valuation of relevance of retrieved documents. The additional task of this profile is to express translation between terminology used by user and terminology accepted in some field of knowledge. This translation is supposed to describe the meaning of words used by user (in user query) in context fixed by the retrieved documents (i.e. user subprofile). Created procedures, essential for building, modifying and using the user profile and subprofiles, were presented. The experiments confirmed, that during retrieval process in Web retrieval system, user words in the query are replaced with the adequate indexing terms from subprofile. It makes the modified query more precise. A user of Web search engine receives support during query formulation, even in the cost of 'hidden iterations' of searching process. The query is modified in such a way that, for the most cases during next retrievals the user receives the smaller set of retrieved documents and the one that consists of more relevant documents. In the future, some experiments need to be done with a bigger group of WWW users, who will retrieve and assess the documents from the Web.

References

1. Ambrosini, L., Cirillo, V., Micarelli, A.: A Hybrid Architecture for User-Adapted Information Filtering on the World Wide Web. In: Proc. of the 6th International Int. Conf. on User Modelling, UM 1997, pp. 59–62. Springer, Heidelberg (1997)
2. Asnicar, F., Tasso, C.: ifWeb: a Prototype of User Model-Based Intelligent Agent for Document Filtering and Navigation in the World Wide Web. In: Proc. of the Workshop Adaptive Systems and User Modeling on the World Wide Web. 6th Int. Conf. on User Modelling, UM 1997. Springer, Heidelberg (1997)
3. Billsus, D., Pazzani, M.: A Hybrid User Model for News Story Classification. In: Proc. of the 7th Int. Conf. on User Modeling, UM 1999, Banff, Canada, pp. 99–108 (1999)
4. Benaki, E., Karkaletsis, A., Spyropoulos, D.: User Modeling in WWW: the UMIE Prototype. In: Proc. of the 6th Int. Conf. on User Modelling UM 1997, pp. 55–58. Springer, Heidelberg (1997)

5. Bhatia, S.J.: Selection of Search Terms Based on User Profile, Comm. of the ACM (1992)
6. Bull, S.: See Yourself Write: A Simple Student Model to Make Students Think. In: Proc. of the 6th Int. Conf. on User Modelling, UM 1997, pp. 315–326. Springer, Heidelberg (1997)
7. Collins, J.A., Greer, J.E., Kumar, V.S., Mccalla, G.I., Meagher, P., Tkatch, R.: Inspectable User Models for Just–In Time Workplace Training. In: Proc. of the 6th Int. Conf. on User Modelling, UM 1997, pp. 327–338. Springer, Heidelberg (1997)
8. Daniłowicz, C.: Modelling of user preferences and needs in Boolean retrieval systems. Information Processing and Management 30(3), 363–378 (1994)
9. Davies, N.J., Weeks, R., Revett, M.C.: Information Agents for World Wide Web. In: Nwana, H.S., Azarni, N. (eds.) Software Agents and Soft Computing: Towards Enhancing Machine Intelligence. LNCS (LNAI), vol. 1198, pp. 81–99. Springer, Heidelberg (1997)
10. Goldberg, J.L.: CDM: An Approach to Learning in Text Categorization. InternationalJournal on Artificial Intelligence Tools 5(1&2), 229–253 (1996)
11. Indyka-Piasecka, A., Piasecki, M.: Adaptive Translation between User's Vocabulary and Internet Queries. In: Proc. of the IIS IPWM 2003, pp. 149–157. Springer, Heidelberg (2003)
12. Indyka-Piasecka, A., Daniłowicz, C.: Dynamic User Profiles Based on Boolean Formulas. In: Orchard, R., et al. (eds.) IEA/AIE 2004. LNCS (LNAI), vol. 3029, pp. 779–787. Springer, Heidelberg (2004)
13. Jeapes, B.: Neural Intelligent Agents. Online & CDROM Rev. 20(5), 260–262 (1996)
14. Maglio, P.P., Barrett, R.: How to Build Modeling Agents to Support Web Searchers. In: Proc. of the 6th Int. Conf. on User Modelling, UM 1997, pp. 5–16. Springer, Heidelberg (1997)
15. Moukas, A., Zachatia, G.: Evolving a Multi-agent Information Filtering Solution in Amalthaea. In: Proc. of the Conference on Agents, Agents 1997. ACM Press, New York (1997)
16. Qiu, Y.: Automatic Query Expansion Based on a Similarity Thesaurus, PhD. Thesis (1996)
17. Salton, G., Bukley, C.: Term-Weighting Approaches in Automatic Text Retrieval. Information Processing & Management 24(5), 513–523 (1988)
18. Seo, Y.W., Zhang, B.T.: A Reinforcement Learning Agent for Personalised Information Filtering. In: Proc. of the 2000 Int. Conf. on the Intellig. User Interfaces, pp. 248–251. ACM, New York (2000)
19. Voorhees, E.M.: Implementing Agglomerative Hierarchic Clustering Algorithms for Use in Document Retrieval. Information Processing & Management 22(6), str. 465–476

Chapter 21
Vertical Search Strategy in Federated Environment

Jolanta Mizera-Pietraszko and Aleksander Zgrzywa

Abstract. Search engines are found the most powerful tools of information systems. However, in case of multilingual systems, they are rather oriented towards shallow techniques. This paper outlines our study into refining query language on the comparison basis of some search engines that utilize different translation models in order to propose a novel search strategy which significantly imposes the Web traffic optimization, in particular of the Deep, or Hidden Web. The framework proposed reveals inadequacies of the trans lingual information processing and provides the benefits for the user interacting with the information system. Our analysis of the context and the syntactic structure enables the user to retrieve sentences in its natural form in at least two languages.

21.1 Introduction

Unquestionably, at the time of the Internet popularity, the users more and more rely on the information from the web. In this respect, search engines are undoubtedly the most efficient tools of the information systems.

Typically, in federated search environment the data models utilized by the search engines differ from those processing local databases either in syntax, or the target language so that the system in unable to index the data efficiently. Irrespective of it, heterogeneous information sources in many formats constitute multilingual collections of billions of documents causing the system efficacy to drop significantly [4]. WebWorldScience, or Google Scholar are some of the systems that perform federated search.

In translingual approaches, predominantly, the source text is translated into all languages of the target database. At that stage, the monolingual searching process is continued for each document as an information entity, to create a ranking list.

Jolanta Mizera-Pietraszko and Aleksander Zgrzywa
Department of Information Systems, Institute of Informatics,
Wrocław University of Technology, Poland
e-mail: `Jolanta.Mizera-Pietraszko@pwr.wroc.pl`,
         `Aleksander.Zgrzywa@pwr.wroc.pl`

Subsequently, the number of the ranking lists is equal to the number of the database languages. Such lists are merged into a resulting list based on the relevance system score [1].

Likewise, a collection of the documents being translations of a source document into all the database languages can be combined with all the other such collections consisting of the documents included as a query set to be indexed against the multilingual database. Following the process, monolingual retrieval is carried out for each of these languages [2].

Also, both a query and the documents can be transformed into language-independent representation of an auxiliary language, also called Interlingua [6].

Another option commonly used is the monolingual, the source language search pre-performance, in which all the documents are translated from their own languages into one language being a source one for starting the searching process [3].

Our study starts from the analysis of the query language profile in some search engines working on different language models to develop a novel retrieval technology that optimizes a web traffic. The methodology of the research relies on identification of the factors that limit the searching process and it is oriented towards the user needs, even the one having a fuzzy idea about the query formulation.

In order to benefit from multilingual searching efficiently, a user is expected to operate the system in a way that allows to keep control over it. That is, he or she should understand in detail how the system works by expressing the need in a form of a query that will result with the precise information required. Nonetheless, to achieve it, the user should have some expertise in:

- the query modes accepted by the particular system,
- the databases searched,
- the maximal query length that can be entered, and
- how the system works in multilingual environment.

Naturally, by submitting queries, the user systematically broadens the knowledge in the field of his or her interest, which results in entering the longer and more and more precise subsequent queries. In our recent research experiments, we observed that in this kind of interaction with the system,a proportion between the relevant responses to the all the items displayed in the ranking list systematically increases. However, on exceeding a certain number of the query words, the system produces no results.

21.1.1   Methodology and a Draft of the Research Overall Framework

The analysis of the factors included in the PageRank algorithm indicates that positioning query words differentiates the ranking list of the system results tremendously. For instance, Google does not index pages larger than 101 KB whereas entering a query consisting of more than thirty-two words gives no results [5] compared to ten-word query limit in 2004.

Our research provides a framework for developing a method that the user is able to benefit from the facilities of the searching process. From this perspective

we study the precision and recall of some multilingual search engines as a result of our queries submitted in the following modes (Query Models):

- Keywords –set of unstructured query words
- Field restricted search – we consider an URL or a title field alternatively
- Boolean operators – most systems have inbuilt operator OR as a standard
- Phrase (called also Exact) search – the results contain the full query phrase
- Proximity search – in the ranking list, the query words are separated with a given by the user number of words
- Fuzzy search – the user has a fuzzy idea about the information need, so that some search systems produce the resulting documents with the synonyms of the query words
- Regular search – the user specifies a syntax that is matched with the document content
- Wildcard search – an asterisk enriches the query words with their variations

Likewise, some of the modes utilized by information systems perform retrieval models like: Boolean, Vector, Probabilistic, Language, Fuzzy model, or Latent Semantic Indexing (LSI).

For example, Google as a semantic search engine, is one of those that employ LSI, which empowers searching with indexing synonyms of the query words. We test the technique using the Canadian Hanzard English and French parallel corpora.

Prepared is a set of English Ad Hoc tasks for which we profile our queries by refining them. In addition, we consider the search modes so as to compare the resulting lists of documents in French. For each class we analyze the number of the relevant responses and the total number of the documents in the class. Moreover, for each Ad Hoc task, we define the relevance criteria as presented in Figure 21.1.

```
T004
     UNESCO World Heritage Sites

     Give the names and/or location of places that have
been designated as UNESCO World Heritage Sites of out-
standing beauty, or importance.
     Relevant documents must mention the name and/or
geographical location of monuments, cities or places
that have been officially designated by UNESCO as World
Heritage sites of outstanding beauty or importance.
Discussions on potential or candidate sites are not re-
levant. It must be clear from the document that the
official UNESCO status is involved.
```

Fig. 21.1 An sample Ad-Hoc task from the set

The query is submitted to the selected systems starting from the one-word query to the full form as in the task. Any changes in proportion between the relevant to total responses to each query class are reported. For every other search strategy the same procedure is carried out. Finally, we repeat the process with another search system.

So we attempt to specify the classification of the multilingual information systems, analyze the changes in proportion between the relevant and non-relevant search results in federated environment, a system precision based on different language models and finally compare the popular cross-language searching strategies utilized by some search systems.

21.1.2   Related Work

Federated environment constitutes a formidable obstacle for the searching process. As a result most of the unlinked pages, very large databases, variety of information resources based on unstructured models and in particular, the multilingual documents are not indexed by the conventional systems. Metadata search is a kind of solution as it performs a deeper crawling. Still, it does not match most of the multilingual documents [6].

An iterative Google algorithm of link analysis produces a ranking list of the search results [7] using a transition matrix and a random walk through the linked pages. Therefore, any alignment of the query words impedes the system responsiveness.

LSI technique improves the retrieval of the multilingual information without implementation of the translation component. A document is processed as a set of the words of the semantic space and created from the source text in the form of a matrix with the values representing a co-occurrence of the query word in this particular document. A measure of the position in the ranking list is computed based on the approximated coverage of the document words' semantic synonymy providing that the number of these words is close to the number of the query words [8].

From the perspective of the worldwide research, comparative analysis of the current indexing techniques in the federated environment, verifies a popular viewpoint of the preferred by the user human-computer interaction methods e.g. entering the queries in the form of the keywords only, or the simple phrases. Specifically, it supports a role of the user expertise about the system as well as about the optimization methodology of the multilingual retrieval process.

21.2   Semantic Mutual Resemblance of Bi-Texts

Resolving ambiguity requires consideration of synonymy, in which a concept is expressed in many forms, as well as polysemy, when different meanings apply to the same term. Assuming that the same words create a limited number of grammatical structures, specifically the latent ones, it is possible to avoid machine translation, language resources, or any human interaction with the system to achieve impressive retrieval results especially in relation to the false positive

(scored as irrelevant to the query) and false negative system responses, representing the numerous missing documents.

Latent Semantic Technique when used for a purpose of analysis called LSA or sometimes Probabilistic (PLSA) [10], otherwise as Latent Semantic Indexing (LSI), captures semantic resemblance in the target language of the conceptual patterns across a database of documents and relates the statistically computed results to each of these documents. In a simple word, it refers the structural patterns within a document to the whole of the collection [11].

Before implementation of this technique to Cross-Language Information Retrieval it used to be exploit to simulation of a human knowledge about the language usage, or estimation of the semantic association degree and lexical similarity.

Google is one of those to have adopted this approach. Created are both query words and documents' spaces that represent query word-document matrix [12].

Let $D=\{d_1, d_2, \ldots, d_n\}$ be a set of documents while $Q=\{q_1, q_2, \ldots q_m\}$, a set of the query words. F forms a matrix of the query term frequency in a document, so that $F=(f_{ij})_{n \times m}$ consists of the term frequency in the rows and the document number in the columns [9]. Each query word is matched against the F matrix. Association between the terms and concepts is computed as a product of the three matrices Singular Value Decomposition (SVD) of F:

- T - an orthogonal term concept-vector matrix of left singular vectors
- R - a singular diagonal values matrix of square roots of document number
- C – an orthogonal concept-document matrix of right singular vector

$$F=T\times R\times C^T$$

The cosine measure between the text terms, document-query, document-document is a normalized dot product of the predefined space elements:

$$\cos(t_1, t_2)=\frac{\sum_{i=1}^{n}(t_1\times t_2)}{\sqrt{\sum_{i=1}^{n}(t_1)^2}\times\sqrt{\sum_{i=1}^{n}(t_2)^2}}$$

where t_1 and t_2 represent the query word-document, text term-text term etc. respectively. The cosine semantic comparison value $\omega^{(1,2)}$ ranges between -1 and 1.

$$-1\leq \omega^{(1,2)}=\cos(t_1, t_2) \leq 1$$

Provided is an example of two sentences, each of which is in a different language:

English Text:
*Latent Semantic **Indexing** relies on a SVD to determine a resemblance between the **bi-texts**.*

French Text:
*Ecrivez **le mot** **recherché** dans l'espace prévu en haut de la page.*

Fig. 21.2 Semantic comparison of the English to French texts

It seems these sentences have nothing in common whether the meaning, sentence structure, or a language is considered. However, the resulting resemblance as a cosine value of these sentences equals to $\omega^{(1,2)}$=-0.6 for any matrix type and all the combinations between the document and a term spaces presented above, except for the document-to-document space when it is $\omega^{(1,2)}$=0.35, due to the English words *Indexing* and *bi-texts* semantic similarity to the French words *le mot* and *recherché*.

Table 21.1 LSI bi-text retrieval

Score	Title of the matching documents	Score	Matching terms
86	Ms. Margaret Mitchell (Vancouver East): Mr. Speaker, I am	100	parliament
84	Mlle Aideen Nicholson (Trinity): Monsieur le President, le	94	strike
84	Miss Aideen Nicholson (Trinity): Mr. Speaker, Bill C-45, an	94	accorde
84	This demonstrates that there has been some change, but the	93	facons
83	I think you are aware, Mr. Speaker, that we in the New	92	parlement

Table 21.1 shows the best matching for the query word Parliament in the Term-by-Document space of the Canadian Hanzard collection. The first score represents the relevance of the term to these documents, while the right-hand score stands for the best matching document terms to the query word Parliament.

Figure 21.3 presents the first document content with the query word in bold to indicate the context of the word usage. Thus, the highest score was given for the term. Accordingly, the lowest score received the document with the best French equivalent Parliament of the query word.

```
>|Ms. Margaret Mitchell (Vancouver East):| Mr. Speaker,
I am very pleased to speak| |against this Bill, which I
consider to be a real disgrace to Members of Parliament
and certainly to the Government. It does not deal with
the real and urgent need for true collective bargaining
rights for employees of Parliament, including employees
of Members of Parliament. At this time, we are speaking
to a motion that would delay this Bill, because we feel
that this Bill should be delayed -- in fact it should be
withdrawn -- because it restricts and does not enhance
the collective bargaining rights of parliamentary em-
ployees. >|
```

Fig. 21.3 The document content that the best matches with the query word Parliament

Though no translation was performed, three out of five matching terms are the French words being the synonyms of the English query word. In this light, the LSI technique provides a shining example of multilingual retrieval that on the contrary to the cross-lingual one, does not exploit translation.

However, the fifth ranking position of a document as the first one with the best French equivalent validates the standpoint about the lack of the translation component that affects the retrieval process.

21.2.1 Multilingual Information Technologies

A great challenge for information technology is to surmount language boundaries, especially for the users speaking their mother tongue only since they are not capable of formulating a query in other languages. Accessing an information in a foreign language requires for them translation of documents, Web pages, or image annotations. Still, some ASCII encoded or HTML with metadata fields digital documents or eventually, spoken word let alone, perhaps in the future, sign languages, cannot be accessible without character decoding technologies [4].

One of the technologies called an inverted index relies on identifying the query terms in the memory table so that all the documents containing it, are retrieved. Automatic compression of the posting files stored on the hard disc turned out to be the most cost effective for matching the most commonly used query terms. This way, stop word removal could be abandoned for the monolingual search.

Applications that implement translation component deploy either the automatic selection of the translated query term in context, or aggregate the most frequently occurring words, or alternatively, consider the term weight for each of the query words.
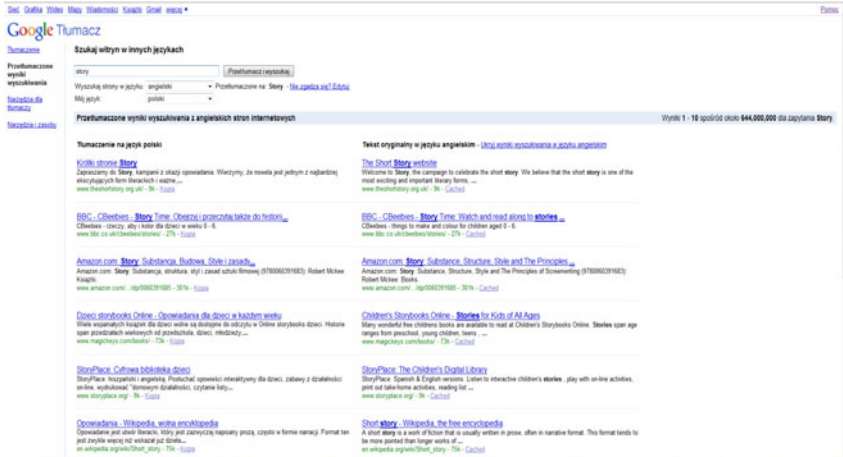
Fig. 21.4 Google Cross-Language Search with a query word "story" in context

As presented in Fig. 21.4, Google offers an interactive translation by listing at first, a number of the English equivalents and synonyms for the word "opowiadanie" to be retrieved and displayed in two languages simultaneously.

Ranking criteria are based on the number of the incoming and outgoing links clicked by a surfer while searching for information [1]. A frequency of visits to the pages is computed as changeable, assuming that the number of the outlinks equals to the inlinks so that the probability of a visit depends upon the state at a time given [13]. The PageRank at step i is computed as a total probability of a number of visits that a surfer goes through until reaching an information.

Given a document d_i a PageRank value for the documents linking it at stage i is

$$P_i(d_i) = (1-\propto)P_0(d_i)+\propto (\sum_{i=1}^{n} P_{i-1}(d_i) \times P_{i+1}(d_{i+1}|d_i))$$

where:

$\sum_{i=1}^{n} P_i (d_i)=1$ the probabilities of all the other pages that link to page d_i equal one
$P_{i+1}(d_{i+1}|d_i)$ – a probability that a surfer reaches page d_{i+1} walking randomly through d_i
α - probability that the surfer clicks a link to this page
$1-\alpha$ – probability that the surfer will type the URL to this page

PageRank emphasizes an importance of the pages and documents that are pointed to by the surfer the most often. Therefore, a position of the relevant document on the ranking list changes respectively as long as the Internet is in use [6].

Yet, matching query words with the documents is performed in a sequence. Consequently, a query word alignment determines the relevance of the particular document.

21.2.2   Multilingual Federated Environment

Granted by the European Commission a French Quero project supported development of an Exalead multilingual search engine that features a unique query refinement in addition to browsing federated environment [14].

Text modules include identification of 54 languages, faceted search, related terms or queries, spell checker, phonetic lemmatization, stop words remover, named entity recognizer, ontology matcher, categorizer, cluster selector, HTML extractor, sentiment analyzer. Synonyms are retrieved in one, or two directions that is a result can be entered as a query word to retrieve the original word again.

This search engine supports all the query models and, on the right hand panel, a user can select the data format (file type), information form, related terms to the query, a percent of the language responses out of total and the countries from which the results are. The processing technology relies on transcribing, indexing and translating the document content.
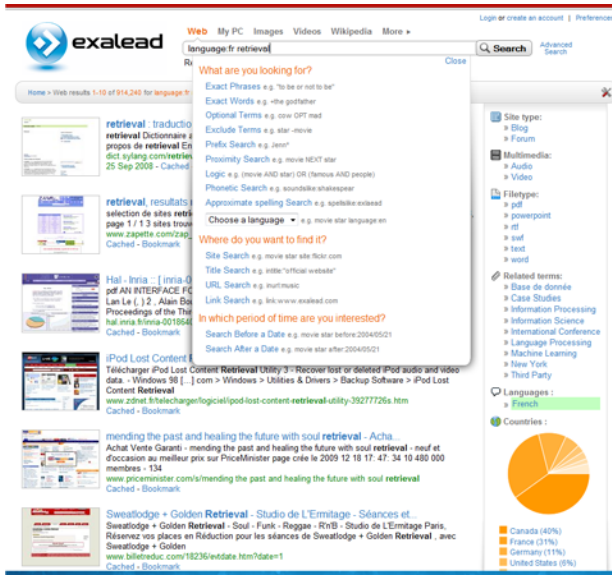
Fig. 21.5 Exalead English-French Cross-Language Search
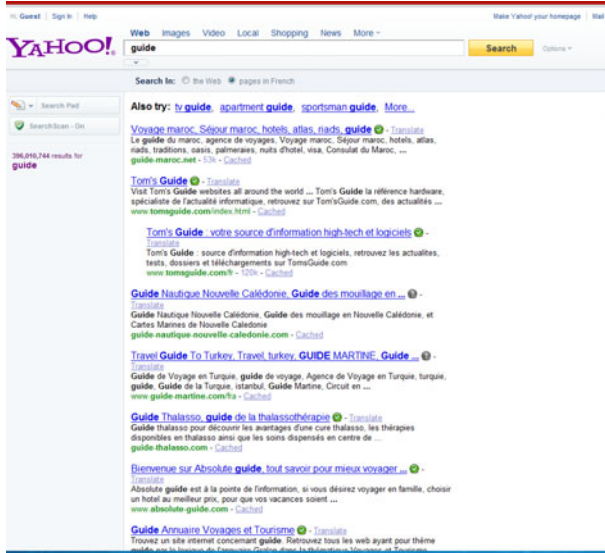
Fig. 21.6 Yahoo English-French Cross-Language Search

21.2.3   Access to the Deep Web

Our third search engine to compare was American Yahoo. It browses numerous static databases in many languages and, in addition to the query models,

refinement is supported by metadata search [15]. Some document formats like e.g. pdf, doc, xls, ppt, ps, not quite often pointed pages, dynamic pages, JavaScript web sites, private or with limited access pages, spiders or crawlers remain missed when conducting traditional search [16]. Since 2005 Yahoo has been exploiting the technology that sometimes, via subscription or login, allows the user to access invisible pages by clicking a search result, which is a challenge for translingual retrieval [17].

Before conducting translingual search, a user selects an option - Search only for pages in (language) - so that the system does not perform any translation model, but only browses the databases in a chosen language.

In spite of the identical spelling of the query word "guide" in both languages, the example in Fig. 21.6 provides the French results in context, which makes it clear that the word is in target language.

21.3 Query Models Efficiency in Translingual Retrieval

Altogether fifty English tasks were entered to the three information systems for transliningual search modeling and refining each of the queries. Before collecting the final results, we eliminated the items being false positive. Since the number of results varies enormously, the entity preferred was a title for the field restricted search.

Likewise, implementation of the query models such as fuzzy search, wildcard and regular searches is determined by the combination of the syntax, asterisks, or the query word popularity expressed by the number of its synonyms. So as to avoid misinterpretation of our analysis results, we agreed on entering the queries in their standard forms without any feed.

For all the relevant results we compute quadratic relevance mean

$$\overline{M}_q(r) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} r_i^2}$$

where: $0 \leq \overline{M}_q(r) \leq 1$:
n – number of queries
r – number of relevant results for each query

From the histogram, a relevance cut off point could be selected to determine proportions between the search engines performing different translation approaches. Overall, the transcription model exploited by Exalead proved the most efficient for the query models analyzed.

The only exception is the fuzzy query model in which case considered were the only query word synonyms. Latent Semantic Indexing deployed by Google improved matching process in comparison to the other information systems.

Interestingly, the most commonly conducted searches like keyword, field restricted or phrase searches turned out to be equally efficient regardless of translation approach.
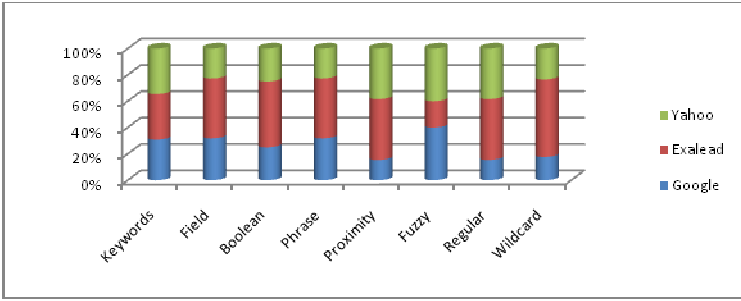
Fig. 21.7 Quadric Relevance Mean of the Query Models

21.4 Vertical Search Strategy

Comparing a couple of information systems that perform different search models and browse variety of databases, a user is supposed to prefer the one that provides the results of his or her interest on the top ranking positions irrespective of the system overall precision or recall.

Therefore, in our experiment we use matrices M_G, M_E and M_Y which rows represent the recalls for each query and the columns that are their vertical search results. The recall, being the matrix element, is measured as a sum of the ranking positions that proved relevant for the user.

$$x_{ij} = \left(log \frac{\sum_{i=1}^{j} p_i}{N_i} \right)^{-1}$$

This way we attempt to analyze the systems performance from two perspectives simultaneously: how the system precision changes while we narrow the query and if it is portable to the other systems.

Table 21.2 A sample from the total number of the results to build the Google Matrix M_G, Exalead Matrix M_E and Yahoo Matrix M_Y

	One-word query	Two-word query	Three-word query
Exalead	17,000,000	465,000	63,200
Yahoo	529,000	73,000	2,380
Google	3,710,000,000	271,000,000	46,900,000

We observe that the proportion between the relevant and total number of results is in increase, thus vertical search produces a set of ranking responses which is becoming more and more dense while narrowing the query. In other words, the number of total results is falling down while the number of the relevant results is approximated as constant. In some cases, we obtain a relatively small resulting set of nearly all the relevant results.
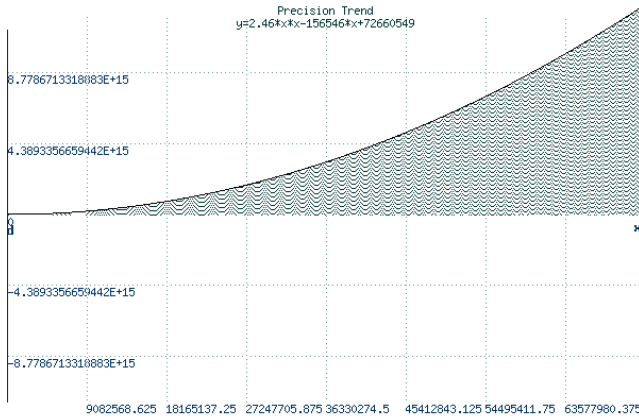
Fig. 21.8 Precision Trend Function of a sample vertical search strategy

$$f(x)=2.46x^2- 156546x+72660549$$

Approximation of the precision trend function, as the next stage, is based on the collated data of the final results analysis. At this step, we apply the least square method. On defining a distribution of a standard deviation of the determined function values from the approximated function values, we estimate a distribution kind.

For a normal distribution, we formulate a minimal sum of the deviation squares, otherwise we compute a weighted coefficient with an inverted proportion to the average value of the standard deviation squares of the function being approximated.

21.5 Conclusion

In contrast to existing approaches, we develop a search strategy that a user can significantly influence the information system efficacy. Vertical search proves beneficial improvement for the three translation models.

Broadening a knowledge about particular issue, a user is becoming more aware of the information need which is transformed by entering the more precise queries. We noticed that the novel technique proposed leads to change in proportion of the relevant to the total system responses.

Depending on the database and the query profile, presumably sometimes also other factors like the system scoring algorithm, or frequency of the query term occurrences in the web documents, it is possible to achieve a ranking list of almost all the relevant results.

In the further stage, we plan to concentrate on the precision trend function in order to develop a methodology for intensifying the set of the relevant results.

References

1. Hiemstra, D., Kraaij, W., Pohlmann, R., Westerveld, T.: Translation Resources, Merging Strategies and Relevance feedback for Cross-language Information Retrieval. In: Peters, C. (ed.) CLEF 2000. LNCS, vol. 2069, pp. 102–115. Springer, Heidelberg (2001)
2. Brashler, M., Ripplinger, B., Schauble, P.: Experiments with the Eurospider Retrieval System for CLEF 2001. In: Peters, C., et al. (eds.) CLEF 2001. LNCS, vol. 2406, pp. 102–110. Springer, Heidelberg (2002)
3. Gey, F.C., Jiang, H., Petras, V., Chen, A.: Cross-Language Retrieval for the CLEF Collections – Comparing Multiple Methods of Retrieval. In: Peters, C. (ed.) CLEF 2000. LNCS, vol. 2069, pp. 116–128. Springer, Heidelberg (2001)
4. Si, L., Callan, J., Cetintas, S., Yuan, H.: An effective and efficient results merging strategy for multilingual information retrieval in federated search environments, Information Retrieval, pp. 1–24. Springer, Heidelberg (2008)
5. Schermen, C.: Google Power – Unleash the full potential of Google. McGraph-Hill/Osborne, US (2005)
6. Kraaij, W.: Variations on Language Modeling for Information Retrieval,, ISSN 1381-3617; No. 04-62 (CTIT Ph.D. - thesis series, Taaluitgeverij Neslia Paniculata, Uitgeverij voor Lezers en Schrijvers van Talige Boeken, Rotterdam, The Netherlands (2004)
7. Lam, W., Chan, K., Radev, D., Saggion, H., Teufel, S.: Context-based generic cross-lingual retrieval of documents and automated summaries. Journal of the American Society for Information Science and Technology 56(2), 129 (2005) ABI/INFORM Trade & Industry
8. Eldén, L.: Matrix Methods in Data Mining and Pattern Recognition. SIAM, The Society for Industrial and Applied Mathematics (2007)
9. Steinhart, D.J.: Summary Street - an Intelligent Tutoring System for Improving Student Writing through the Use of Latent Semantic Analysis, Faculty, of the Graduate School of the University of Colorado, Department of Psychology, Ph.D. Thesis (2001)
10. Park, L., Ramamohanarao, K.: Efficient Storage and Retrieval of Probabilistic Latent Semantic Information for Information Retrieval. The VLDB Journal (2009)
11. Wolfe, M.B.W., Goldman, S.R.: Use of Latent Semantic Analysis for Predicting Psychological Phenomena: Two Issues and Proposed Solutions. Behaviour Research Methods 35, 22–31 (2003)
12. Corft, B.: The Modern Algebra of Information Retrieval. Springer, Heidelberg (2008)
13. Hock, R.: The Extreme Searcher's Internet Handbook – A Guide for a Serious Searcher, 2nd edn. Randolph Hock, US (2008)
14. Arasu, A., Novak, J., Tomkins, A., Tomlin, J.: PageRank Computation and the Structure of the Web: Experiments and Algorithms., Technical Report, IBM Almaden Research Center (November 2001)
15. Baush, P.: Yahoo Hacks. O'Reilly Media Inc., Sebastopol (2006)
16. Kasman Valenza, J.: Power Research Tools – Learning Activities and Posters, Joyce Kasman Valenza (2003) ISBN 0-8389-0838-1
17. Su, W., Wang, J.: Automatic Hierarchical Classification of Structured Deep Web Databases. In: Aberel, K. (ed.) WISE 2006. LNCS, vol. 4255, pp. 210–221. Springer, Heidelberg (2006)

Chapter 22
Music Information Retrieval on the Internet

Zygmunt Mazur and Konrad Wiklak

Abstract. The chapter discusses selected issues in the field of music information retrieval on the Internet. Differences between music information retrieval and text based information retrieval are indicated. Ways of extracting information from audio files are discussed. ISMIR (Internet Systems of Music Information Retrieval) based on MIDI files, algorithms for identification of audio files based on so called fingerprints and ways of updating audio file metadata are discussed. New algorithms for positioning results of music files searches are proposed and essential elements of contemporary Internet Music Information Retrieval Systems are indicated.

22.1   Introduction

The main goal of every information retrieval system is to find results for user imputed query in collection of documents. A document is not only a book, journal, article, electronic library (e-books), Web resources, patent or copyright certificates, but also separate parts of the text, such as clause, chapter, paragraph, e-mail message, newspaper article, etc. In recent years we have been dealing with enormous dynamics of data growth, the data are easily and mass generated through electronic media, electronic mail and internet services. The documents are not just files in textual form. The documents can also be audio or graphic files. In the information retrieval systems documents are stored in unstructured form. During process of classification, documents are mostly bound by automatic indexing process with descriptors – terms, keywords. To search for relevant documents for the query it is crucial to use such methods as collecting organization and searching

Zygmunt Mazur
Wrocław University of Technology, Institute of Informatics, Wrocław, Poland
e-mail: Zygmunt.Mazur@pwr.wroc.pl

Konrad Wiklak
Wrocław University of Technology, Institute of Informatics, Wrocław, Poland
e-mail: konrad.wiklak@gmail.com

(very often even approximate) huge sets of resulting documents, that let specify degree of compatibility (similarity of document's from collection subject with a query specified by user). In the last decade Internet information retrieval influenced very fast progress of information technology. At the end of twentieth century on the search engines market were leading companies Yahoo! and Altavista. After 2000 came era of Google. That company developed some standards of Web information retrieval. The Google is universal search engine, but in practice it shows that in some domains, such as graphics or sound searching, is better to use technical search engines. The subject of the chapter are the methods used to find information contained in sound files stored in the Internet.

22.2 Differences and Similarities between Music Information Retrieval and Text Information Retrieval

Today, most of existing search engines base their work on processing text documents. The issue of text information retrieval has already been analyzed, but still some new issues are appearing, for example related with new methods data exploration, that are used in extracting statistical data from text and it's later processing. Although still in Internet search engines couple of parameters still are covered by commercial secret, it is commonly known, which algorithms are the base of searching and positioning result documents.

The most important operation providing information retrieval is documents and queries indexing. The indexing includes selecting of subject or object and expressing it in information retrieval language. In ensuring indexing uniformity very important role is automation of indexing. The database of search engine has appropriate for itself possibilities of query designing.

In most of search engines text information retrieval gives very good results. Unfortunately retrieving information from sound files produces some other problems, that do not exist in methods of text information retrieval.

Each sound file contains – in the compressed form or not – recorded frequencies of sound signal. Depending on file format, such file contains sample rate, sample size, file length and additional data. Some formats of sound files, such as MP3 and OGG provide entry additional metadata, that allow inserting, editing and reading information about song title, artist, album, album year, genre or even copyright directly from the file. With sound file metadata is possible to treat the issue of searching music files, located on the Internet, as a special case of text information retrieval. Such approach is used widely in many MP3 search engines. Unfortunately, metadata inserted by users can contain any of the content and could not correspond to real data values. For this reason, a good search engine should provide additional way to verify correctness of the inserted metadata.

At this point, the similarities between text and music information retrieval ends. Because song is a sequence of sounds, that is illustrated with musical notation, we may consider to retrieve that sequence of sounds (or notes) – played by the specified instrument. Such information, which is another form of describing sounds than musical notation, may be useful for people who learn playing song for specified

instrument, but they do not have got music sheets. Similarly, it is possible to give, as input search engine data, sound sequence – using appropriate query notation or query by humming, whistling or singing part of the song, and finding songs, that we want to retrieve.

22.3   The Method of Extracting Information from Audio Files

In this subsection problems of extracting music information are discussed. As already mentioned, in case files that contain metadata, the information retrieved by searching robots – so-called *crawlers* or *spiders* – can contain incorrect data. Therefore is obligatory to verify metadata with real values. It can be achieved through manual update by a group of domain experts. However, due to cost of experts employment and a huge number of music files, this solution is inefficient. Another possibility preventing from retrieving incorrect audio file metadata is automated file metadata update with using of appropriate software. At this point another problem appears – how correctly indentify file using only information about subsequent frequencies, that are stored in that file? This problem is resolved in subsections 5 and 6.

The music search engines basic tasks are for example searching audio files using specified query music – a query string, that is inserted by user, describing a part of song – or choosing sequence of sounds from given audio file. The issue of retrieving sounds pitches stored in digital audio signal is task of *sound analysis* module. To obtain, from recorded frequencies of audio signal, a sequence of played sounds, it is necessary to get knowledge about so called *sound spectrum*. The phenomenon of sound spectrum can be explained on example vibrating string of guitar or piano. During the vibrating on entire length, the string of musical instrument generates a basic sound frequency – so called *pure tone frequency*. In fact not only entire string vibrates, also vibrate half of string, generating sound with frequency equals doubled basic tone frequency, also vibrates third parts of string generating sound equals tripled basic tone frequency etc. Therefore output audio signal for sound of given pitch is in fact sum of sound components, generated from sound generating material, which instrument is built.

The components of sound, included in the output audio signal, for a specified sound pitch, that are equal n-times pure tone frequency are called *n-harmonical*. In practice important for sound analysis is specified initial scope of *harmonical series*, not necessary beginning from the base tone frequency. Knowing the spectrum of the sound, generated by specified music instrument, is possible to extract from the recorded digital audio signal sequence of sounds played in song. The problem appears in singer voice analysis. Human voice is much complex instrument and to identify pitch of it, is necessary to know how human speech analysis works. Therefore the two well-tuned pianos will generate sounds with similar *timbre*, but the sounds of two different people will be completely different. That occurs, because the *sound timbre* results directly from the audio signal spectrum. The way how people hear specified sound depends on number of its *harmonical series* with the high and low frequencies. Is possible to perform sound analysis on low noise level audio signal generated by single music instrument. It does not

matter if the multiple sounds (*a chord*) have been played at the same time – based of the knowledge of sounds, that are generated by given instrument, spectrum is possible to proceed sound analysis. The problems appear in case percussion instruments, because that instruments often do not generate only a tone, but also a different kinds of sounds like noises.

The most significant problems arise in case playing multiple instruments at the same time. A modern music concert is a good example of that kind of situation. Typically simultaneously playing at least two guitars – lead and bass guitar, drums, often piano or keyboard and also singer is performing the song. Because of different characters and kinds of sounds generated by instruments – as it was mentioned before percussion generates noise and the guitar generate tones with *harmonical series* in spectrum – an analysis of each instrument separately is almost not possible. The resulting audio signal is a sum of amplitudes sound wave generated by each of the instruments. This follows, that to retrieve sequence of notes played by each instrument will be required extraction the frequencies of sounds generated by specified instrument from audio signal. Also this operation will need knowledge about the spectrum of sound wave generated by that instrument. Unfortunately for today there is no efficient and fast algorithm, that could handle with this task. Therefore current algorithms, that are used to identify the music files, use analysis the whole sound file – the sum of frequencies generated by instruments during playing the song. So there is no possibility to extract sequence of sounds, played by only one of the instrument playing in the song. A solution could be recording each instrument to separate audio files or developing a new audio file format, that would contain data of each recorded instrument on a separate track/canal. But that would cause huge increase of disk space usage by music files in new format and the same will occurs in case using separate file for record single instrument data.

22.4 Internet Music Information Retrieval Systems Based on MIDI Files

One of the audio files format, that are best to obtain music information from the Internet, is MIDI format. In fact MIDI is not real audio file format – it does not contain frequencies of audio signal, but a set of commands to process by synthesizer, including data about sound pitch, which is generated at the moment. MIDI standard were most popular from the late eighties to late nineties in the twentieth century. Then in every computer magazine a hundreds of MIDI files (with ".mid" extension) were added on CDs. Additionally with MIDI standard are compatible all of the instruments, that synthesize sound – most of them are popular instruments called *keyboards*, which are keyboard instruments, that could replace another instruments. With *keyboard* is possible to freely play like the piano, the guitar or the trumpet. However, well working human ear could easily feel the difference between synthesized sound, played on *keyboard*, and real sound, played on original instrument Unfortunately even the best sound synthesizer still could not fully replace real instrument. Generated sound, by synthesizer, is as artificial

as the way of its creation. This finally decided that files in MIDI format, have been replaced with today popular formats MP3, OGG or uncompressed WAVE. However MIDI files still could be useful today – especially in games or karaoke programs. MIDI is ideal file format for music information retrieval. Every file is divided into *n channels*, where *n* is integer in scope *[1,16]*. Channels are divided into tracks. Each channel could contain unlimited number of tracks. In the channel on selected track any musical instrument could be synthesized – it is in common to use channel 10 (and also often 9) with percussion instruments only. On each track are located so called MIDI file *events*, that define beginning and end of track, duration time and pitch of synthesized sound and which instrument is synthesized on current track. The most important of events are:

- NoteOn – begins playing current sound with given pitch
- NoteOff – ends playing current sound with given pitch

Other events are less important. Based on the events is possible to read duration time of current sound. With all this information, although same NoteOn events are enough, is possible to easily obtain played sounds by given instrument. The great advantage of MIDI files is fact, that it is impossible to occur any reading or recognizing sound errors – all data about, which current sound of given pitch and played by given instrument is generated, are simply stored in file. Therefore, at first glance it seems, that every information retrieval systems based on the recorded sounds, should be using only MIDI file format. Unfortunately this format has also one big disadvantage – converting format WAVE to MIDI, MIDI to WAVE or another sound file format is very complex, and currently there is no software, which could efficient provide converting between MIDI and other audio formats. That results in fact, that systems based on MIDI format, working only with MIDI files or searching for other music file formats ,after initial MIDI search, using metadata stored in system.

Internet music information retrieval systems based on MIDI are good example of the use different ways to obtain music information, in particular:

- Query based on user-specified part of sequence of sounds, using special notation or generated by software written in Flash or Java – for example simulating piano keyboard. This solution undoubtedly gives best results. – with proper length sequence of sounds played in song – but only if we exclude sound duration time, because pace is subjective song feature and depends individually from performer or conductor. In truth in this case of queries sound time duration could be useful, but this results in more data to store and much more data to proceed and less number of query results – it can exclude searched song from results, because it was performed with another pace – pace is determined by fractions of seconds.
- Query based on user activity: humming, whistling or singing part of song. Of course longer part than results should be better. In practice it appears, that most of human population has no vocal abilities, therefore that case of query rarely brings good enough results.
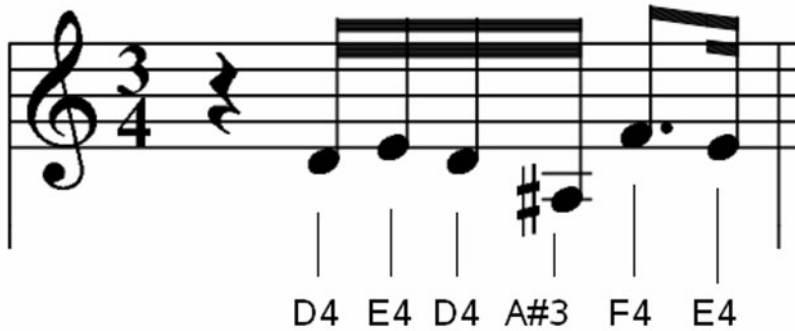
Fig. 22.1 An example of music notation used in MIDI-based music information retrieval system

Additionally in both methods instead of special notation of sequence pitch of sounds is possible to use *Parsons Code* notation. The *Parsons Code* is transcript of melodic contour, it describes relationships between pitches of two successive sounds from the song. *Parsons Code* uses following symbols:

- "*" – symbol which indicates beginning of the song and also first played sound in song. It can appear only once in the whole song recorded with Parsons Code.
- "U" – symbol which indicates that next sound was played with higher pitch, than previous sound
- "D" – symbol which indicates that next sound was played with lower pith than previous sound
- "R" – symbol which indicates that next sound was played witch the same pith as previous sound

An advantage of Parsons code is that is possible to eliminate flat and sharp errors occurred while singing, humming or whistling the song. Disadvantage is that many songs could contain similar melodic contour, even though they were played with different sounds pitches, therefore query returns to many matching results. This could cause that searched song is lost somewhere between redundant results. Therefore, Parsons Code should not be used to songs retrieval based on special notation that contains sequence of pitches of sounds query, inserted by user. Parsons Code could be used with queries by humming, whistling or singing any part of the song.

An example of existing Internet music information retrieval system based on MIDI format is Musipedia *(http://www.musipedia.org)*. This system allows to creating queries, by using one of the following solutions:

- by giving the title, author, album, and also this solution provides songs search by playing a part of the song on special Flash application, which simulates piano keyboard
- by whistling or humming a part of the song (song is converted to Parsons Code)
- by playing song with one note – this query is based on rhythm and pace of song
- by manually inserting Parsons Code

It is possible to search for the music files in Musipedia repository and also in the whole World Wide Web. The test show that most effective is first search query solution, while second and last usually return good results. Unfortunately first solution uses time of sound duration, which lowers the set of correct results, as was mentioned earlier. However completely failure is solution based on rhythm and pace of song – it returns no correct results or the correct result is missing somewhere between redundant results. Musipedia is one of the first Web music systems information retrieval on the wide scope.

22.5 Algorithms for Songs Identification Based on Spectral Analysis and Fingerprint Generation

The most popular audio files formats in Internet are certainly MP3, WAVE and OGG. Metadata contained in most of the audio file formats, are basis for today's MP3 search engines. However, there is one serious problem connected with metadata contained in sound files. Metadata often are not equal real values or it can not contain information – only a part of metadata is inserted. Often is necessary to fill file metadata based on file contents. Another useful for users service, which uses that kind of file identification could be obtaining data about song – title, album, artist, album year, etc. – from fingerprint extracted from part of audio file recorded from for example radio broadcast. Fingerprint extraction could be processed on local user computer, but then extracted fingerprint is sent to service server, which returns as results found music files. To achieve described goals, it was necessary to develop efficient and fast enough algorithms for extracting features from audio files and for searching based on that features.

Now will be presented different approaches, that illustrate evolution of issue searching for music files, based only on audio file. First step of each algorithm is checking whether the file is in compressed format or not, and then, if file is in compressed format, uncompressing the audio file to raw PCM (Pulse Code Modulation) format – often it is just a WAVE file. Next, after extraction of audio frequencies recorded in file, is possible to define basic algorithm of file search. The basic solution is binary search algorithm. However, compare of two audio files is firstly time consuming and secondly sending files on server to compare them with others will cause server overload. This type of data transfer is possible only on mobile networks operators.

Knowing all of the problems connected with binary search, attempted to construct file searching based on calculated hash function value. The idea of this algorithm is quite simple to understand. The client-site software calculate hash function value and send that value to server to compare it with other hash values stored on server, using the similarity relation between. Hash function have one enormous problem – they are sensitive for file structure changing. A single bit difference between two files results completely different hash value – event files contain the same songs. Therefore idea of build music file identification system based on binary representation was abandoned and algorithm concepts focused on use of

audio signal frequencies. Important activity is spectral analysis of audio signal, that is saved in music file (as PCM).

The algorithm using so-called *fingerprint* [1] consists of the following steps:

1. Decompressing audio file to raw uncompressed audio format (PCM).
2. Division audio file into so-called *frames* – constant length fragments of audio signal, with constant specified size. The important feature of frames is 50%-70% overlapping each of two bordered frames. The overlapping is caused by fact, that the recorded part of song could be shifted in time in relation to original audio file. Frames generation takes place with using proper windowing function. In case Philips implementation of this algorithm the Humming function is used as windowing function.
3. Transformation of frames. In order to transform audio signal to a form, that permits extracting features, that are necessary to determine similarity between two audio files, for each frame-window appropriate transformation is applied. Often the transformation is a Fourier Transform, through successive components of the audio file spectrum are obtained. Another transformation are proposed in[1,2].
4. Features extraction. A sample feature obtained in extraction process could be *Power Spectral Density (PSD)*.
5. Normalization of obtained results
6. Generation of fingerprint

Fingerprints, generated this way, are stored with audio file metadata in search engine database. In newer systems, for audio files classification, artificial intelligence elements are used – for example neural networks. Different versions of described algorithm are used by many well-known companies like Philips. Results of tests of algorithms properties are described in [1,2].

The upgraded versions of extraction fingerprints algorithm are used in:

- In programs like TrackID – using TRM technology, developed by Relatable company. However in commercial solution, this technology often did not correctly matched files to metadata, and now is replaced by another solutions.
- In MusicDNS – technology developed by AmpliFind Music Services Company. The MusicDNS is based on the black-box rule. User sends query to server, that hosts MusicDNS service, and as a response the PUID is returned. The PUID is an unique identifier, that is obtained through complete analysis of audio file. Based on PUID identifier is possible to process search for music files in one of most popular Internet music databases, like MusicBrainz. The MusicBrainz database provided also its own software, that uses MusicDNS, which is used to update metadata in users sound files. This technology gives very good results.

22.6   The Method of Updating Missing Metadata in Audio Files

In this section subject of updating metadata for music files is expanded. The basic method is using one of existing Internet music databases, that store information

about songs. That databases are for example: FreeDB, Discogs, MusicBrainz, and other similar Web databases. Based on already possessed knowledge about title, artist, album, album year, genre or even the name of company that released the album, is possible to find missing metadata. This is relatively good solution, but only if enough knowledge about the music file metadata is possessed. Unfortunately often all file metadata are missing – no information except sound file is possessed. In that case is possible to use one of existing music file identification systems (like MusicDNS), that based on audio fingerprint file identification – like Picard software, that uses MusicDNS and MusicBrainz.

In case when we possess original CD or DVD with audio tracks, there is another, older than fingerprint identification, possibility to identify metadata for that songs. It is possible to calculate the disk identifier – DiscID, based on the length of each audio track in the disc. DiscID works now with many popular Internet music databases and audio players – like Windows Media Player, Itunes, etc.

22.7 Algorithms for Positioning Search Results Music Files URLS

We will present here a new algorithms for positioning search results *links* to music files. A link mean here an extracted URL to music files, based on one of the following html/xhtml tag, that are used on WWW pages:

- A – reference to object (link)
- EMBED – multimedia site extension (this tag is not defined and compatible with W3C standard, but is supported by popular Internet browsers, like Internet Explorer and Firefox)
- OBJECT – a tag, that enables placing on Web site audio or video file

These methods of attaching audio data to WWW site provide direct access to attached files, which enables possibility to download file on local disk and then perform another operations on it. Indirect file access, like streaming or download script calling, are not included in the algorithms. Indirect access often prevents from direct downloading and saving file on a local computer.

Now we propose algorithms of positioning URLS referencing to music files. These algorithms could be used by all music files Internet search engines, to proper sorting the search results. For "normal" – text based – pages have been developed already many well working algorithms based on connections-links between Web sites. These algorithms are: HITS, Page Rank and Salsa. Unfortunately these algorithms couldn't be applied to audio files. That occurs because searching process is done in completely different domain, than text documents. All of the algorithms, that positioning WWW sites with text-information have to cover how many WWW sites, and the most important – what (the quality criteria) WWW sites refer to positioning Web site. In case of audio files that criterion do not always give best results. For example positioning based on results of average of result values returned by algorithms Page Rank, HITS or SALSA can be used, but that solution completely misses such important features like correct metadata or

completeness of the song. More valuable is file containing the whole song, than a file containing just a part of song or additional "zone of silence" after the ending of song. As an additional criterion sound file size can be used.

From above analysis we propose three solutions, that can be used in algorithms for links referencing to music files positioning:

- The smaller is sound file size, the faster file will be downloaded to perform next operations on it. An advantage of this approach is saving the transfer rate. A disadvantage, that downloaded file is highly compressed and that will cause problems with playing quality.
- The larger is sound file size, the better quality will be file. The advantage of this approach is that in most cases downloaded file is best playing quality. However the bigger file size causes larger bandwidth usage.
- The compromise between the quality of playing audio file and the transfer rate. As the file size will be closer to specified value (it can be, for example, the arithmetical mean of size of indexed files), the rank of the file will be better. A properly selected optimal size file parameter, could result in significant improvement of transfer rate and quality of downloaded files, than two other methods presented above.

Considering the fact that if the file contains correct metadata, then it is a real music song, achieving of good parameters of file quality and transfer rate could be a base for final rating the quality of music file. In fact is not important where on the Internet file is located. Important is the information, that is contained in audio file. The more data are accurate and valuable, the file evaluation will be better (with higher value). The advantage of this approach is fact, that there is no possibility to falsify the results, by constructing Web sites with a specific structure of links connections between sites, to improved rank of searched file. In this way is easy to "cheat" search results positioning system for example in Page Rank algorithm, which is used in Google search engine. However if another criterion of positioning audio files is needed, then is possible to use as additional criterion sorting in descending order by the numbers of links leading to current sound file. That solution shows that if file is very popular, then it is susceptible to falsify its position by constructing Web sites containing many links to this file. In this case is good idea to develop some hybrid solution with one of text-based Web search results positioning algorithms with includes calculating weight for each WWW site, that contains link to music file and calculating weighted mean for each positioned sound file based on previously calculated weights for Web sites. To calculate weights for WWW sites is used one of text-based algorithms like Page Rank. Next is necessary to, as was mentioned before, calculate the weighted mean using all weights corresponding to sites, that contain links to currently positioned file. Result of this calculation is used to rank the position of audio file. This result should be treated as another criterion of sorting Web search results – at first the result files should be sorted with quality and correctness of metadata criterion, and then with use of calculated weighted mean. That gives us a certainty, that link to file is located only on Web sites with high quality of text information. However this solution requires high level of computational complexity and large memory usage. If positioning files with quality

criterion have been used, then large number of links to file do not falsify the results file rank and in that case it is not necessary to use hybrid algorithm. The algorithm based on music files quality does not retrieve any information from number of links. The positioning using music files quality is not susceptible to any influent of number links to the file, but only on the order of indexing Web sites by Web spider and on the metadata, which are stored it that file.

Based on proposed solutions, three algorithms of positioning result music files are proposed:

- *No positioning of music files* – it is the simplest solution. It significantly reduces the cost of implementation and work of the proposed Internet music information search system and its complexity. However the search results are not always will be correspond of the users expectations. That probably will cause lack of popularity of this search engine. It is not recommended solution.
- *Positioning of music files based on the number of links leading to file* – it is relatively simple solution to implement. During the crawling, when the Web spider retrieves additional information for database, for example: how many links leading to file are contained in Web sites. However this method is vulnerable to falsify the results of algorithm, by preparing many Web sites, that contain links leading to positioned audio file. That implicates necessity of assigning weights to sites, with using one of text-based positioning algorithm – the hybrid solution. In the simplest form it will be analysis, performed during the music file search, of results returned by one of the popular text-based search engines like Google. Using of our own hybrid methods will results in enormous complexity of required calculation and large amount of memory usage. The sound file quality criterion, in that version of algorithm, is number of links leading positioning music file.
- *Positioning of music files based on the correct metadata, completeness of the song, song quality and the transfer rate* – this solution ensures that user will not have to manually fill or update metadata in downloaded file. The downloaded file contains complete song, that will be played in proper quality or – in case of different criterion – will not the connection overload and not consume too much data transfer, during downloading the file. This solution gives best results, but it requires analysis of each song time duration, identification of song metadata based on audio file and analysis of the file size. If another criterion of results positioning is required then numbers of links to music file can be used, however it is optional.

22.8 Proposals for Quality Measures of Contemporary Internet Music Information Retrieval Systems

This section discusses the conditions, that have to be satisfied by modern Internet music information retrieval system. We have divided our considerations in two parts – first defines features that must be satisfied by module, that is responsible for new music files retrieval and the second part defines system's features from the user's requirements point of view.
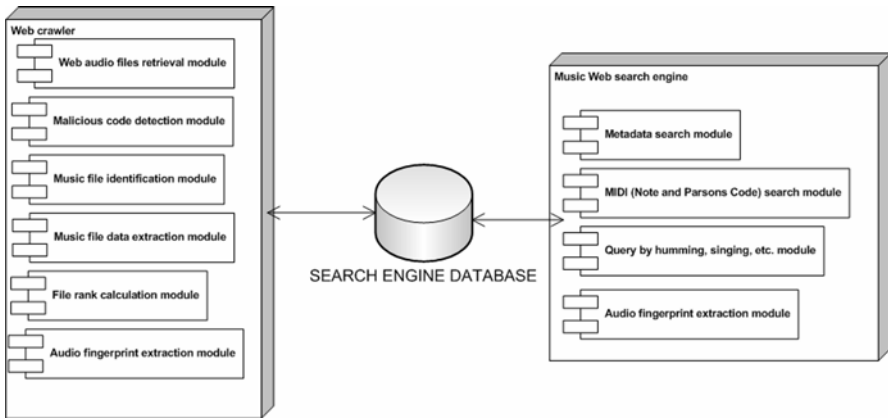
Fig. 22.2 Basic Internet music information retrieval system architecture

22.8.1 The File Retrieval Module

Every Web music information retrieval system should have at least one (single or multithreaded) *Internet spider (crawler)*, that search and retrieves music files from the Web. The process of searching the Web and indexing the files is a continuous process. After finishing the file indexing process is required to restart the crawling to search and index new files, to update changes in files and to delete not up to date links. The *database, that stores the links to music files*, should not store redundant data. Additionally Internet spider should has *module of detection already indexed Web sites* – to avoid unnecessary re-analysis and loops in stored Web sites links. As a crawling start point, the *initial Web sites set* should be prepared. Next sites will be indexed based on the Web sites structure – connections (links) between sites. The system must contain appropriate *module to checking whether the metadata and links stored in search engine database are up to date*. As additionally upgrade of crawling process could be assigning *priority* to Web sites, and implementation of crawler, that uses these priorities. As an example of criterion of Web sites priority will be fact, that WWW sites, that contains links to audio files should be processed more frequently than those that not contain such links.

A very important factor affecting the quality of the indexed sound file, is the presence of malicious code. This is a problem, that regards almost every MP3 search engines. Is possible to simple infect computers with malicious software while the audio file is downloaded. That happens because the search engines do not check files in case of infection by malicious code. A good music files search system *must check indexed files for malicious code presence, using one of many existing Web services*, developed by professional companies that fight with malicious software and provide Web links scanners. The process of scanning the file should be proceeded *before start downloading of that file and before any other operations on that file*. If a virus or other threat is found, then is required to immediately stop any operations with current processed link. Additionally such kind of

link could be inserted into another table in database, that contains malicious links. With this information, rather than scanning each link is possible to check whether the link is not in that table. Another required system behavior is to check – if only that kind of operation is possible in current audio file format – whether the *audio file copyrights allow to index the file by Web crawlers.* If the file passes both of these conditions, it is time to download and *extract required information from the file.* In this step of the algorithm is useful to *check the correctness of metadata –* for example with using one of the Web music files identification services like MusicDNS –and if necessary update the metadata information of the current file in database. The next step should be *obtaining and calculating parameters required to calculate file rank in search results* and *calculating the file rank.* A file rank, like also other data corresponding to music file, should be periodically updated.

22.8.2 The Music Search Engine Module

The main element of every modern music search engine is WWW site, that contains text field, which allows user to search for music files based on metadata. The search results should at first include all occurrences of string inserted by user, additionally next also results that are only partially equal searched string could be included. The results should be displayed, after sorting according to links ranks – that had been computed with positioning algorithm, in *tabular form.* The *columns should define displayed metadata,* that describes the audio files like song title, album name, artist, year of album release, music genre etc. The *rows contain audio files data, returned by search engine,* according to columns headers Additionally each row should contain, with using *buttons or hyperlinks, several additional options* like: playing online selected file, downloading or buying selected file – it depends on copyrights.
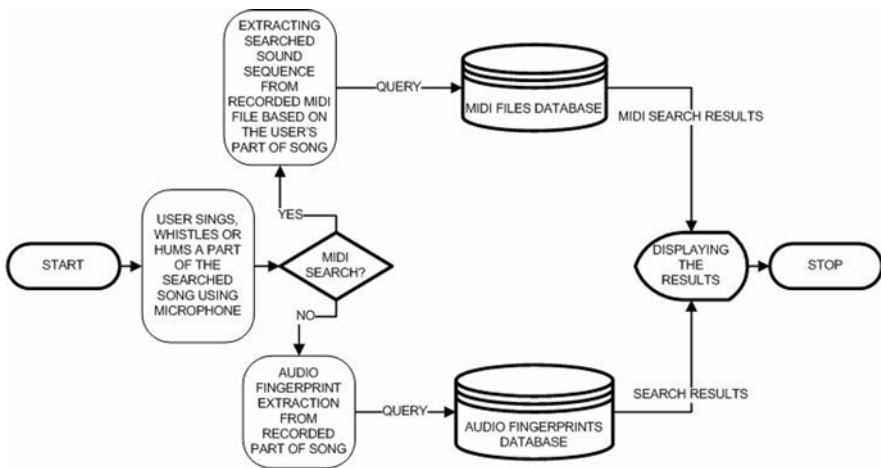
Fig. 22.3 The illustration of query by humming, singing or whistling algorithm

Once the basic search mechanism is properly defined, it is time to expand our Internet music information retrieval system. Is possible to implement some additional *search mechanism, that work with MIDI files*. Based on fingerprint files identification algorithms is possible to implement music files search engine, that uses *query by singing or humming, cooperating with social networking website*. To add new song, use has to personally sing it before uploading the file. Also every of the site users could sing any of the songs stored in portal. For every singed song fingerprint is generated and stored into search engine database. With the help of software written in Flash or Java, placed somewhere on the search engine WWW site, a user of the social networking website could sing or hum a part of searched song. Generated through that operation fingerprint is compared with fingerprints stored in the search engine database. This solution gives much better results than comparison with the original music file, that contains in addition to singer voice, other instruments, fingerprint. An example of practical implementation of presented concept is Midomi music search engine (*http://www.midomi.com*).

22.9 The Trends in Development of Internet Music Information Retrieval Systems

The chapter described problems and different ways to upgrade modern Internet music information retrieval systems. This includes algorithms of positioning the search results, improving the search engine efficiency, improving the quality of files metadata and making search process more attractive for user, by expanding system with new search options, like query by humming. All of the described solutions can be implemented today. Therefore the question arises – what brings the future? How will be constructed Internet music information retrieval systems of tomorrow? With some degree of probability it can be concluded that development of music search systems will be follow in the direction of using real time search algorithms. Web music information systems retrieval will use the social network websites in same way as Google search engine already does - processing in real time sing data collected from Twitter and Facebook. An example of this trend is already mentioned Midomi search engine – but it uses its own social networking website. However there is no obstacles to establish cooperation with major social networking sites, that provide music sharing, such as LastFM or Myspace. Another great step in development of music search engines will be popularization of new MPEG 7 standard. It is a standard, based on XML technology, which describes every data that can be found in audio file – from metadata to sound frequencies. Further improvements of music information retrieval systems will include lowering time of search The main question arises – will the solution, based on using social networking websites, result in real improvement quality of searching music on the Internet?

References

1. Cano, P., Battle, E., Kalker, T., Haitsma, J.: A review of algorithms for audio finger-printing. In: IEEE Int. Workshop on Multimedia Sig., Proc. (December 2002)
2. Doets, P.J.O., Lagendijk, R.L.: Extracting quality parameters for compressed audio from fingerprints. In: ISMIR 2005 (2005)
3. Homenda, W.: Intelligent computing methods in music information processing, Exit
4. ISMIR conference main site (2007), `http://www.ismir.net`
5. ISMIR conference proceedings, `http://www.ismir.net/proceedings/`
6. Haitsma, J., Kalker, T.: A highly robust audio fingerprinting system. In: 3rd Int. Symp. on Music Information Retrieval, ISMIR (2002)
7. Allamanche, E., et al.: Content-based identification of audio material using MPEG-7 low level description. In: 2nd ISMIR October 2001 (2001)

Abbreviations

ISMIR – Internet systems of music information retrieval

Chapter 23
Verifying Text Similarity Measures for Two Layered Retrieval

Andrzej Siemiński

Abstract. The goal of the paper is to assess the usefulness of various text similarity measures for the two layered Internet search. In that approach the first layer is a generic Internet search engine. The second layer enables the user to evaluate, reorganize, filter and personalize the results of first layer search. It is run on a local work station and can fully exploit the so called user dividend. Crucial for that stage is assessing text similarity between text segments. The papers discusses classical, statistic text similarity measures as well semantic, WordNet based semantic measures. The results of an experiment show, that without word disambiguation techniques the semantic approaches can not outperform statistic methods.

23.1   Introduction

The generic Internet search engines such as Google are excellent tool for finding a whole lot of pages as answers to user's queries. However the engines do not assist in an equal manner the user in selecting relevant pages. Even the most notable achievement on that field (The Page Rank) orders the pages in a global manner, common to all users which is mixed blessing. The URL address are useful only for the websites known to the user – a requirement not always fulfilled. The displayed sometimes text snippets are more useful but they are just randomly selected text windows with a term from a query. This is certainly not satisfactory. Far more useful would be the presentation a self contained text fragments such as a sentence or a paragraph that are pertinent to the user interests.

The main contribution of the paper is the introduction of the concept of a two layered, self –adopting retrieval algorithm. In that context the idea of the user dividend is introduced and explained. Crucial for that stage of retrieval is assessing

Andrzej Siemiński
Wrocław University of Technology, Institute of Informatics,
Wybrzeże Wyspiańskiego 27, 50-370 Wroclaw, Poland
e-mail: `Andrzej.Sieminski@pwr.wroc.pl`

text similarity between text segments. The papers analyzes both classical statistic text similarity measures as well semantic, WordNet based semantic measures.

The paper is organized as follows. The Section 23.2 outlines the concept of the two layered retrieval. The concept relies on assessing the similarity of texts. The paper verifies the usefulness of both traditional statistical and newer semantic algorithms for assessing text similarity. They are described in the Section 23.3. The semantic algorithms are based on WordNet which is briefly presented in the next section. To evaluate the performance of these algorithms an experiment was conducted. Its results are discussed in the 23.5 Section section. The concluding Section 23.6 indicates further research areas.

23.2   Two Layered Retrieval

For answering ad hoc queries the performance of common search engines could not be surpassed. The scope of indexed data and technical infrastructure make them immune to any competition stemming from university research area. Therefore the first layer relies on the regular web search. The search may be using the interface of the web search engine more preferable it will use web service offered by the engine; e.g. he Google company has published Google Web Toolkit. It enables developers to build and optimize browser base applications. The ranking of pages reflects generic page properties: the value of its host server or text similarity to the user query. The search engine is responsible for handling many millions of queries per day so the output, albeit prepared using sophisticated algorithms, follows the principle "one size fits all".

The aim of the second layer is to personalize the first layer output and make it more readable for the user. The processing is done on a local work station and therefore the knowledge of both user preferences and processing power could easily surpass the capabilities of a generic search engine. The layer can fully exploit the so called user dividend [1] – a set of features that could be utilized on a local workstation but are not available for a generic search engine.

The scope of actions performed at the second layer include:

- Prefetching best pages.
- Filtering out unnecessary pages.
- Displaying meaningful text fragments having the highest similarity to the user needs.
- Employing advanced lexical, syntactic or semantic text processing.
- Modifying user interest profile.
- Processing of the obtained results.

Downloading the pages before they are presented to the user has two advantages: improves response time (page is shown almost instantly) and far more important the pages text is available for further processing as described below.

The filtering out of some pages includes both the elimination of useful pages that are already known to the user and accordingly tagged or pages known in advance to be not interesting e.g. originating from user disdained servers.

Generic search engines display only a text snippet containing one of the query terms. The second layer could display a sentence or if necessary a paragraph most similar to the query. Having such a detailed information a user is better equipped to asses the usefulness of a page.

Local processing of texts makes possible a wide variety of activities starting with substituting words by their base forms, syntactic tagging or exploiting semantic similarity. Initially the user interest profile is made out of his/her query. The users prefer to formulate their queries in a simple manner and are generally unwilling to formulate sophisticated queries. This widely known phenomena. The two following factors is mainly due to two factors: the lack of familiarity with the advanced search options of search engines and the rigid nature of the options that are unable to catch the subtle nuance of user information needs. The second layer is capable to infer user preferences by analyzing his/her behavior. An extremely long time spent on reading a paragraph, specific user actions such as downloading a page, saving it on a local system or printing it out clearly indicate a profound interest in the page and therefore its content could be used to augment the query in a manner similar to that described in [2] and [3].

The second layer can fully benefit from the so called "user dividend": a collection of properties that are either not available or very difficult to obtain for an external search engine. In this case they include mostly: a capability of precisely recording user interaction with the search engine what makes it possible to infer about user preferences and the availability of processing power located on a user work station.

23.3 Statistical Full Text Search

The full text search is almost necessary for Internet due to its dynamic nature of Internet text resources and their immense size. There are attempts to provide high quality Internet search through the work of community of volunteer editors. The most notable project in that field is the ODP (Open Directory Project) which uses a hierarchical ontology scheme for organizing site listings. Listings on a similar topic are grouped into categories, which can in turn include smaller categories. Their total number is over 590,000 categories The motto for the project is "humans do it better" [4]. This year the number of sites exceeded 4,5 millions with more then 85 thousands cooperating editors. Despite such considerable resources dedicated the project its popularity is marginal comparing search engines like Google or Bing.

The traditional full text search does not take the advantage of syntactic nor semantic properties of natural texts and treats text as just bags of words. A text similarity measure should take into account two basic text properties:

- texts vary considerably in their lengths;
- words are more popular then the others.

Over the years the cosine measure is with or without the idf weighting facto have proved to provide high quality, stable performance [5].

Despite its popularity the word based full text search has certain deficiencies. A recent paper [6] contains a extensive list of them. Some of them are posed by shortened forms of terms or variant spellings could elevated by careful preprocessing while others like figurative language, the so called "aboutness problem" or the disambiguation of personal names are clearly beyond the scope of even most advanced NLP algorithms The middle ground are the problems stemming from the usage of synonyms, exploiting semantic similarity in general or word disambiguation. They could be handled by more refined methods. The next section describes the way in which a WordNet data base could be used to include the semantic factor into text search.

23.4   Semantic Full Text Search

The described above measures take into account only the occurrence profile of words. The inclusion of semantic text properties was not possible until the creation of comprehensive data bases describing the meaning of words. The most notable is the WordNet. It was created and is being maintained at the Cognitive Science Laboratory of Princeton University [7]. It groups more then 150 000 words into sets of synonyms called synsets, provides their short, general definitions, and records the various semantic relations between these synonym sets.

The WordNet can be interpreted and used as a lexical ontology in the computer science sense. Various algorithms for assessing semantic word similarity are described and compared in [8].

In what follows the processing was limited to nouns in their base forms. The reasons for that are twofold: the word similarity measures are developed mostly for the nouns and secondly they convey the text meaning and are commonly used for indexing. An addition bonus was the shortening of texts what was important taking into account the substantial complexity of the required computations. The compared texts are represented as sets of their words.

In the first three presented below approaches the calculation of the semantic similarity of two texts α and β starts with the comparison the similarity of each word w from α with any word from β.

In what follows maxSim(w, α) denotes the maximal similarity of between the word w and any of the words form α.

23.4.1   BestSim Measure

The first approach starts with the calculation of the sum(β,α) function which returns all words from β or α. augments the sequence α with words from the β. Next the values of sBest (α, sum(α, β)) and sBest(β, sum(β,α)) are calculated. The sSim function replaces each word w from the second parameter with the most similar word from the first parameter. The function SSWS is in turn responsible for replacement of the words by value of the maxSim function. The obtained values replace the frequency factor in the cosine measure. The resulting measure is

denoted by *BestSim(α, β)*. A following example taken from [9] should clarify the way the function is calculated.

α: "french basketball team" , β: "english football team" ;
sum(α, β)= "basketball english football french team"

Let us further suppose that wns(English, French)= 0.80, wns(foolball, basketball)= 0.73 and for all other non identical terms the wns function has value of 0.

st(β,α) = (basketball, english, football, french, team)
sBest(α, β) = (basketball, french, basketball, french, team)
sBest(β, α)=   (football, english, football, english, team)
SSWS(α, β) = (1.0, 0.8, 0.73, 1.0, 1.0)
SSWS(β, α,) = (0,73, 1.0, 1.0, 0.8, 1.0)

The standard similarity $cos(α, β)$ = 0,333 whereas *BestSim(α, β)*=0.97 a value which intuitively better reflects the relationship that exists between the two texts. The usefulness of using the word similarity is even more evident when we consider another set of texts: "French football team" and "soccer player group". The original cosine measure does not detect any similarity (not a single word appears in both texts) whereas the BestSim measure produces a value of 0,93 which is undoubtedly more appropriate.

23.4.2 The SimSum Measure

Let us consider three word sets:

α = {pilot}
β= {airplane, cockpit, airport}
γ= { airplane, apples, plums}

The values of BestSim((α, β) and BestSim((α, γ) are equal. The similarity between the pilot and airplane is the highest and therefore the pilots' similarity to cockpit and airport is ignored. Most humans would however judge text β as being more semantically related to α then γ. The SimSum measure tries to alleviate the deficiency by replacing the maxSim(w, α) a more elaborate function. The sumSim(w, α) function represents the cumulative similarity of the word w to all of words from $α = \{s_1, …, s_n\}$.
The recursive formula to calculate sumSim(w, α) is defined below:

   a. $simSum(w, \{ s_1, …, s_n\}, 0)$=0;
   b. $simSum(w, \{ s_1, …, s_n\}, i)$= $simSum(w, \{ s_1, …, s_n\}, i-1)$; for i=1,…,n
   c. $simSum(w, \{ s_1, …, s_n\})$= $simSum(w, \{ s_1, …, s_n\}, n)$.

The ordering of elements in the sequence α is not significant, is could be easily proved that all orderings lead to the same result. From the practical point of view ordering the elements according to decreasing similarity to w is useful as the consecutive iterations contribute less and less to the final score and therefore the process could be stopped early without influencing the final result significantly.

23.4.3 ExtSim Measure

Another approach to measure the similarity of two text segments was described in [8]. In an apparent difference to above formulas it does not augment the texts with vocabulary from the opposite text. It takes into account the idf weighting factor see the below formulas:

$$ExtSim(T1,T2) = \frac{extMax(T1,T2)}{2} + \frac{extMax(T2,T1)}{2}$$

where:

$$extMax(T1,T2) = \frac{\sum_{w\in\{T1\}} \max Sim(w,T2)*idf(w)}{2\sum_{w\in\{T1\}} idf(w)}$$

As in the case of previous measure the similarity score has a value between 0 and 1. Its value of 1 indicates identical text segments, and a value of 0 means no semantic overlap between the two segments.

The disadvantage of all of about introduced algorithms is their level of computational complexity. Computing similarity of two texts A and B involves finding for each term in text A the most similar term in text B. The computational complexity is thus proportional to n*m where n and b are the number of terms in respective texts. The calculation of term similarity is in itself a pretty complex operation. During the experiment to speed up the processing the similarity level of commonly encountered term pairs were cached. Even so the processing was substantially slower then in the case of methods based on the cosine measure.

In the next two sections two attempts are described that aim at incorporating the WordNet based term similarity without the necessity to calculate the similarity of term pairs.

23.4.4 Synsets Based Measures

The first attempt consists in replacing each terms by a list of its synets. A synset consist of words that have similar or identical meaning so they could be replaced by a synset identifier without loss in information content. The synset identifiers are then treated as words and the calculation of cosines based similarity measures follows the usual path.

The WordNet database includes different 202582 nouns and the number which are assigned to almost 80 thousand synsets what promises a great deal of unification. Unfortunately the popular words tend to belong to large number of synsets. Disregarding word-plays or jokes a noun has in a sentence one meaning so it should be replaced by one precisely selected synset. To achieve that word disambiguation techniques must be employed. Replacing words by synsets is akin to the traditional indexing.

23.4.5 Semantic Groups Based Measures

In this approach the clusters of semantically related words are created. The square matrix containing the WordNet word-word similarity was created and then the k-means algorithm was used to identify word groups. The algorithm was selected because is attempts to find the centers of natural clusters. The number of groups was set to 30 and the number of iterations to 50. The distant Euclidian distance was used as a distance measure. In the experiment the grouping was done by the Statistica package.

The clustering of poses significant computational problems and therefore following steps were taken:

- Only nouns present in the test data were taken into account;
- The words occurring less then 2 times were left out;

For the test data set the above steps were sufficient, in the real life application one can consider identifying the groups separately for nouns having the same head synset. If a word was not included in the similarity matrix then it was replaced by its base form.

23.5 Verification

The test data set contained about 1500 sentences that were extracted from 10 papers on two diverse subject areas: picture processing and the applications of WordNet. The size of the test data set is not large. It is due to the fact, that the experiment should mimic actual work at the second layer. The papers were rather lengthy so in the term of the sentence number they are equivalent of some 20-30 WWW pages retrieved by a generic search engine seems to be a sensible choice.

At the initial stage the text was preprocessed in order to eliminate non natural language elements such as tables, formulas or graphical data. At that phase the abbreviations were substituted by the full text and a dots appearing in floating point numbers was replaced by commas. Finally the text was cut into sentences. A sentences was a piece of text separated by dots, exclamation or question marks. In what follows the resulting text is known as a raw text.

During the study the raw text was transformed into sequences of: base forms, nouns, synsets identifiers, groups identifiers.

In order to pick out nouns the sentences were tagged by the tagger developed at the Stanford University [10]. The transformation into synset and group identifiers was described in sec. 4.4 and 4.5 respectively. The semantic similarity of words was calculated using the algorithm described in [11], the implementation remarks are in [12].

The following similarity measures were tested: simple cosine, cosine IDF, SimExt, SimBest, SimSum.

The first two measures were calculated for the raw text all 4 transformed sequences. The calculations ware accomplished using the methods adopted from the focused crawler project developed at the Indiana University [13]. The last three

similarity measures were used only for the nouns extracted from text raw text. The similarity measures were calculated by methods written for the purpose of experiment according the specifications from sections 4.1 and 4.2.

In what follows the combination used text type and similarity measure is called an environment. The total number of environments was equal to 13.

The aim of the evaluation is to find out which environment produces similarity values alike human judgment. In order to find it out a following experiment was conducted. At start a set of 5 sentences was selected. The sentences are used as queries. For each environment and each query the following steps were executed:

- Calculating the similarity between the any of the queries and all on the sentences in the test data set.
- Collecting an ordered sequence of 7 "best" sentences were collected, in what follows it is referred to as Eqs – environment query sequence.
- The set of all sentences occurring in at least in one of the Eqs was created.
- The a human expert was asked to eliminate from the above set the sentences that did not fit the query, the resulting set is denoted by EQS.
- The distance between the sentences selected human and EQS's was calculated using the below formula.

Let n denote the number of selected sentences in an Eqs and Eqs(i) the a sentence on the i-th position in an Eqs. The value Eqs(i) is calculated as follows:

$$Val \ (Eqs \ (i)) = \begin{cases} (n-i)^2, & Eqs \ (i) \notin EQS \\ 0, & otherwize \end{cases}$$

Val(Eqs) -the distance of Eqs to EQS is given by the following formula:

$$Val \ (Eqs \ ) = \left\{ \frac{\sum_{i=1}^{n} Val \ (Eqs \ (i))}{\sum_{i=1}^{n} (n-i)^2} \right.$$

The Val(Eqs) distance measure is equal to 0 if the environment was not capable of finding a single sentences accepted by a human expert and is equal to 1 if all of its sentences are accepted.

Two Fig. 23.1 and Fig. 23.2 summarize the results. The former one depicts the average of the Val(Eqs) for all considered environment query sequences while the latter shows the average query – sentence similarity level. The names attributed to individual environments are self descriptive.

The results clearly show the value of the well known cosine measure. With or without ids normalizing factor it achieves top results. The basic word processing – transformation to base word forms or picking out only nouns is useful. In the experiment such a processing has not influenced much the quality of retrieval but it tends to increase the similarity level. The not significant influence of the idf factor is probably due to the relative small size of test data set. The usefulness of more advanced word processing is less clear. The level of word – word similarity in cos_cluster, cos_idfcluster, simSum and simBest environments is clearly to high.
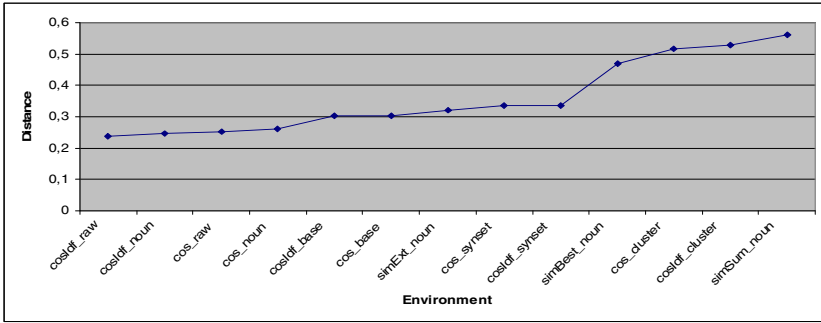
Fig. 23.1 Distances to human evaluation for different similarity measures.
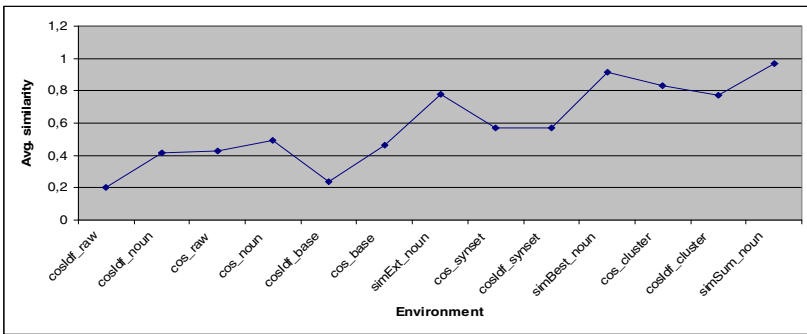
Fig. 23.2 The average similarity of sentences for different similarity measures.

As a result the level of query – sentence similarity is high and what inevitably leads to poor precision. The simSum environment was reported to achieved good precision but the experiment was conducted on a different type of data. The study dealt with the measuring the link texts and the text on a destination page. Link texts are short and therefore the broadening of a term sense was advantageous. The resulting decrease of precision was mitigated by a adaptive weighing schema.

The only one semantic algorithm to perform reasonably good is the simExtNoun. Its dissimilarity level is acceptable and the average similarity of result is very high.

Using such algorithms without word disambiguation algorithms is not justified. One should bear in mind that the computational effort necessary to compute the value of similarity e.g. for simExtnoun is far greater then any of the cosine based measures. Therefore combining word disambiguation with synset coding is probably the most promising attitude.

On the other hand the clustering of words according their semantic properties does not offer much ground for improvement. The number of clusters is an input parameter and it is known that its inappropriate choice may yield poor results. The used number of clusters was arbitrary selected and is most certainly not the best one but optimizing the number is does not look to be particularly promising.

The much hoped for word disambiguation technique can not be used in that case and one must not forget about the complexity of clustering sets of many thousand words.

23.6 Conclusions

The two layer retrieval system offers the user the possibility to better evaluate search results produced by a generic search engine. Its second advantage is its ability to incorporate in an automatic manner user preferences. The prerequisite for such a system is the availability of effective text similarity measures. The traditional work on the area dealt with either with comparing short queries with relatively long texts or clustering of such long texts.

The conduced experiment, albeit far from being extensive, shows the power of traditional cosine based measures. Evaluating the results we have to take into account not only the but also relatively low values of average similarity. It means that the measures ware capable of selecting the indisputably proper texts. For such texts the computer and human similarity judgment were identical. Such measures are far less productive in selecting texts that the similarity was not so obvious hence the low values of average similarity.

The other approach to asses the text similarity involved including semantic into the process. In general the semantic bases approaches are capable of selecting a far broader spectrum of texts but the precision of such retrieval is not satisfactory. It is assumed that to improve the precision the word disambiguation techniques have to be employed. The path is followed in the next stage of the research.

Another interesting area is the comparison of the inflected and uninflected languages [14]. The semantic similarity uses WordNet data base. Although the English WordNet has the highest language coverage but in recent years substantial work on creating WordNet data bases for inflected languages such as been done.

References

1. Siemiński, A.: The potentials of client oriented prefetching. Intelligent technologies for inconsistent knowledge processing. In: Nguyen, N.T. (ed.) Magill: Advanced Knowledge International, cop., pp. 221–238 (2004)
2. Cox, K.: A Unified Approach to Indexing and Retrieval of Information, DOC 94-10/94 Eanff, Albera, pp. 176 -181 (1994)
3. Cox, K.: Searching by Browsing. University of Canberra, Australia. PhD Thesis
4. Sherman, C.: Humans Do It Better:Inside the Open Directory Project, (July 2000), http://www.infotoday.com/online/OL2000/sherman7.html. (2000)
5. Manning, C., Raghavan, P., Schütze, H.: An Introduction to Information Retrieval. Cambridge University Press, Cambridge (2008)
6. Beall, J.: The Weaknesses of Full-Text Searching. The Journal of Academic Librarianship 34(5), 438–444 (2008)
7. Miller, G.A., Beckwith, R., Fellbaum, C.D., Gross, D., Miller, K.: WordNet: An online lexical database. Int. J. Lexicograph 3(4), 235–244 (1990)

8. Mihalcea, R., Corley, C., Strapparava, C.: Corpus-based and Knowledge-based Measures of Text Semantic Similarity. American Association for Artificial Intelligence, 775–780 (2006), `http://www.aaai.org`
9. Siemi ski, A.: Using WordNet to measure the similarity of link texts. In: Nguyen, N.T., Kowalczyk, R., Chen, S.-M. (eds.) ICCCI 2009. LNCS, vol. 5796, pp. 720–731. Springer, Heidelberg (2009)
10. `http://nlp.stanford.edu/software/tagger.shtml`
11. Seco, N., Veale, T., Hayes, J.: An Intrinsic Information Content Metric for Semantic Similarity in WordNet. In: Proceedings of the European Conference of Artificial Intelligence (2004)
12. `http://www.codeproject.com/KB/string/semanticsimilaritywordnet.aspx`
13. `http://www.informatics.indiana.edu/fil/is/JavaCrawlers/`
14. Piasecki, M., Szpakowicz, S., Broda, B.: A WordNet from the ground up. Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław (2009)

Chapter 24
Verification of Open Source Web Frameworks for Java Platform

Dariusz Król and Jacek Panachida

Abstract. A comparative analysis of the two most popular open source web frameworks for Java platform is done. The aim of the paper is to present modern software environments designed for implementing web applications and also to make final recommendations about web framework, which should be used in developing web application. The subjects of the analysis are Spring MVC and JavaServer Faces. The solution of the problem relies upon theoretical analysis of available framework features and upon empirical studies on implemented application designed to support managing a pet clinic.

24.1 Introduction

At the moment there exists over 40 web frameworks for the Java platform, e.i. Cocoon, Echo, JavaServer Faces, Maverick, Spring, Struts, Tapestry, Turbine, WebWork. The selection of appropriate framework is very difficult task. This paper can help to avoid potential problems.

The intrinsic quality of software system in a real world is its ability to evolve and adapt. The software is still modified to new requirements and new frameworks are created that can augment or enhance an existing system. Because of this, the programmers are in a dilemma, which one should be used. Existing software tools do not solve this problem because there does not exist the universal verification method. For each new framework we should develop new verification process.

In this study, we investigate the two most popular environments: Spring MVC and JavaServer Faces. It is clear that obtained results could not be direct transferred

Dariusz Król
Wrocław University of Technology, Institute of Informatics,
Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland
e-mail: `Dariusz.Krol@pwr.wroc.pl`

Jacek Panachida
Wrocław University of Technology, Faculty of Computer Science and Management,
Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland

to any other framework, but we believe that this method could be used in a similar fashion also to other systems from the list of existing web frameworks.

The remainder of this paper is structured as follows. Section 24.2 provides an overview of application performance study of three selected user groups. Section 24.3 discusses essential issues from the point of Java metrics. Templates and Trinidad components study are described in Sections 24.4 and 24.5. Finally, Section 24.6 concludes the presentation and provides final conclusions.

24.2 Application Performance Study

Performance study was conducted on final version of Petclinic application [9]. Each test involved three different group of users: users that mostly browse data- Group 1, users that execute tasks connected with updating - Group 2, and mixed group in proportions 4 to 1 of previous two groups - Group 3.

Table 24.1 Numbers of requests according to user group

Measurement	Group 1	Group 2	Group 3
Max. no of requests	478/190	263/172	317/184
Max. no of requests in 3 sec.	54/38	52/35	37/26

Numerator replies to Spring MVC, denominator – JavaServer Faces

Table 24.1 presents numbers of requests according to user group. Value in numerator describes result for Spring MVC [1, 12] and in denominator for JavaServer Faces [2, 6]. Data in first row apply for maximal measured requests per second, second for requests in fixed response time (3 sec.). In case of failed response test was finished.

The main observations are the following:

- For first user group there is significant difference. It results from nature of JavaServer Faces applications that attempt to mimic behaviour of desktop applications. Each request directed to JavaServer Faces application has cookie with the average size of 5kB. It has significant impact on performance. The cookie include information about state of page components.
- The second group consists of users that made operations with database usage e.g. add new visit or change owner's information. This type of operations effected in reduced performance, nearly two times. The decrease of requests is mostly noticeable for Spring MVC application that changed from 478 to 263. For JavaServer Faces values are almost the same. It probably result of cookie size and higher web server load due to database operations.
- The third group was created to simulate ordinary system usage. Most users only browse data applications. There was only small difference between second and last group. The reason is that database operations are far more resource consuming than common browsing actions.

- The second row presents similar results but for requests in fixed response time. As well in this case Spring MVC application was better, it has more requests per second versus JavaServer Faces application with the same group of users. In each group ratio of Spring MVC and JavaServer Faces requests are nearly the same and is equal to 1.4.

Table 24.2 Average response time according to user group

Measurement	Group 1	Group 2	Group 3
Max. no of requests	0,4/1,9	0,6/0,8	0,54/1,4
Max. no of requests in 3 sec.	0,25/0,5	0,12/0,27	0,19/0,36

Value of numerator replies to Spring MVC, denominator – JavaServer Faces

Table 24.2 presents average response time depending on user groups. Value in numerator describes result for Spring MVC and in denominator for JavaServer Faces. The biggest difference is for the first group to the advantage of Spring MVC framework. The lowest difference is for second group. Value for the third group various for about 50% for each framework [11].

24.3   Java Metrics Study

The aim of code metric measure [3] was comparison of projects size and quality of each project.

Table 24.3 Metrics of code for application packages

Metric	Data	Core	Spring MVC	JSF
LOC	1048	450	1195	977
LOC of methods	259	93	477	407
No of classes	23	10	29	21
No of interfaces	9	9	0	1
No of packages	5	3	7	7
No of methods	137	47	115	130
No of fields	29	20	29	41

Table 24.3 shows size of each project. Analysis shows also data for common modules such as data and core. Data module is responsible for data storage and core module for business logic.

The main observations are the following:

- Application created with Spring MVC has more lines of code than JavaServer Faces application. Difference in Method Lines of Code metric is only insignificant lower. The smallest module is core that does not have to implement complex business logic that is necessary for standard CRUD (create, retrieve, update, delete) application.

- Number of packages and interfaces is nearly the same. The exception is number of classes which is bigger in Spring MVC application. Again core module is the smallest. There is only one interface used for implementation of utility beans for JSF application.
- The only metrics that JSF has higher values is number of fields and methods. This is a result of usage component model in JavaServer Faces application. Each class of component is Plain Old Java Object (POJO), which encapsulate its state. Rich components encapsulate many fields to preserve their state. It results also in higher number of getter and setter methods per class.

Table 24.4 Metrics of code with functional classification of packages

Metric	VAL	WEB	UT	CON	LST	FORM
LOC	59/–	233/109	67/149	150/63	93/336	573/300
LOC of methods	34/–	91/31	30/68	64/28	32/156	223/121
No of classes	2/–	6/4	2/3	3/2	7/6	9/5
No of interfaces	0/–	0/0	0/1	0/0	0/0	0/0
No of methods	4/–	22/22	0/3	20/4	0/54	60/45
No of fields	0/–	7/4	0/1	7/0	7/17	15/19

Numerator replies to Spring MVC, denominator – JavaServer Faces

Another comparison takes into consideration functional aspects of application packages. These were divided into six groups. Code that is responsible for validation (VAL), code of controllers (WEB), code of utility classes (UT), code of converters (CON), code for display list of objects (LST) and code for handle forms (FORM). Table 24.4 presents metrics of code with functional classification of packages. Value in numerator describes result for Spring MVC and in denominator for JavaServer Faces.

The main observations are the following:

- JavaServer application does not have validators. All necessary validation is performed using Trinidad validation tags in web page templates.
- In case of utility classes Spring MVC is better as it has two times lower number lines of code. The difference is a result of code that was necessary for implementation of beans management and class for storage of used language version of application.
- Process of conversion of data was easier in JavaServer Faces application. There was no necessity for implementation of converter for list of medicines.
- Comparison of others values is hard because of difference in models used in each application. WEB module in case of Spring MVC application is set of traditional controllers. In JSF application role of controllers is concentrated in code of components, which has also other responsibilities. In JavaServer Faces application controllers from WEB group are facade for business logic code.
- Last element is code related to web forms. Approach of JSF application that is similar for desktop application results in two times smaller value of total lines of

code metric. In fact part of code that is responsible for form handle is located in LST group.

- Others metrics like number of children, depth of inheritance tree or number of overridden methods show some interesting places in applications. In Spring MVC application code of forms have the biggest value of depth of inheritance tree (average 8.78) and number of overridden methods (3.11). For JSF application average depth of inheritance tree is 1. Metric for number of children and number of overridden methods equals 0. This is result of lack of necessity of extending of class of JavaServer Faces framework.

Table 24.5 Complex metrics of code for Spring MVC and JavaServer Faces application

Metric	Spring MVC	JSF
Specialization index	1,862	0
Afferent coupling	0,857	2,429
Efferent coupling	4,143	2,714
Abstractness	0,016	0,036
Instability	0,829	0,664
Normalized Distance	0,187	0,3

The results of complex metrics are presented in Table 24.5. Each value is average of packages for whole application. We use the following metrics:

- Specialization index based on proportion of number of overridden methods and depth of inheritance tree. In connection with lack of overridden methods in JavaServer application value of this metric equals 0.
- Afferent coupling indicates on average packages responsibility. Code of application that was created with Spring MVC has less than two times responsibility in comparison to JavaServer Faces application. In Spring MVC application the biggest responsibility have utility classes and converters, in JSF application utility classes and controllers.
- Efferent coupling is the number of other packages that the package being measured is dependent upon. In Spring MVC the higher value of metric have code of forms and controllers, in JSF application code of controllers and lists of objects. The result of a subtraction between Afferent and Efferent coupling equals 5 for Spring MVC application and 1 for JSF application. It means that in Spring MVC application majority of packages depends on minority of other packages. Situation is different for JSF application, here values of Afferent and Efferent coupling are almost the same.
- Abstractness of both applications is comparable low. There was much more concrete class in comparison of abstract one.
- Instability as quotient of efferent coupling and sum of all coupling is bigger for Spring MVC application. The most instability is due to utility packages.
- Normalized Distance from Main Sequence represents the best balance between abstractness and stability of package. In this case better is Spring MVC as result of bigger instability with smaller abstractness.
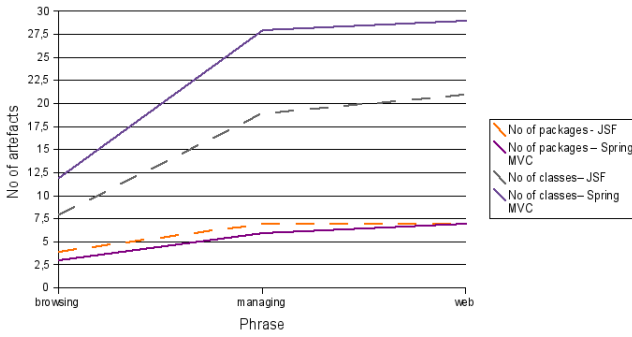
Fig. 24.1 Changes of code metrics during project development

Figure 24.1 presents changes in size of project during implementation of web module. Second phase related to addition of web forms made biggest changes in code base. Third phase connected with localization of application [10] was insignificant for both project sizes.

24.4 Page Templates Study

In order to provide a comparison, metrics for page templates were introduced. There are five types of views: paged list, details of object, form of object modification, form for adding visit (wizard style), simple list with fixed number of rows.

Figure 24.2 presents average numbers of tags of dynamic contents per view type. Application created with JavaServer Faces use more tags [5]. The biggest difference is in case of simple list that was implemented in JSF application with usage of paged list. Trinidad library does not have dedicated solution for simple types of list. Another difference is for paged list, list in JavaServer Faces application is more complex. The smaller difference was in case of modification form despite the fact of presents of validations tags of JavaServer Faces application.
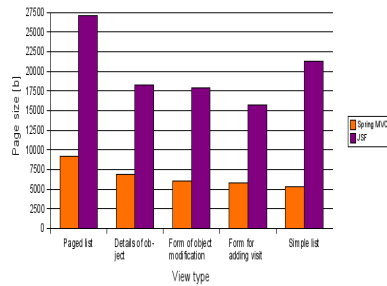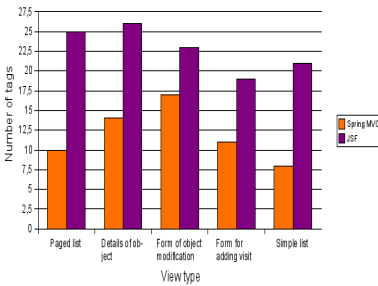
Fig. 24.2 Number of tags according to view type   **Fig. 24.3** The size of generated HTML page

Figure 24.3 presents the size of generated HTML page. The size of a page template is almost the same and equals about 7kB. But after transformation of dynamic tags size into HTML code differences increase four times. The biggest difference is in case of paged list but it is because of complexity of list component. The main reason in all types of views is result of additional JavaScript in JavaServer Faces application [7].

24.5   Accessibility and Maintainability Study of Trinidad Components

Accessibility [4] of Trinidad components was measured by validation of HTML code and by verification of correctness of functioning components in modern internet browsers. This evaluation apply only for visual components.

In the first test each browser obtained points that depended of correctness of functioning tested components. Two points if component was fully functional. One point if component was partly functional. No points if component did not work. Results are presented in Table 24.6. Figure 24.4 presents the correctness of components according to internet browsers. Components work the best in Internet Explorer browsers. The results for Firefox are slightly worse, but all components still work. There are only insignificant problems with chooseDate, inputData and selectBooleanCheckbox components. The worst was Opera browser. Three components (chooseDate, inputData and selectBooleanCheckbox) did not work. Another two (navigationTree and inputNumberSpinbox) had problems with proper rendering.

Table 24.6 Correctness of functioning Trinidad components according to Internet browser

Internet browser	No of points
IE 6.x	62
IE 7.x	62
Firefox 3.x	59
Opera 9.x	50

Fig. 24.4 Correctness of Trinidad components   **Fig. 24.5** Types of errors

Next test was performed using W3C validator for HTML 4.01 Transitional specification. During testing some components returned internal error with message "the server encountered an internal error() that prevented it from fulfilling this request". Component selectOneChoice was returning message "javax.faces.FacesException: SelectItem with no value". Finally in test took part 31 components.

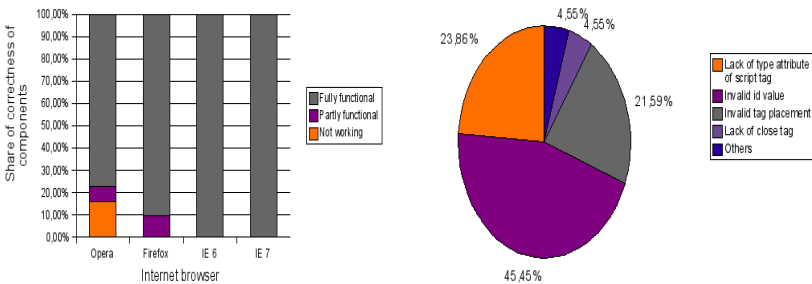During test it turned out that over 95% of errors belong to one of four group. Figure 24.5 presents percentage share of errors according to their types. The most common error applied for invalid value of id attribute, that starts with underscore or dollar sign. Second category was lack of type attribute in script tag. Third applied to invalid tag placement and last applied to lack of close tag for caption and table tags. Over 1/3 of components did not have errors, 13% had one error and 20% two errors. There was about 1/4 components with number of errors that include between three and seven and over 6% with over seven errors. Components with the most number of errors are inputColor with 19 errors, chooseColor with 13 errors and selectManyCheckbox with 7 errors.

Measure of maintainability [8] of application base on number of files that have to be added, modified or appended according to some new elements in application. Action add means add new file, modification means change content of some file that already exists and append means change file but without changing old content.

Table 24.7 Maintainability of Trinidad applications and number of changes regard to project files

Function	List	Form	Localization
Add	2/2	3/1	1/2
Complete	2/1	2/1	1/1
Modify	0/0	0/0	0/0

Value of numerator replies to Spring MVC, denominator – JavaServer Faces

Table 24.7 presents results of maintainability of Spring MVC (numerator) and JavaServer Faces applications (denominator). Measured values relate to add new paged list, web form and localization of application.

The main observations are the following:

- To add new paged list it is necessary to update (append) petclinic-servlet.xml, add new mapping for controller and a method handler. Then add new file with controller and implement handler. Next action involves creating of page template and adding new mapping of logical view name (file views.properties).
- In case of JSF application first declaration of managed bean must be done (faces-config.xml). There is no necessity for mapping any new declarations. Next is creation of managed bean implementation. Last actions involves creation of page template.
- In comparison of adding pages list, JavaServer Faces application require less work to achieve paged list. Only one modification is required and there is no need to define new URL mappings.

- Creation of web form in Spring MVC application involves as before adding mapping and new controller method, but additionally validator class is required.
- For JSF application there is only necessary to add new component method and add new page template. In case of simple forms there is not necessary to add custom validators. Often Trinidad validator tags are enough.
- Form creation using JavaServer Faces is easier. In Spring MVC application it is required to repeat steps known from previous action and add new validator. In JavaServer application is often enough to add new template and method.
- Localization of application in Spring MVC involves adding special beans to configuration file and adding message translation in separated files for different language. Language switching is done by beans.
- In JSF application first adding information about using languages is required. Next action involves adding files with message translations. Trinidad has out of box translation for various language. To add language switcher implementing new classes to code base is required.
- JavaServer Faces is better in message translation. Spring is better in case of language switching, JavaServer Faces doesn't have support for this feature.

24.6   Conclusions and Future Work

The selection of the appropriate framework is complex and important task of software programming. From our empirical study, we found that:

1. Spring MVC represents standard approach to developing web applications, well-known for development using Struts framework. JavaServer Faces uses component- and event-based model, similar to Microsoft ASP.NET.
2. Spring MVC based application have better performance than JavaServer Faces application. It is consequence of JSF components used in presentation layer.
3. Project made with Spring MVC was bigger in sense of code metrics of lines of code and other artifacts. Classes of JSF project had more methods and fields as a consequence of component model.
4. Measurement of complex metrics indicated better design of Spring MVC application in concern of package dependencies and OOP techniques.
5. Pages generated with JavaServer Faces are even four times bigger than similar pages of Spring MVC application.
6. JavaServer Faces framework is better in creation of web forms but worse in comprehensive localization of application.
7. Trinidad components work properly in most of the modern web browsers.

Although our study investigated only two systems, we think that the observations from the study provide reasonable basis for further validation process. Further research is currently being undertaken to extend the application by additional functionality such as reporting or authentication and to evaluate Open Source Ajax Frameworks: Google Web Toolkit versus Direct Web Remoting.

References

1. Arthur, J., Azadegan, S.: Spring Framework for Rapid Open Source J2EE Web Application Development: A Case Study. In: Proceedings of the Sixth International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing and First ACIS International Workshop on Self-Assembling Wireless Networks (2005), `http://dx.doi.org/10.1109/SNPD-SAWN.2005.74`
2. Bergsten, H.: JavaServer Faces. O'Reilly Media, Sebastopol (2004)
3. Chidamber, S.R., Kemerer, C.F.: A metrics suite for object oriented design. IEEE Trans. Software Eng. 20, 476–493 (1994)
4. Chisholm, W., Vanderheiden, G., Jacobs, I.: Core Techniques for Web Content Accessibility Guidelines 1.0 (2000), `http://www.w3.org/TR/WCAG10-CORE-TECHS/` (accessed June 1, 2010)
5. Chusho, T., Ishigure, H., Konda, N., Iwata, T.: Component-based application development on architecture of a model, UI and components. In: Proceedings of the Seventh Asia-Pacific Software Engineering Conference (2000), `http://doi.ieeecomputersociety.org/10.1109/APSEC.2000.896719`
6. Deugo, D.: Techniques for Handling JSF Exceptions, Messages and Contexts. In: Proceedings of the International Conference on Internet Computing. CSREA Press (2006)
7. Dunkel, J., Bruns, R., Holitschke, A.: Comparison of JavaServer Pages and XSLT: a software engineering perspective. Software–Practice & Experience 34, 1–13 (2004)
8. International Standard ISO/IEC 9126-1 Software Engineering Product Quality. Part 1: Quality model, Technical report, Geneva (2001)
9. Krebs, K., Leau, C., Brannen, S.: The Spring Petclinic Application (2007), `http://static.springsource.org/docs/petclinic.html`
10. Parr, T.J.: Web application internationalization and localization in action. In: Proceedings of the 6th International Conference on Web Engineering (2006), `http://doi.acm.org/10.1145/1145581.1145650`
11. Selfa, D.M., Carrillo, M., Del Rocio Boone, M.: A Database and Web Application Based on MVC Architecture. In: Proceedings of the 16th IEEE International Conference on Electronics, Communications and Computers (2006), `http://dx.doi.org/10.1109/CONIELECOMP.2006.6`
12. Seth, L., Darren, D., Steven, D., Colin, Y.: Expert Spring MVC and Web Flow. Apress, Berkeley (2006)

Part V
E-Learning Platforms

Chapter 25
E-Learning Usability Testing Platform

Adam Wojciechowski and Pawel Meller

Abstract. E-learning is a constantly developing manner of distant courses distribution. With a technological progress its effectiveness and usability controlling abilities has increased and should be measured in order to provide more efficient web solutions. In this paper, multi-modal, user tracking, web usability testing platform is presented. Presented platform provides Afterwards selected Moodle based e-learning implementation usability testing was performed revealing its drawbacks and misfunctionality.

25.1 Introduction

Contemporary still more and more universities are providing additional resources and courses using e-learning platforms. One of the most popular is a Moodle[1] platform. At the same time distant learning solutions must catch up with prevailing technological trends. It results with constant web project layouts changes and functionality reorganization. However it still remains a question concerning modification correctness and their influence on usability. This paper tries to summarize web usability testing process. At the beginning selected organizational rules have been distinguished and adequate testing methods were presented. Provided analysis is necessary to determine which elements of interface may be challenging for users, while the tests show, where exactly the problem is, and if suggested changes improve the usability or not. Approach towards usability testing is a complex subject and may differ among scientists. However when testing group and method are selected carefully the test should show how an average user will perform while trying to achieve a sample task. The tasks are circumscribed in a way, that would represent a list of normally performed actions that a user might want to execute on a web

Adam Wojciechowski and Pawel Meller
Institute of Computer Science, Technical University of Lodz, ul. Wolczanska 215, 90-924 Lodz, Poland
e-mail: `adamwoj@ics.p.lodz.pl,meller@stud.ics.p.lodz.pl`

[1] http://www.moodle.org

site, using problematic areas that were determined in analysis phase. During the years of research, new technologies and methods appeared to provide better analysis and testing possibilities. Some of them require usage of additional tools, that record user's action with computer, so it is possible to analyze the interaction more deeply. Several of these solutions include special applications, that not only provide recording options (such as recording of visited websites, mouse movement, etc.), but also make the planning and execution of the tests easier. Unfortunately, the market for such software products is still very small, so the available solutions may not be suitable for all situations. That is why, in some cases, it may be better to develop a dedicated application that would provide all the necessary functions, and will be more suitable for selected e learning tasks.

25.2 Web Usability Testing

Throughout all types of websites, starting from simple HTML pages up to Reach Internet Applications (RIA), developers and designers tend to create pages that would succeed in delivering information. According to Jakob Nielsen, web analyst and one of the pioneers of web usability, users visit website for its content and everything else is just the backdrop [7]. The design is there to allow the people access the content. Unfortunately the point of view of developers and normal users has often been completely different, which led to creation of websites containing useful information but presented in such a way, that none of the users can get to it. Therefore, the need of a tool, that can measure easiness the user can get, what has been sought, was present for a long time. In order to measure user experience the information, the user receives from a web page, should be grouped into several areas of interaction: first impression of a web site (content recognition), navigating through the web structure (navigational system) and determining the area, where the specified piece of information can be found (organizational system). Content recognition describes whether the content of a page is clear to the user, basing on the first look of the page. Despite the belief of developers and designers [5], users do not always scan through all content on a web page before giving up in most cases, the first impression is enough for them to judge, if they arrived on a page they were looking for [4, 9]. The necessary features of a homepage and ways of presenting them are varied. Jeff Johnsons list [4] can be used as a general guide for homepage planning. Additional details regarding graphical approach to a web page first impression can be found in Stobinska paper [9]. Due to huge amount of data, presented on a page, they should be organized in a special organizational scheme corresponding to the nature of the information. It is a hard task, because normally every user has his own sorting or labeling habits [6] . However since most of people use parallelly some ordinary source of knowledge (i.e. phone book) some standards can be assumed, that visitors are used to, and these standards can be applied to website's structure. Basing on [6] two main organizational groups can be distinguished: exact organizational scheme and ambiguous organizational scheme. First of them contains methods that are self-explanatory, based on data types being processed (e.g. alphabetical, chronological

or geographical). Ambiguous schemes are based on more natural ways of sorting
and provides better results for large amount of data. Implementing second scheme
requires creation of rules for sorting and grouping and is more difficult to execute
then than the first model. However, in some cases, the rules may lead to undecidable
situations, requiring conduction of some manual grouping or adjusting the rules. Ex-
emplary ambiguous organizational rules may concern: topics (like in yellow pages
where type of service is provided), tasks (like in Microsoft Windows menu, when
limited amount of high-priority tasks is anticipated), audience (where content can
be divided among group of users), metaphors (where level of abstraction is used for
representing the content like toolbar or system icons) or hybrid solutions. If con-
tent structure is organized, a problem in the way to a good website, is navigation
scheme. It answers the question where to place which elements of the structure, so
they will be easy to find. Peter Morville and Louis Rosenfeld [6] define two types
of navigation systems: embedded navigation system and supplemental navigation
system. Embedded navigation systems are placed inside web pages, providing con-
nections between content, whereas supplemental navigation systems are additional,
external web pages, such as indexes, sitemaps and search systems, containing an
overview of all pages that a web site contains. Embedded navigation system can
be further divided into three groups/subsystems: global, local and contextual nav-
igation. Every group represents different connections. First one provides access to
global structure, second one enables access to other pages in selected content group
and third one covers all hyperlinks to related content. All three groups should coexist
within a web page in order to visualize its navigation system. However, according
to Morville [6], these three major subsystems are necessary but not sufficient in
themselves. This is because not all users are willing to use the standard navigation
while trying to get information. In some cases (e.g. support product pages) it is eas-
ier for them to search information directly, rather then browsing through the struc-
ture. Then supplemental navigation systems (e.g. search engines) are the last chance
for the user when navigation structures fail. In this point it should be remembered
that traditional database searches are written that a single incorrect search term will
mean that you get no match [2]. That is why a good practice in such situation is
providing a suggestion system, which can help users to get the needed information.
Not only by correcting spelling mistakes but also enabling to narrow or broaden the
search results by adding or removing some keywords.

25.3   Testing and Measuring Web Usability

Web usability is a complex term, describing many aspects of user interaction and
thus is not possible to measure it directly and to express it with a single mark. Instead
of that, testing task can be divided into smaller parts, each concerning other part of
user interaction, making it easier to analyze and provide precise results. Tests can be
divided into two main groups: user based and expert based. First group contains all
tests that are performed by a group of users providing useful information about how
they understand the website and how they are able to interact with it. When dealing

with user based tests, choosing the right participants and preparing correct instructions for users are very important. Afterwards the method of performed tests should be selected. Test participants can be engaged in a personal survey, group interview or individual usability testing [2, 8]. The last, individual approach, seems to be the most valuable, since many user's actions can be captured, analyzed and processed. The standard types of tests performed with users can be found in [2, 10]. Whereas at early stages of project development, it can be cheaper and easier to choose tests performed by an expert. Besides the result can be influenced by the knowledge of the expert, it requires less preparation and provides satisfactory results. Some popular expert tests are presented in [2, 10, 11]. After setting up the testers, testing schedule should be prepared. Before it is handed to the user, several variables need to be defined, i.e. [3, 1]: scope, purpose, schedule and location, participants, scenarios, questions, data to be collected, hardware and software requirements and additional people roles. Whereas in surveys or expert usability reviews the data is already in proper format, in usability tests users actions have to be measured and referred to the task or scenario. Metrics that describe such relation are called performance metrics [10] and consist of five basic types: task completion, time required to complete the task, mistakes made during tasks completion, efficiency describing effort expended during task completion and finally learnability, expressing changes of performance. Regardless of the metrics the test results should be measured and captured. In such cases think aloud protocol [2], videotaping or special hardware/software recording is used for data collecting. All of described methods and metrics, that involve participants, testing the product by means of computer, require some ways of data acquisition. It is always possible to sit next to the participant and take notes or videotape him during the test, but there are also more complete systems, that provide some additional functions like: click tracking, mouse tracking, eye tracking, log analysis, screen recording or browser recording. All of them try to capture some of the communication between a user and a computer. There are many tools that provide one or two of mentioned functionalities, but most of them are not free. Additionally it was impossible to find a solution that would provide all or most of client side possibilities. That is why a dedicated multichannel application was designed and developed. Its details are presented in the next section.

25.4   User Activity Tracking Platform

Testing web usability with users requires special software recording their actions. A special testing platform was designed mainly for the subject of Moodle usability testing, but with some minor tweaks, it can be easily used for other web-based or application GUI-aimed testing purposes. The application is written in C # 2.0, so it requires a Windows based machine, with .NET framework installed. These are the only limitations, that currently can delimit the appliance of this plat-form. The application provides a simple graphical interface, which enables easy configuration of all its options. Its GUI is shown in figure 25.1.
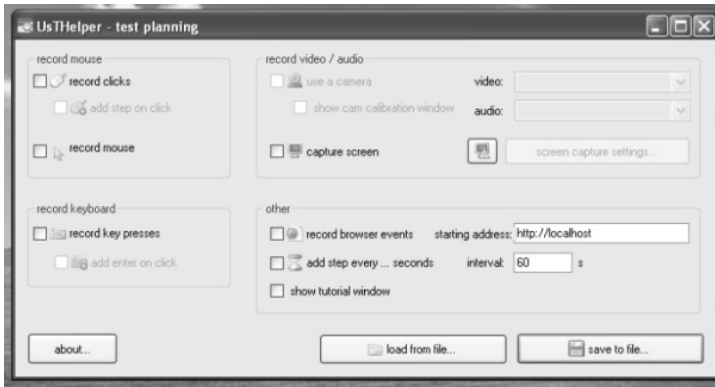
Fig. 25.1 Testing platform interface

Set of the features of the application include browser events recording, mouse movement and click recording, keyboard recording, screen capture and audio/video (webcam or normal camera, compatible with Microsoft DV) recording. Since the recording is started by the user (using an automatically created desktop shortcut), time measurement can also be performed and provide additional, accurate timing data. The ease of use and deployment was very important, to ensure the possibility of concurrent testing of many users at the same time - testing users usually involves deploying the application on many computers, so any additions improving this process are always handy. For this reason, the application provides an xml based system for storing setting files and uses the desktop shortcuts. This not only enables the supervisor to easily run the same tests on different machines, but also provides easy access to modification file (XML), if necessary. Further-more, the application provides a simple wizard, that can provide the user with some introduction before the test. The instructions are automatically loaded from an external XML file and can be easily configured to contain as many steps, as necessary. The user is informed about the progress of the test using Windows native balloon tips (fig. 25.2a), so he knows exactly, when the test begun and when his work is completed. Additionally all aspects of the application are transparent to the user, not distracting his attention.

The application minimizes the number of actions required as input from the user e.g. the recording session is finalized by closing the browser window. Such improvements are more natural and reduce the risk of e.g. fault timing being record-ed, when the user does not remember or know how to close the application. All the collected information is useless, if it can not be read and processed. Application outputs two types of data: session overview and xml test log. First type is less detailed, but provides a good overview of the recorded session. It contains a list of all pages, that were recorded, along with accompanying screenshots of the visited pages. Recorded mouse movement and clicks are overlayed on the screen-shots, providing an additional source of information about the user activity (fig. 25.2b). The application also provides access to the data in an untouched version, by returning an XML file with all data. The XML file's structure makes it easy to navigate within steps, and
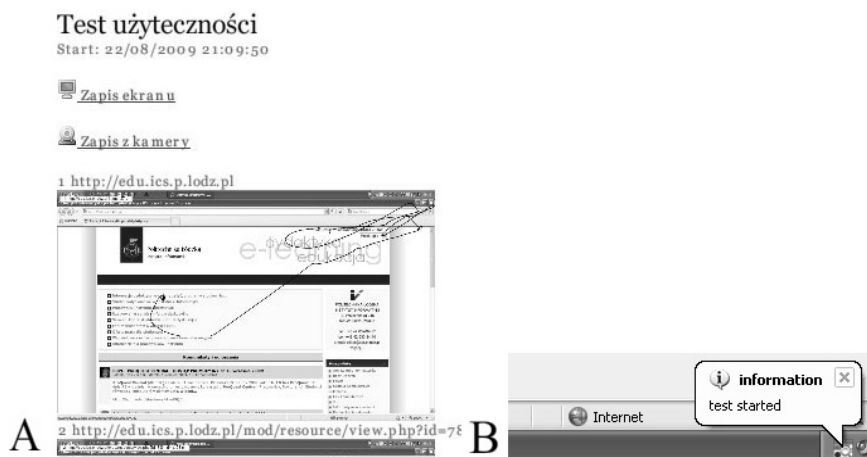
Fig. 25.2 a) Sample session summary, displaying session date and time, links to recorded videos, and a list of pages, that were visited during the session b) Sample message displayed by the application

possible to import the collected data to an external program. The size of files was also an important issue, especially when a camera was used to provide user recording (even 3 MB per second of recording), so the application had to perform video compression.

25.5 Tests

The goal of the test was to evaluate Institute of Computer Science e-learning page (http://www.edu.ics.p.lodz.pl) by means of developed application. The testing phase included three testing scenarios. Every scenario was created for one unique usability issue, which was determined at the phase of website analysis. Each issue was connected with different aspect of usability first one covered content recognition, second inline navigation (local navigational systems), and third one concerned supplemental navigational systems. For each of the problems, two versions were tested. First version was a copy of the original web site. Having identified the problems, solutions for them have been created and applied, creating the second tested version with suggested improvements. Research by Nielsen and Molich [11] and Virzi [10] showed, that number of test participants should be between 5 and 10 people to discover correspondingly 80 and 90 percent of problems. Carol M. Barnum provides [1] a good summary of these two theories stating that 7 ($\pm$ 2) people is the mantra for structured writing and other methods for organizing information, whereas 5 ($\pm$ 2) participants is the mantra for the number of participants needed in a usability test. Considering number of tests, versions and group count, a total number of 30 users were examined. The participants were chosen from first year students, as they were inexperienced moodle e-learning platform users. During the test, all

Fig. 25.3 Text styles used for first scenario a) page with scarcely marked hyperlink, b) page with strongly marked hyperlink

participants were using the same browser, screen resolution and operating system. The scope of first scenario was to test the contextual navigation. The original web page uses a style, where hyperlink text looks exactly the same as normal text, making it impossible to distinguish one from another (fig. 25.3a). In a first version participants were presented a sample page, which contained a number of lines of text with inline hyperlink, which they were supposed to visit. As a result, in first tested version, the hyperlink, placed within a text, was very hard to notice. The goal of the test was to read provided text and eventually find and click mentioned hyper link. The main variable, that was tracked, was the time needed for task completion. The participants were given additional clues (the part of the page, where the link is located) after 100 seconds, so they could finish the task. The tests have shown, that participants using the first version needed an average of 134.8 seconds to finish the task, while the users, that were presented the modified version (fig. 3b), needed an average of 49 seconds. Since the task required finding a link on the main page and clicking on it, the total number of steps, that should be made by user, is two. The number of steps for task completion was also recorded, and showed, that the modified version required only 2 pages for task completion, while the users from the first group needed from 2 to 7 visits before finally reaching the right page . Second scenario was aimed at content recognition. The participants were presented the home page of the Institute e-learning platform (first group original page (fig. 25.4a); second group - modified version (fig. 25.4b)). After analyzing its con-tent basing on the first impression (browsing through the structure was not allowed) they were asked to complete a survey containing questions about the con-tent of the page. Set of questions is provided below.

1. Does the website provide information about the kind of content, that is available on it?
2. Does the website provide resources for studying?
3. Does the website provide information about exams timetable?
4. Does the website provide information about how to contact your lecturer?
5. Does the website provide information about how to apply to study Computer Science?

Each of the five questions expected answers yes, no or don't know and were then rated, giving +20% for each good answer, 20% for wrong answer and no change for 'don't know' answer. First version, which had no, or very little information about the page, showed an average of 8% success rate, while second version, which had information about the content, resulted in 48% success rate. Third test was targeted

Fig. 25.4 Institute homepage tested in second scenario; a) original; b) modified

Fig. 25.5 Course assignment page; a) original; b) page with visible search;

at a simple task that all users were performing on yearly basis - course assignment. Course selection page displays the list of courses, grouped by categories, but with no automatic sorting sample page is shown in figure 5a. This resulted in the list being difficult to navigate, and what was more surprising, no search system was available on this page, even though it was enabled on other pages. Similar to previous approach, half of the users were showed the first version with no modifications and search disabled, the second half of users were shown a version, that included search and auto completion. An average time of course enrollment without auto completion was 81 seconds while for the supported version it was 63 seconds.

While the first and second scenario generally proved, that the proposed changes have improved the usability of tested platform, the third scenario brought some surprising results. Both versions showed similar average time required to complete the task. After reviewing the screen captures (similar to figure 25.2b), that were recorded by the application during each test, a reason for this similarity has been found. Some users ignored the absence of search on the first page, and used alternative browser's search instead. On one hand, it proves the importance of search being available, and that the users rely on it, but on the other hand, it shows, that sometimes users have their own methods for dealing with common usability problems. During the tests, application proved its usefulness, giving easy access to all data, that was recorded. Especially in third scenario, the use of screen recording was very important for the right analysis of the user's actions. Not all activities were fully analyzed, some of them, like mouse movement and clicks, can provide additional information and lead to even more successful problem solving, but were not relevant to the subject of the tests. A great field for further improvement is webcam analysis, that could extend the data collected by the application from only what user does to also what user perceives.

25.6   Conclusions and Future Work

The web usability analysis, presented in the paper, not only provides well organized procedures for usability evaluation, but helps in suggesting valuable improvements as well. Moreover, in the paper, there is introduced specialized soft-ware solution that helps in users interaction tracking. Newly constructed application can monitor keyboard and mouse changes as well as record simultaneously user body, voice and face activity. Usually mentioned functionality was delivered by a set of commercial applications, whereas spectacular possibilities of our solution, let test very complicated interaction process by means of one dedicated platform. In our research, provided software was used for testing our Institute of Computer Science e-learning platform revealing and afterwards confirming certain drawbacks. Multiple modes of user tracking helped to analyze and understand some inaccuracy that was hard to solve basing on only one tracked activity. Testing platform, worked out in our research, is not limited to e-learning websites, but tested aspects have just proved its functionality. According to authors it can be directly used for other web usability tests. Finally it must be remarked that popularity and accessibility of web usability testing tools, in which presented paper obviously helps, can make the web content more accessible and less vulnerable to mess and disorganization.

References

1. Barnum, C.: Usability Testing and Research, Longman, New York (2002)
2. Brinck, T., Gergle, D., Wood, S.D.: Usability for the Web - Designing Web Sites that Work. Morgan Kaufmann, San Francisco (2001)
3. Dumas, J.S., Redish, J.C.: A practical guide to usability testing. Intellect Ltd. (1999)

4. Johnson, J.: Web Bloopers - 60 Common Web Design Mistakes and How to Avoid Them. Morgan Kaufmann, San Francisco (2003)
5. Krug, S., Black, R.: Don't Make Me Think! A Common Sense Approach to Web Usability. Que (2000)
6. Morville, J.: Information Architecture for the World Wide Web. O'Reilly, Sebastopol (1998)
7. Nielsen, J.: Designing Web Usability. Peachpit Press, Berkeley (1999)
8. Baron, R.S.: So Right It's Wrong: Groupthink and the Ubiquitous Nature of Polarized Group Decision Making. In: Zanna, M.P. (ed.) Advances in experimental social psychology, vol. 37, pp. 219–253. Elsevier Academic Press, San Diego (2005)
9. Stobinska, M., Pietruszka, M.: The effect of contrasting selected graphical elements of a web page on information retrieval time. JACS 17(2), 113–121 (2009)
10. Tullis, T., Albert, W.: Measuring the User Experience - Collecting, Analyzing, and Presenting Usability Metrics. Morgan Kaufmann, San Francisco (2008)
11. Wharton, W., et al.: The cognitive walkthrough method: a practitioner's guide. In: Nielsen, J., Mack, R. (eds.) Usability Inspection Methods, pp. 105–140

Chapter 26
E-Learning in Teaching the Object Oriented Programming

Jerzy Kisilewicz

Abstract. Developed e-learning materials and their use in teaching object-oriented programming course are presented in this paper. Presentations, films, laboratory instruction, multimedia textbook and handbooks are available to students via e-learning system Moodle. The training quizzes, quizzes for laboratories and the examination are created by Moodle individually for each student. All quizzes use the questions of common databases. These quizzes are resolved remotely and they are automatically evaluated by the Moodle platform.

26.1 Introduction

Computer equipment and, in particular the universal access to the Internet pose a new, qualitatively different learning opportunities. T. Küchler [3] says that e-learning has become a significant element of modern teaching and is seen as an important criterion for evaluating the activities of universities. Research conducted by M. Striker and K Wojtaszczyk [7] confirm the great interest of teachers in using a variety of e-learning platforms though, how writes G. Penkowska [4, 5] characterizing the Polish e-learning, the development of the e-learning courses is particularly labor-intensive.

At many universities, many of the e-learning courses were created. A. Rybak and W. Półjanowicz [6] described the concept of education in the three e-learning courses. They emphasize the large interest of the students on the topic. J. Krzyżek [2] has described three programs to develop teaching materials and he has characterized the Moodle, how it can be useful to teach mathematics.

The aim of this paper is to describe the developed e-learning materials such as presentations, quizzes, videos, instructions and multimedia handbooks, as well as

Jerzy Kisilewicz

Chair of Computer Systems and Networks, Wroclaw University of Technology

Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland

e-mail: `jerzy.kisilewicz@pwr.wroc.pl`

the aim is to present, as these materials have been used in teaching the course *Object-oriented programming* in the Technical University of Wroclaw.

The author of this paper teaches object-oriented programming for approximately 400 stationary students and approximately 150 part-time students. The stationary students are divided into three groups and lecture is repeated three times. In parallel with the lecture, students must pass laboratory exercises carried out by different teachers for groups of 15 to 30 students. E-learning is not an independent form of the education, but it supports the regular classes.

The developed e-learning materials are helpful to both students and teachers. These materials facilitate the assimilation of knowledge to students and allow to examine the degree of mastery of knowledge, minimize the effects of possible absences of part-time students at the university, because they allow students for distant learning. These materials help teachers in the classroom, help in monitoring the current progress of students and help in the final evaluation of students. The developed materials create the system of mutually complementary and cooperating items, made available through e-learning platform Moodle. Moodle is a free software. Moodle is often used and it is easy to use [1].

26.2 Materials for Students

General information materials, multimedia presentations for lectures and for laboratories, and training quizzes are made available to students through e-learning platform Moodle.

General information materials:

- the conditions for credit course and assessment principles,
- a list of primary and secondary literature,
- the rules and the scenario of the examination quiz,
- the topics of subsequent lectures and exercises.
  Multimedia presentations for lectures and for laboratories:
- PowerPoint presentations with audio explanations,
- instructional videos for laboratories,
- instructions for laboratory exercises,
- a multimedia handbook.
  Training quizzes:
- the quizzes to the lectures, containing the guidance comments to answers.

The conditions for credit course specify the necessary formal requirements for the student, to receive the credit course. They determine also the minimum knowledge required to receive a positive evaluation. This document specifies all the detailed conditions for getting credit and it gives the principle of determining the final grade within the course.

The rules and the scenario of the examination contain the information on: the number of questions, time to complete the answers, the rules for scoring the

responses and the threshold to pass the examination. This document includes: what the student must have each other when solving the test, what he can to do and what is him not allowed to do. This document also includes the practical advices on how to efficiently solve the test, what the student must prepare before the test, and what he must do after closing the test.

The schedule of the lectures contains for each of the lecture: the topic and the date of this lecture, the file with the presentation to this lecture, the instruction for the laboratory exercise, the optional instructional video and the training quiz for this lecture, or quiz for this and several previous lectures.

The PowerPoint presentation contains the slides presented during the lecture. The texts of the explanations, are very brief. The audio explanations are added to the presentation, synchronized with the presented animations. Adding sound to a presentation, significantly increases the size of the file. The size of selected presentation files with sound and without sound are included in Table 26.1. Students can print the images contained in this presentation and during the lecture to focus on teacher explanations. They can write an own comments, without spending time redrawing illustrations.

Table 26.1 The size of selected presentation files with sound and without sound

Presentation No	The size of the file without the sound	The size of the file with the sound
1	1.11 MB	10.1 MB
2	1.06 MB	8.3 MB
3	1.90 MB	15.7 MB

Instructional videos show step by step, the implementation of selected tasks. These tasks are different than those provided for the implementation of the lab, but are similar and cover the same issues. They are the moving films, that show the computer screen which was recorded when you create a program or the selected parts of the program. These videos contain audio commentary with explaining what you do.

Films were made by capturing the screen of the monitor and audio commentaries using a free program CamStudio v.2.00. The capture parameters were optimized in order to get the smallest files, even at the expense of a significant deterioration in the quality of the recording. The file sizes were minimized, to reduce the time of remote access to them via the Internet, because these films are made available to students via Moodle. For parameters: Capture Frames Every 200 ms, Playback Rate = 5 frames/second, and Set Key Frames Every 100 frames, the obtained files have the size of the order of 3 to 8 MB per minute recording (average of 5 MB/min). The size of selected files and corresponding recording times are contained in Table 26.2. *The multimedia handbook* is a guide to e-learning materials. Handbook presents material that is required, to pass the course. This handbook

Table 26.2 The file sizes and the recording times, for the selected films

Film No	File size	Recording time	File size / minute
1	28.3 MB	5 min 46 sec	4.908 MB/min
2	36.6 MB	4 min 52 sec	7.520 MB/min
3	21.2 MB	3 min 12 sec	6.625 MB/min
4	21.9 MB	4 min 22 sec	5.015 MB/min
5	31.5 MB	6 min 33 sec	4.809 MB/min
6	32.6 MB	8 min 53 sec	3.670 MB/min
Average	28.7 MB	5 min 36 sec	5.117 MB/min

Table 26.3 The quizzes and the numbers of questions

Quiz No	Number of questions	The topic of the quiz
1	20	Basics of C programming language
2	10	Basics of C++ programming language
3	15	Classes and their methods
4	10	Derived classes, inheritance, polymorphism
5	10	Event-driven programming for Windows
6	10	UML, exception handling, templates, preprocessor

is dominated by the descriptions, examples, explanations and interactive links to the multimedia documents such as presentations, videos, quizzes, etc.

The training quizzes are being solved by the students via the Internet under the control of Moodle. The test questions are divided into categories. The questions which are assigned to one category relate to specified subjects. Moodle creates an unique test each time using the random questions, with the fixed categories. The tests are designed so that no more than half of all questions from the category are chosen. Therefore the student must resolve the quiz several times to learn about all the questions. The list of quizzes, including the random numbers of questions, contains Table 26.3. The list of categories gives Table 26.4. For each category the table contains: the category subjects, the number of questions in this category, and the number of questions chosen for the examination test and for the training quizzes.

The scenario of the training quizzes is similar to the scenario of the examination-test, that students can truly master the technique of solving the test. The assessment criteria are the same for all tests. The time limit allowed for a question is several times greater for the training quiz. Solving the training quiz, student can repeatedly accept the answer to some question, and each time he receives an assessment of the responses and sometimes he gets the suggestions for possible investigations. The student can solve one training quiz many times and a test of examination he can solve only once.

Table 26.4 The categories and the numbers of questions

No	Topic	Total	Exam	Q.1	Q.2	Q.3	Q.4	Q.5	Q.6
1	Basics of C programming language	41	2	12					
2	Structures and pointers in C	18	2	8					
3	Basics of C++ programming	24	4		10				
4	Classes and their software	34	5			15			
5	Derived classes and inheritance	21	3				7		
6	Polymorphism and abstract classes	10	3				3		
7	Programming for Windows	18	2					10	
8	UML	7	2						4
9	Exceptions, templates, preprocessor	12	2						6
	Total:	185	25	20	10	15	10	10	10

26.3 Materials for Teachers

The following materials have been developed for the lecturer and for teachers working in laboratories:

- the multimedia presentations for lecturers,
- the tests allowing to enter the laboratories,
- the examination test,
- scoring rules and points counting rules for laboratory,
- the spreadsheet for calculating the final grade for the course.

The multimedia presentations for lecturer contain the same graphic content, which contain the presentations for students. These presentations does not contain the sound, because during the lecture the graphic is explained by the lecturer.

The tests allowing to enter the laboratories are developed by teachers working in laboratories. These tests use a common database of questions for training quizzes and for examination. Teachers can also add their own questions to the quizzes.

Students solve *an examination test* in laboratories, remotely over the Internet. During the test, the identity of students is checking, as well as independency of solving. Moodle gives a random set of questions for each student. The questions are divided into thematic categories, and the random system draws the specified number of questions from each category. The examination test and the training quizzes use the same database of questions. Any suggestions about the accuracy of answers are excluded. After completing the answers, student shuts test or this test will be closed after a timeout. After closing the test, the student automatically receives an information about the acquired points and he receives a grade for test. Student can no longer change the closed test.

The database of the questions includes only those questions whose answers can be automatically evaluated. These are questions such as: multiple choice, single choice, matching questions, true/false questions, numerical questions. In case of the multiple choice questions, more points are assigned to the basic answers. In

case of the incorrect answers student gets the negative points. To the fundamental error is assigned more points. To pass the test, you must obtain at least 60% of possible points.

The scoring rules for the laboratory shall determine the number of points awarded to student for his achievements. These principles are the same for all teachers and all student groups. The maximum number of points that a student can receive for the laboratory is equal to the maximum number of points per examination test.

The spreadsheet for calculating the final grade for the course contains the rules converting the points to the final grade for the course. This grade is determined by the sum of points for the examination, laboratory and other achievements, such as attendance at lectures. These rules allow you to easily set the thresholds so that the number of grades awarded: A (5 and 5+), B (4+), C (4), D (3+) and E (3) were successively 10%, 25% , 30%, 25% and 10% of all positive grades.

26.4 System of the e-Learning Materials

The course *Object-Oriented Programming* consists of lectures and laboratory exercises that are subordinated to the lectures. To receive course credit, you must receive credit, both lecture and laboratory exercises. Grade of these two credits in the form of points obtained are entered into a spreadsheet. Into this sheet are entered some other student achievements affecting the final evaluation, e.g. the number of attending lectures. The final grade is calculated using the rules stored in a spreadsheet.

The presentation is created for each lecture, to is use by a lecturer. The presentation supplemented by audio explanations, is posted in Moodle available to students.
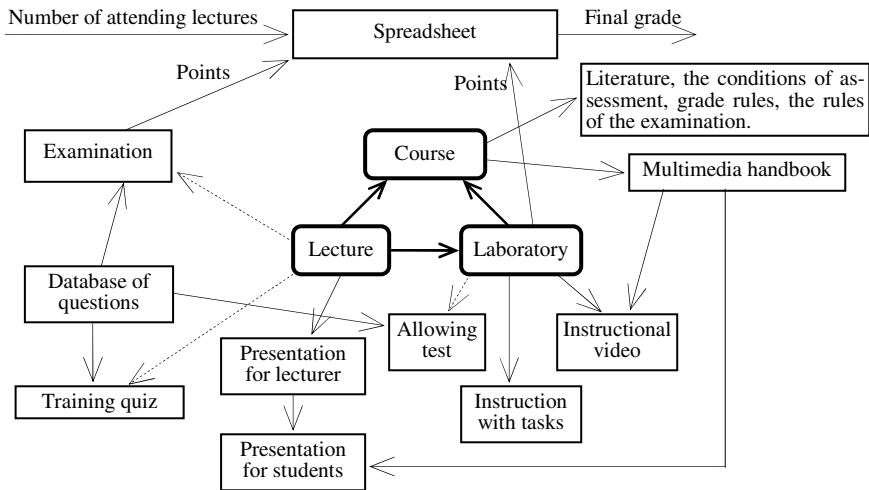

Fig. 26.1 The organization of the course *Object Oriented Programming*

The multimedia handbook refers to this presentation. With a single lecture or with a group of 2-3 lectures is associated category of questions and the training quiz. Students can solve this quiz many times.

A practical illustration of the lectures are laboratory exercises. For each exercise is worked out the instruction with a set of tasks. For selected tasks there are developed instructional videos. Instructional video shows the process of creating applications similar to the application in a given exercise. The multimedia handbook refers to these videos.

Concerning the relationship between the course materials, lectures, laboratory exercises and the quizzes are shown on Figure 26.1.

26.5   Conclusion

E-learning system which supports the teaching of object-oriented programming is used already almost 3 years and it is popular among students. About 97% of stationary (full-time) students and about 95% of part-time students use the available on Moodle presentations and other materials. About 48% of students use with the e-learning materials regularly at least once a week and approximately 77% of students at least once every two weeks. Almost all the students solve the training tests. Most students solve each test several times. In this academic year the stationary students solved each training test an average of 7 to 8 times, and the part-time students solved an average of 5 to 6 times.

Until 2007, the students had solved the sets of tests, which were printed on paper. In the year 2008, the existing questions have been entered to Moodle and the e-test was introduced, which is generated individually for each student. With the introduction of the e-tests, the students changed the approach to these tests. The theft of quizzes, the trade in set of questions and the trade in templates of answers, have become meaningless. Students began to practice in solving the training quizzes. As a result, in comparison with the years 2006 and 2007, in the year 2009 the number of students who did not pass the examination fell almost tripled from 22.6% to 8.6% for stationary studies, and fell about six times from 69% to 11% for the part time studies.

References

1. Cole, J., Foster, H.: Using Moodle, 2nd edn. O'Reilly Community Press, Beijing, `http://issuu.com/iusher/docs/usingmoodle2` (accessed April 9, 2010)
2. Krzy ek, J.: Tools used to create teaching aids in support of e-learning in school. E-mentor 33 (in Polish),
   `http://www.e-mentor.edu.pl/`
   `wersja_do_druku.php?nmer=33&id=711`
   (accessed April 8, 2010)

3. Küchler, T.: The effective implementation of e-learning projects in higher education:
    two case studies. E-mentor 21 (2007), (in Polish),
    ```
    http://www.e-mentor.edu.pl/
    wersja_do_druku.php?numer=21&id=465
    ```
    (accessed April 8, 2010)
4. Penkowska, G.: Polish e-learning in expert opinions part I E-mentor 20 (in Polish),
    ```
    http://www.e-mentor.edu.pl/
    wersja_do_druku.php?numer=20&id=433
    ```
    (accessed April 8, 2010)
5. Penkowska, G., Polish e-learning in expert opinions part II. E-mentor 21 (in Polish),
    ```
    http://www.e-mentor.edu.pl/
    wersja_do_druku.php?numer=21&id=457
    ```
    (accessed April 8, 2010)
6. Rybak, A., Półjanowicz, W.: The concept of training students in the field of e-learning.
    E-mentor 31 (in Polish),
    ```
    http://www.e-mentor.edu.pl/
    wersja_do_druku.php?numer=31&id=674
    ```
    (accessed April 8, 2010)
7. Striker, M., Wojtaszczyk, K.: Barriers to implementing e-learning on the example of
    university, Part 2. E-mentor 32 (in Polish),
    ```
    http://www.e-mentor.edu.pl/
    wersja_do_druku.php?numer=32&id=693
    ```
    (accessed April 8, 2010)

Chapter 27
Analytical Framework for Mirroring and Reflection of User Activities in E-Learning Environment

František Babič, Ján Paralič, Peter Bednár, and Michal Raček

Abstract. This chapter deals with evaluation of user activities and their participation in collaborative processes realized within the supporting virtual environment. The main goal behind proposed solution is to provide standalone package with all necessary functionalities to obtain data from examined virtual collaborative system in form of logs. For this purpose the following has been designed and implemented: a repository with predefined log format representing source historical data for analyzes; supporting middleware services for proposed analytical approaches; end-user tool for time-line based mirroring and analyses of user activities. These basic functionalities are extended with possibility to extract various summative statistics [10] about performed user activities and possibility to export data in predefined format (e.g. MS Excel) for analyses in third party tools as e.g. IBM SPSS Modeler. Described analytical framework has been designed, implemented and tested mainly within KP-Lab System that represents new interesting application in the domain of virtual environments or e-Learning systems. Proposed architecture was designed as generic platform with possibility to be integrated with other systems such as Moodle or Claroline in order to reflect different user practices.

27.1 Introduction

Reflection and mirroring of performed user activities in virtual environment is important step in complex evaluation of user behavior for different purposes, e.g. research or education. The first research direction in this case covers investigation of

Ján Paralič
Department of Cybernetics and Artificial Intelligence,
Technical University of Kosice, Slovakia

František Babič and Peter Bednár
Centre for Information Technologies, Technical University of Kosice, Slovakia

Michal Raček
PÖYRY Forest Industry Oy, Finland

interesting patterns of behavior that can be used as "best practices" or "worst practices" in future work. We analyzed the necessary functionalities for effective realization of collaborative processes and designed suitable technological solutions as supporting integrated analytical platform. The second research direction is evaluation of virtual environment usage in education conditions in order to analyze student's effort, quality and level of participation engaged in their activities.

The main motivation behind proposed analytical framework is to provide necessary functionalities for these two directions as standalone package with open communication links to various virtual or e-Learning systems. Typical e-Learning system offers virtual user environment with various integrated end-user functionalities for realization of collaborative activities based on selected learning goals. These activities consist of user's actions around shared objects of interest in different formats as documents, videos, learning packages, wiki pages, etc. All these objects evolve in time, so it is important to monitor and store performed events in order to obtain source historical data for analytical purposes.

This chapter presents the whole architecture of proposed solution, its integration within various collaborative environments and some partial results of its continual evaluation. The content consists of four main sections and starts with introduction, which briefly presents motivation and describes related works. The second section deals with technical issues as architecture, log format and proposed end-user functionalities. The third section describes proposed analytical approaches and the last one provides conclusion.

27.1.1 Related Work

The analytical tools based on data from virtual or e-Learning systems offer different views on performed activities, user interactions and used resources based on various users' expectations and requirements for analysis. Three main approaches can be identified in current research and practice. (1) Export of historical data from examined system in order to analyze data using the third party tools. (2) Integration of existing analytical procedure directly into the system. (3) Design and implementation of own analytical approaches. We followed the last alternative, whereas we can recall also to the actual state of the art in this domain represented mainly by the following research lines.

- Educational Data Mining [2] represents utilization of suitable data mining methods for education purposes based on their adaptation to relevant conditions, e.g. classification methods, association rules or clustering.
- Process mining provides functionalities for extraction of potentially useful information from event logs that are obtained from investigated system. The first main step is to extract a process model from an event log, the second one represents detection of discrepancies between a modeled process (created process model) and an event log - actual process status [11]. The main representative of this direction is application called ProM[1].

---

[1] http://prom.win.tue.nl/tools/prom/

- Muehlenbrock and Hoppe [8] propose activities in virtual environment as a basis for a qualitative analysis. This approach has been termed as action-based collaboration analysis and was implemented as a plug-in component in the generic collaborative system. Integrated plug-in provides features for monitoring user actions and analyses focused on reasons of conflicts or revisions identification based on a plan recognition approach.
- Semantic Spiral Timelines (SST) represents an interactive visual tool aimed at the exploration and analysis of information stored in investigated collaborative learning systems [4]. It provides an interesting way of presenting the events in form of spiral timeline that contains sequences of color-coded events.
- GISMO is based on Information Visualization methods [3] and provides usage analysis in virtual learning systems through suitable graphical representation [7]. This visualized representation offers only basic information about users and their activities, detailed description is missing. GISMO is implemented as a module on the visualization layer in Moodle, so all analysis can be performed by teacher himself directly in the system.
- Interesting and similar approach to the proposed solution is described in [12]. This approach covers constrain-based analytical approach for pattern discovering, i.e. defining filters during the pre-processing phase that reduces the search space; constraints during the mining phase (association rule mining, sequential pattern analysis, clustering and classification) accelerate and control pattern discovering. Constraints at the evaluation phase make obtained patterns user friendly and simpler to evaluate. The results are visualized through intuitive graphic charts and tables in order to make discovered patterns easy to interpret.

Described analytical approaches were identified within detailed state-of the art analyses in relevant research domains and used as inspiration for design of proposed analytical framework. Inputs to this phase contained also expectations and requirements from potential end-users. We also took into account possible and available technological solutions, as well as some ideas for future progress. As a result, four design goals have been defined and implemented prototype has been evaluated with respect to them. 1) Support for explorative analysis of collaborative learning or working processes; 2) Taking into account the external events (often very important to be taken into account when analyzing collaborative process); 3) Support for multiple perspectives and comparisons of different groups working e.g. in the same course; 4) Meaningful and comprehensible visual presentations of analyzed data that should easily be customized to the information needs of various categories of users.

27.2 Technical Description of the Analytical Framework

Proposed analytical framework was designed and implemented as standalone package that contains the following modules (architecture of possible integration is showed on Fig.27.1):

- Generic logging services on middleware layer for obtaining data about events in the virtual environment.
- The obtained data are stored into separate database called awareness repository in predefined format. The event database was implemented in MySQL database in order to provide fast and simple access with sufficient scalability.
- This source of historical data is used through analytical services and the results are presented within presentation layer.
- Presentation layer of this solution is represented by end-user tools that can be integrated directly in the environment or used standalone.
- These basic functions can be extended with information from content or user repository in order to obtain data about users and their profiles; or information about all existing versions of the content.
- Communication between repository, middleware and virtual environment is implemented through web services based on principles of Service-Oriented Architecture (see also [5]).
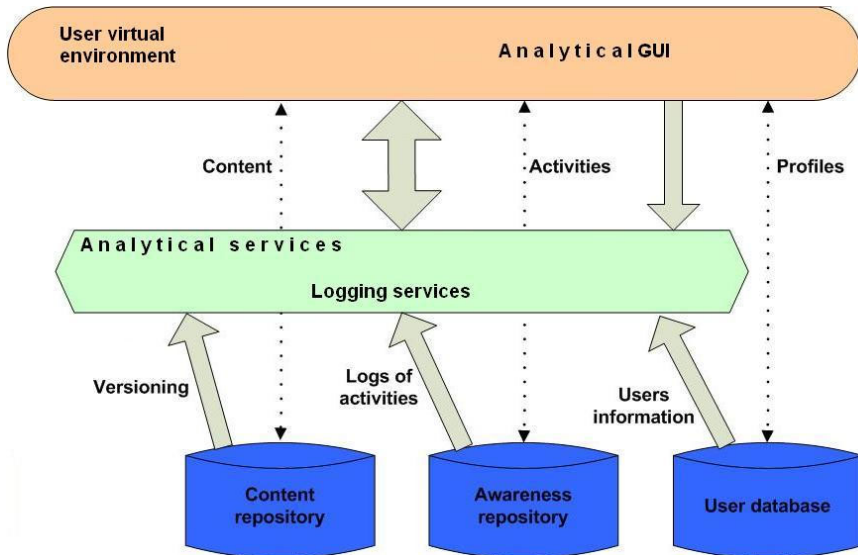

Fig. 27.1 Architecture of the proposed integration approach with examined system

27.2.1 Data for Analyses as Log of Events

The main motivation aspect behind proposed log format was to provide complex, simply extendable information source. Log format can be modified (adjusted) based on several criteria such as inspected system, user expectations or requirements, i.e. to add new parameters or eliminate existing ones. Actual version of log format that was used for testing and evaluation purposes is described below:

[ID, Type, Actor, Actor Type, Actor Name, Entity, Entity Type, Entity Title, Belongs to, Time, Custom properties, Custom data]

- *ID* – unique identifier of the log entry;
- *Type* – a type of performed actions, e.g. creation, modification, deletion, etc;
- *Actor* – unique identifier of actor that performed given event;
- *Actor Type* – user role that is delegated based on relevant part of the user environment;
- *Actor Name* – user name obtained from user management based on his system logging information;
- *Entity* – unique identifier of the shared object that motivates given event;
- *Entity Type* – type of shared objects, e.g. task, document, link, wiki page;
- *Entity Title* – concrete title of related shared objects
- *Belongs to* – unique identifier of relevant part of user environment where this event was performed;
- *Time* – time when the event was logged into database (represented in the following format: year-month-day HH:MM:SS);
- *Custom data and properties* – these parameters are used in situation when end-user application will store some properties or data that are typical for it.

Presented log version is the second one and was designed as sufficient generic format.

27.2.2 Experimental Integration with Selected Systems

Generic architecture and log format were experimentally tested within selected representatives of virtual or e-Learning systems as KP-Lab System, Moodle, Claroline and Dokeos (further details in [1]).

In the case of KP-Lab System, logging mechanism was connected on the middleware layer to obtain data from user environment in cooperation with monitoring services on this level. The experiments realized within Moodle cover the design and implementation of a new web service that is responsible for communication with internal Moodle logging system to obtain data from its internal database based on required log format. Integration with Claroline system was oriented to updates and changes in internal Claroline logging API to extend existing functions. This approach was not successful by reason of low support for possible extensions of this API, insufficient documentation of the whole system and necessary changes in source code. In the case of Dokeos, new event handler for monitoring purposes was created and connected with proposed logging services of analytical framework.

Realized experiments declared generic character of proposed solution and its potential for future usage in education or research conditions.

27.3   Analytical Approaches

The main motivation behind proposed analytical approaches is to provide complex view of different activities realized in virtual user environment which meets the design goals described in section 27.1.1 above. It means that designed and implemented technical framework provides several basic functionalities for simple usage by students and teachers, but also some advanced level of features for researchers that need deeper level of understanding. For these purposes two various approaches were designed, implemented and currently are evaluated. The next section will describe only one analytical approach called time-line based analyses that can be primarily used for mirroring and reflection. This approach to the analysis of user activities or the whole processes is to consider them as a series of different actions in a chronological order, possibly with different levels of granularity, where some subsets of them may have crucial importance. Such carefully (manually) selected subsets of actions are called *critical patterns*. These patterns usually lead to some critical moments, which can mean, for example, a significant progress, discovery of new knowledge/approach, or on the other hand they may indicate non-success of a particular process or its early finish. Such kind of patterns may also conceptually represent interesting learning paths emerged within particular user activities – either being positive (something like best practices), or negative (worst practices).

27.3.1   Time-Line Based Analyses

The simple example for time-line based analyses (TLBA) is to evaluate the whole learning course. The course contains several students' groups with given learning goals. Each group has its shared space in virtual user environment with possibility to use all available tools. The course starts with customization of related virtual space and creation of initial plan of the whole process with defined responsibilities, decomposition into simple tasks, identification of relevant resources and definition of time constraints. Created plan is dynamic so it is possible to change it based on newly emerged conditions, i.e. to identify necessary changes on-the-fly in order to improve it. The activity of each group is based on daily activities in virtual environment as creation or modification of shared objects, communication through synchronous or asynchronous tools, upload of related resources, linking them, annotating, commenting, etc. The course will finish with presentations of final group products with some potentially interesting findings. The next step will be evaluation of group or individual activity during the whole course. This evaluation will be based on monitored and stored data that will be now visualized (see Fig. 27.2) by teacher as sequences of student actions with descriptions and relations to particular shared objects on selected time-line. Created time-line presents work progress of each group or individual student in defined time period. Alternatively, the analytic tools can be used during the course by teachers as well as students to reflect on on-going activities and make corrective actions, if necessary.

The main functionalities provided in the time-line based analyses are the following.

- Sequences of events are visualized in chronological order via defined (one or more parallel) time-line(s). The time-lines represent visualization of all interactions and relations based on user requirements; and provide possibility to filter the list of showed properties for each visualized element.
- Possibility to define and store interesting patterns (pattern is a set of selected events or elements) from the time-line through suitably specified set of constraints (conditions).
- Basic time-line visualization consists of automatically collected events that are performed in monitored system. In some situations it is necessary to include elements called external events (performed outside monitored system) that are relevant to analyzed process. This operation is performed manually by the user.
- Possibility to add comments to events on the time-line (both automatically logged as well as external events, added manually) is another way how to support multiple perspectives on analyzed processes.

Design and development of these functionalities were lead by first design goal, where openness of shown data (means to filter, highlight, comment, etc.) brings new possibilities to infer information and make sense of data collected. With regard to the second design goal, TLBA users are able to add external events to the timelines to indicate that some outside event has taken place and are in relations with internal events performed in virtual environment. Despite TLBA basic functionalities described above that undertake partly third design principle as well, TLBA encourages users to design and share their own analytic queries by building, searching and sharing event 'patterns', by which third design goal is met. A pattern is represented as a set of events where different restrictions can be placed upon them, i.e. time constrains, performed by user, multiplicity etc. TLBA visual appearance is meant to be easy to grasp and simply to understand, so it would serve large group of users both occasional and frequent, which is in line with the fourth design principle.

27.3.2 Evaluation within KP-Lab System

KP-Lab System represents modular and flexible collaborative system with several integrated tools for different working or learning practices, e.g. operations with shared objects, modeling of various types of processes, virtual meeting support, visual modeling functionalities, integration of some Google Apps as Calendar or Docs, usage of semantic features as vocabularies or tagging and many others. Detailed information about this system are described in Lakkala et al., 2009 and Markkanen et al., 2008. The system was designed and is still being developed within EU IST project called KP-Lab[2].
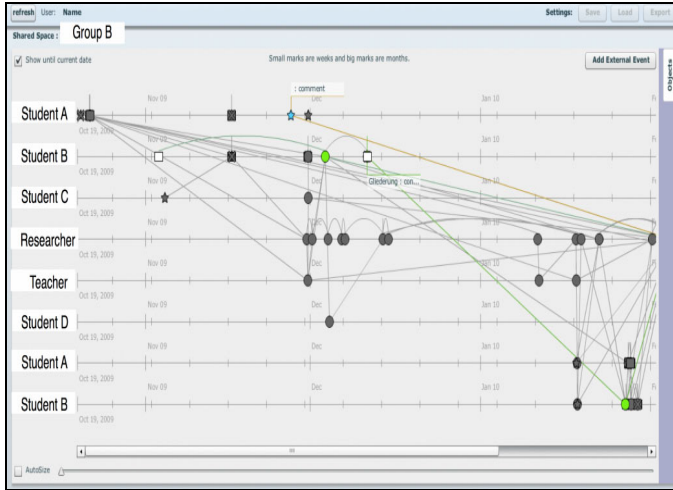
---

[2] http://www.kp-lab.org/)

Fig. 27.2 Overview of time progress within selected shared space representing one learning course

In this case the analytical framework has been integrated on the middleware layer in order to obtain historical data describing performed events in KPE, user virtual environment of the whole KP-Lab System. Current extent of awareness repository represents more than 80 thousand log entries of events from KPE.

Evaluation procedure within KP-Lab System consists of several phases that resulted into actual prototype (see Fig. 27.2) available on[3]:

- Experimental evaluation of first prototype that mainly consisted of logging services, awareness repository and basic analytical features. These experiments were oriented mainly on technological solution, integration with the whole system, correctly implemented logging services on the side of end-user tools.
- Identification of problematic aspects as quality and consistency of stored historical data, performance issues within first iteration of pilot courses and field trials (Norway, Finland, Netherland or Austria). Quality of the obtained data is very important condition for successful realization of analytical procedure.
- Updates in relevant implementation based on positive or negative findings in order to improve offered functionalities. This phase was oriented mainly on bug fixing on middleware layer based on previous testing.
- Second iteration of user's evaluation, mainly oriented on getting familiar with offered end-user features and obtaining meaningful and helpful knowledge. Results of this evaluation step represent several usability reports that describe identified problems on the presentation layer.
- Continuous improvement of actual prototype with the aim to provide stable and effective solution ready for future use in real conditions.

---

[3] http://2d.mobile.evtek.fi/shared-space/

Time-line based visualization (see Fig. 27.2) provides time overview of performed user activities within selected shared space with relevant shared objects and their description. The whole overview is organized around individual users and various types of activities are distinguished with different graphic shapes, i.e. creation as star, opening as circle, and modification as rectangle.

27.4   Conclusions

Mirroring and customized reflection of performed user activities during learning or working collaborative processes represent important features for general evaluation or deeper research investigation. Proposed analytical framework represents standalone software package with all necessary features on repository, middleware and presentation layer. Implemented analytical approaches provide various views on obtained data to support and improve mainly teacher's or researcher's work with possibility to select which type of analysis they want to use: simple graphical visualizations or complex time-line visualization.

Evaluation procedure can be divided in the technological and usability part. The aim of technological experiments is to test generic aspect of proposed solution, its potential for integration with various virtual or e-Learning systems and performance issues. The usability testing is oriented on quality of stored historical data, its consistency and advanced value for users, visualization on the presentation layer and general usability for different task in real conditions. Actual prototype is continually improved mainly on the usability side.

References

1. Babic, F., Wagner, J., Jadlovská, S., Leško, P.: A logging mechanism for acquisition of real data from different collaborative systems for analytical purposes. In: SAMI 2010: 8th International Symposium on Applied Machine Intelligence and Informatics, Herlany, Slovakia, pp. 109–112. IEEE, Los Alamitos (2010)
2. Baker, R.S.J.D.: Data Mining for Education. In: McGaw, B., Peterson, P., Baker, E. (eds.) International Encyclopedia of Education, 3rd edn. Elsevier, Oxford (2009)

3. Card, K.S., Mackinlay, J.D., Shneiderman, B.: Readings in Information Visualization, using vision to think. Morgan Kaufmann, Cal (1999)

4. Gomez-Aguilar, D.A., Theron, R., Garcia-Penalvo, F.J.: Semantic spiral timeline as a support for e-learning. Journal of universal Computer Science 15(7), 1526–1545 (2009)

5. Habala, O., Paralic, M., Bartalos, P., Rozinajova, V.: Semantically-aided Data-aware Service Workflow Composition. In: Nielsen, M., Kucera, A., Miltersen, P.B., Palamidessi, C., Tuma, P., Valencia, F.D. (eds.) SOFSEM 2009. LNCS, vol. 5404, pp. 317–328. Springer, Heidelberg (2009)

6. Markkanen, H., Holi, M., Benmergui, L., Bauters, M., Richter, C.: The Knowledge Practices Environment: a Virtual Environment for Collaborative Knowledge Creation and Work around Shared Artefacts. In: Luca, J., Weippl, E.R. (eds.) Proceedings of ED-Media 2008, World Conference on Educational Media, Hypermedia and Telecommunications, Vienna, pp. 5035–5040. AACE, Chesapeake (2008)

7. Mazza, R., Milani, C.: Exploring usage analysis in learning systems: gaining insights from visualisations. In: Workshop on usage analysis in learning systems at 12th international conference on artificial intelligence in education, Amsterdam (2005)

8. Muehlenbrock, M., Hoppe, U.: Computer supported interaction analysis of group problem solving. In: Proceedings of the Computer Support for Collaborative Learning (CSCL) 1999 Conference, pp. 398–405. Stanford University, Palo Alto (1999)

9. Lakkala, M., Paavola, S., Kosonen, K., Muukkonen, H., Bauters, M., Markkanen, H.: Main functionalities of the Knowledge Practices Environment (KPE) affording knowledge creation practices in education. In: O'Malley, C., Suthers, D., Reimann, P., Dimitracopoulou, A. (eds.) Computer supported collaborative learning practices: CSCL 2009 conference proceedings, pp. 297–306. International Society of the Learning Sciences (ISLS), Rhodes (2009)

10. Parali , J., Babi , F., Wagner, J., Simonenko, E., Spyratos, N., Sukibuchi, T.: Analyses of knowledge creation processes based on different types of monitored data. In: Rauch, J., Ra , Z.W., Berka, P., Elomaa, T. (eds.) Foundations of Intelligent Systems. LNCS, vol. 5722, pp. 321–330. Springer, Heidelberg (2009)

11. Van der Aalst, W.M.P., et al.: ProM: The Process Mining Toolkit. In: Proceedings of the BPM 2009 Demonstration Track, Ulm, Germany. CEUR-WS.org, vol. 489 (2009)

12. Zaïane, O.R., Luo, J.: Towards evaluating learners' behaviour in a Web-based distance learning environment. In: 2nd IEEE international conference on advanced learning technologies, ICALT 2001 (2001)

Chapter 28
The Paradigm of Screencasting in E-Learning

Marek Kopel

Abstract. This paper focuses on producing screencasts for e-learning purposes. Creating screencasts by students in order to document their work in completing a task may be a good alternative to traditional tests in asynchronous learning. Presented in this paper experiment concerns analyzing student screencasts by extracting the media metadata. Metadata such as video resolution, fps or audio sampling frequency are compared in order to find most popular parameters of screencast created by regular screencasters, in this case represented by students.

28.1   Introduction

The term "screencast" was coined in 2004 at Jon Udell's blog [6]. Readers of the blog proposed terms an then in voting chose one. For the definition of "screencast" let us use the explanation from Jon Udell who motivated people to name this experience in [5]:

> A screencast is a digital movie in which the setting is partly or wholly a computer screen, and in which audio narration describes the on-screen action. It's not a new idea. The screencaster's tools–for video capture, editing, and production of compressed files–have long been used to market software products, and to train people in the use of those products. What's new is the emergence of a genre of documentary filmmaking that tells stories about software-based cultures like Wikipedia, del.icio.us, and content remixing. These uses of the medium, along with a new breed of lightweight software demonstrations, inspired the collaborative coining of a new term, screencast.

As mentioned in the quotation the idea of screencasting is not new. One may see its origination in video conferencing tools. Microsoft NetMeeting which was bundled with Internet Explorer 3 was available for every Windows user since 1997. Using MS NetMeeting as described in [3] without any additional software or system

Marek Kopel
Wroclaw University of Technology, Wybrzeze Wyspianskiego 27, 50-370 Wroclaw, Poland
e-mail: marek.kopel@pwr.edu.pl

configuration allowed desktop sharing. Desktop sharing means live streaming of a remote desktop view, usually in addition to streaming a remote camera video showing a conference participant. At the same time a cross-platform solution for desktop sharing was developed: Virtual Network Computing (VNC) presented in [2].

Although desktop sharing may also be called screencasting the concept described by Jon Udell differs from desktop sharing by a few assumptions. First one is taking control over the remote desktop. In VNC solutions its most uses concern working on a remote computer via GUI. If the experience named screencast is compared to filmmaking then taking control over a desktop may be compared to directing. In this case remote directing by a viewer of a documentary would be something unnatural.

Another assumption made for desktop sharing and not necessarily true for screencast is the aspect of real time. For remote controlling live streaming of the desktop view is mandatory. Since screencasts are mostly meant for asynchronous use they are usually offered as Video On Demand (VOD). Publishing screencasts in services like YouTube or Vimeo allows its viewers more flexible use, especially when the screencasts present educational content, like tutorials or how-tos. As a result of the non-real time aspect screencast can be edited offline just like any show that is not a live transmission.

28.2 Educational Aspects of Screencasting

Publishing desktop actions usually has an educational purpose. Educational aspects of desktop sharing were discussed in [1]. Recording those actions along with a narration for later presenting is what screencasting is about. It allows asynchronous learning which is usually an attribute of e-learning. E-learning based on screencasting not necessarily has to concern software. Although animation of user actions on a desktop–as an extension of static screenshots–is invaluable in software bug tracking, tutoring, concept demos and helpdesk services, screencasts may also be used in any other discipline than computer science. Recorded actions over an editor may present problem solving for later consideration. Be it a text, equation or graphics editor the solved problems may be independent of their digital(computer) nature of presentation. Usually, instead of an editor, an interactive whiteboard is used. It minimizes the problem of using keyboard and mouse for people used to pen and paper or chalk and blackboard. One of most known sets of screencast tutorials is Khan Academy [4]. It supplies a free online collection of more than 1,400 videos on mathematics, chemistry and economics. The videos consist mostly of texts, equations and diagrams handwritten in real time with "color pens" on a black background. Of course each video includes a narration of Salman Khan explaining the concept being written. This gives a virtual experience of listening to a teacher at a blackboard during a lecture.

On the other hand screencast may be recorded by students to document their work. In 2009 at Wroclaw University of Technology making screenshots or

recording a screencast was proposed to student while doing their Digital Media Processing course tasks. The tasks concerned using media editors. In case of graphics, students used editors like GIMP or Photoshop for activities like retouch or photomontage. In case of sounds, editors like Audacity were uses for mixing audio recordings. Documenting their work with screenshots/screencast students offer more insight into their work than just the final result. When grading the work teacher provided with such documentation may asses student's knowledge of used tool (editor). It gives also a better chance for proving independence of student's work.

An interesting relationship between screenshots and a screencast comes out while using them as documentary media. To match screencast's documenting details screenshots need to be annotated and edited to make up for the animation of user actions they are lacking. Therefore producing documentary screenshots takes more effort than just pressing 'record' in case of a screencast. On the other hand human perception works faster for static images. Screenshots allow faster overlooking and focusing on interesting parts, which is different from the synchronization with timing of a screencast playback. Thus screencasts are easier to produce than screenshots, but they are harder to consume (analyze).

28.3 Experiment

During mentioned earlier Digital Media Processing course students were given a task to produce a screencast. The only limitations were: the screencast's length–circa 5 minutes and its format–ogg with theora and vorbis compressions. The latter requirement was inspired by the conformance of codecs used by browsers supporting HTML5 <video> tag [7]. Publishing screencasts in ogg container with theora compression for video and vorbis for audio is as easy as writing a simple HTML code shown in Fig. 28.1. The experience of playing them is as good as at any video site like YouTube or Vimeo. The main difference is that no additional player of Flash plugin is needed. In case of ogg the browser serves as a player.

Metadata were extracted from produced screencasts using Python library Hachoir. Example metadata extraction gives the following output:

```
>hachoir-metadata asd.ogg
Common:
- Duration: 5 sec 128 ms
- Producer: Xiph.Org libVorbis I 20070622
- Producer: VLC media player
- MIME type: audio/vorbis
- Endian: Little endian
Audio:
- Channel: stereo
- Sample rate: 44.1 kHz
- Compression: Vorbis
- Bit rate: 128.0 Kbit/sec
- Format version: Vorbis version 0
```

```
Video:
- Image width: 800 pixels
- Image height: 600 pixels
- Pixel format: 4:2:0
- Compression: Theora
- Frame rate: 15.0 fps
- Aspect ratio: 1.01333333333
- Comment: Quality: 12
- Format version: Theora version 3.2 (revision 1)
```
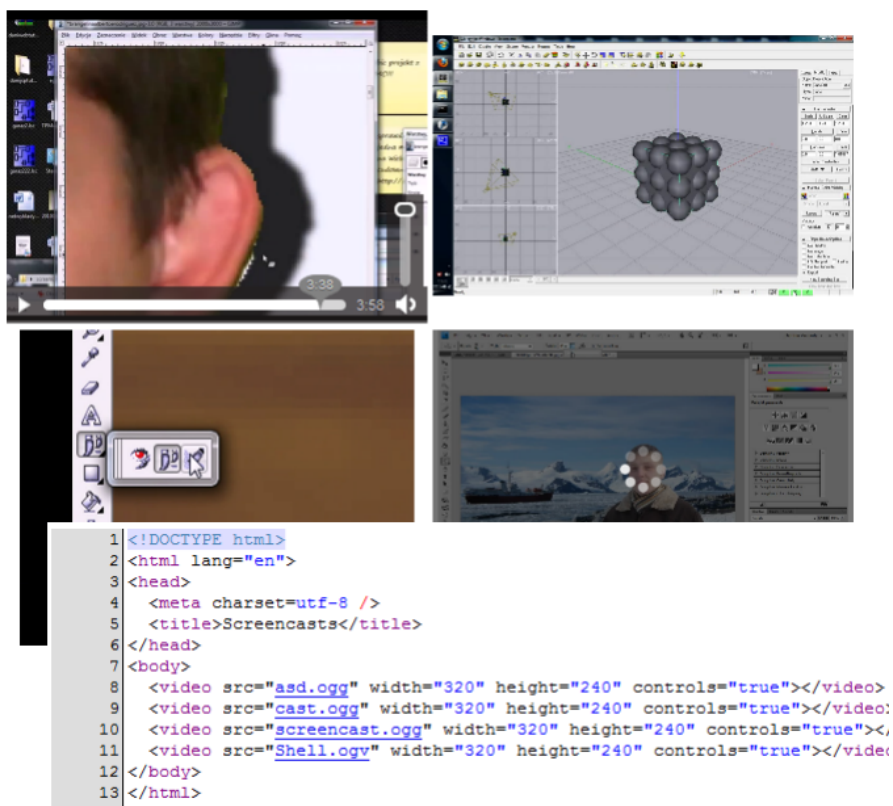


```
 1  <!DOCTYPE html>
 2  <html lang="en">
 3  <head>
 4    <meta charset=utf-8 />
 5    <title>Screencasts</title>
 6  </head>
 7  <body>
 8    <video src="asd.ogg" width="320" height="240" controls="true"></video>
 9    <video src="cast.ogg" width="320" height="240" controls="true"></video>
10    <video src="screencast.ogg" width="320" height="240" controls="true"></video>
11    <video src="Shell.ogv" width="320" height="240" controls="true"></video>
12  </body>
13  </html>
```

Fig. 28.1 Screencasts displayed in Firefox (playing the upper left one, buffering the lower right one, pausing the other two) and the page's source code using HTML5 `<video>` tag

Not all metadata parameters were present for all screencast files. Also they could come in different order. Simple bash and awk scripts were written to automatically convert the extracted metadata to CSV (comma-separated values) format. The scripts also added files' sizes to the metadata for analysis.

28.4 Results and Discussion

First analyzed parameter of screencasts produced by students is video resolution. Although students were instructed to use as low resolution as needed they often used their full screen resolution for normal work as their screencast resolution. Still, as shown on Fig. 28.2, the most frequently used resolution was VGA, i.e. 640x480. Also most of the resolutions used were lower than XGA, i.e. 1024x768, but 30% of them used horizontal resolution (width) of SXGA, e.i. 1280. Sometimes height of those SXGA wide screencasts was lower. This was the result of cutting off from video Windows taskbar or a title bar of a window in full screen mode. One thing that is in favour of using full screen resolution is zooming. Tools like Camtasia Studio allow setting, automatic or post processing, zooming of selected parts of the screen. e.g. a dialog box that just poped up. Zoomed part of screen becomes more clear and discards all aspect of the screen that did not allow focusing on subject. However zooming was rarely used and most of the high resolution screencast were blurred and unreadable, especially when presented in lower resolution e.g. on a digital projector.
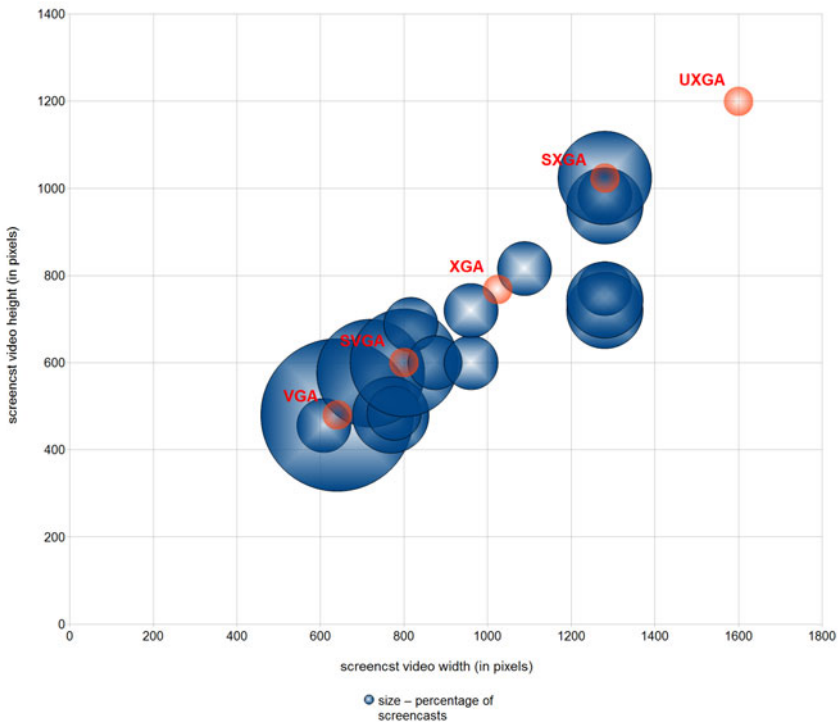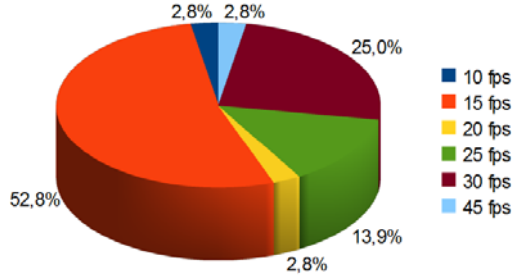

Fig. 28.2 Screencasts' resolutions. Blue bubble's position indicates the resolution used in screencasts. The bigger the bubble the bigger the percentage of screencasts using this resolution. Red bubbles' positions indicate popular screen resolutions

Fig. 28.3 Percentage of
screencasts using video with
given frame rate expressed
in FPS (Frames Per Second)

Another parameter of screencasts' video that was analyzed was frame rate. Playing back recorded user actions over a desktop usually does not require frame rate as high as in movies, i.e. 30 fps. Fig. 28.3 shows that analyzed screencast mostly used 15 fps. A quarter of them used 30 fps, which may be required when screencasting same high quality animation or a movie being played in the shared desktop. Almost 14% of screencast used 25 fps which is a rate adapted from analog television. Other frame rates were barely used.

Audio for a screencasts is usually a voice narration. Although human voice is proved to need lower sampling frequency than music to be clear and understandable it is a good practice to always use best frequency i.e. 44 kHz–twice the highest frequency human ear can detect. As shown on Fig. 28.4 analyzed screencasts in most cases follow this practice. Only 11,5% screencast used lower sampling frequency, i.e. 22 kHz, which was used for voice recordings in times of low efficiency audio compression algorithms.

All the screencasts were recorded using only one microphone, which suggests using mono audio. However only one third of screencasts used 1 channel of audio.
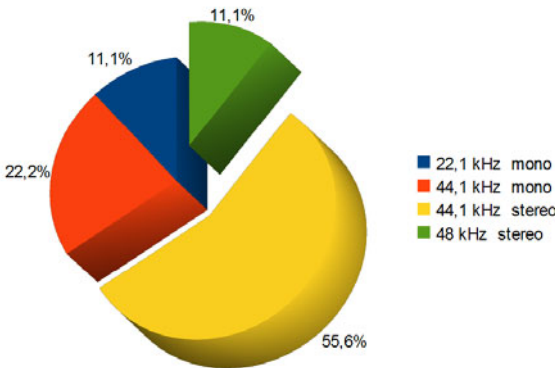

Fig. 28.4 Percentage of screencasts using audio with given sampling rate and number of channels
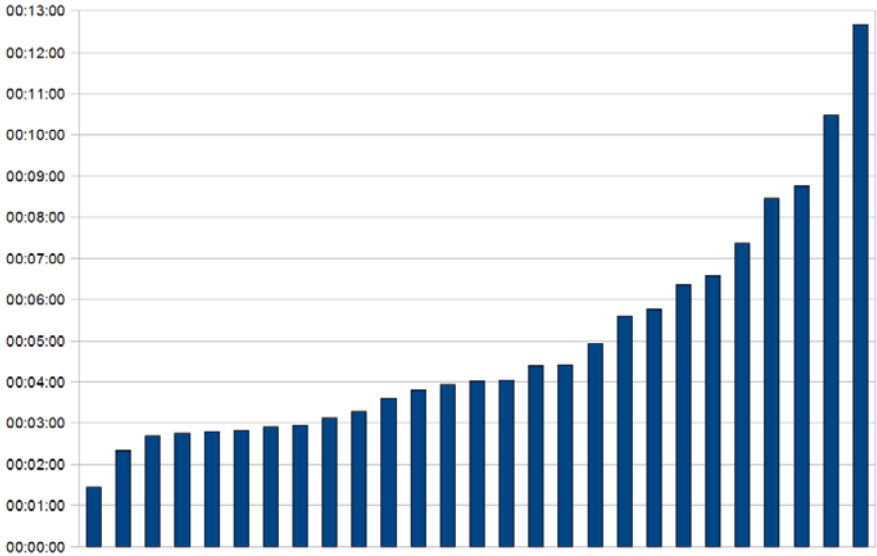
Fig. 28.5 Playback time of screencasts from the experiment

Using stereo may suggest post production mastering of audio and adding effects that require more channels, e.g. reverb. Since audio recordings for analyzed screencast very often were missing any editing the most likely cause of using stereo is not customizing default settings of an audio recording tool.

Beside codecs, the other limitation for screencasts created as a student task was time. Length of each screencast should have not exceeded 5 minutes. This time is enough for simple tutorial and not long enough to bore viewers and lose their attention. Although most of students managed to keep their screencast short enough (see Fig. 28.5) one third of the screencasts were longer than 5 minutes.
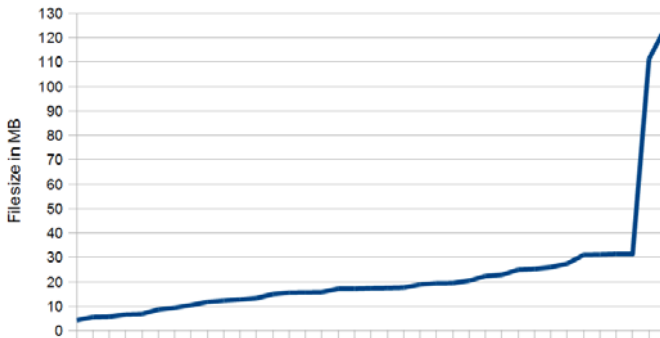

Fig. 28.6 Filesizes of screencasts from the experiment

Playback times are hardly correlated with sizes of screencasts' files. Even though the same media codecs were used, different resolutions, frame rates, audio sampling frequencies and number of channels implicated different bitrates for screencasts. This, in addition to different qualities chosen during compression, made file sizes independent of their playback times. Nevertheless the filesizes are rather regularly distributed over 5MB–30MB interval, beside a few much larger files (see Fig. 28.6). The larger filesizes are mostly results of too high frame rate setting.

28.5 Conclusions

Using screencasts allows easy publishing of user actions in computer environment. Adding to such video a frame showing the narrator via a webcam stream makes this media very powerful tool for asynchronous learning. Along with cheap distribution over Internet, the reusable aspect of screencasts makes them a low cost, high efficient educational tool. Many bloggers and podcast creators use this medium as a basis for their tutorials, demos or just showing off. The more popular screencasting gets the more use cases it fits. A combination of static screenshots or slides with live narration may also be classified as screencast, but thanks to site slideshare.net it gained a new name: slidecast.

In this paper an experiment concerning screencasts created by students was presented. The experiment's goal was to analyze the screencasts' metadata to find most popular parameters used while producing screencasts by regular users. To sum up the experiment output: the most frequently used resolution for screencasts is VGA (640x480). Most common frame rate for screencast's video is 15 fps. Audio in most cases used 2 channels (stereo) with 44 kHz sampling. The length of screencasts was imposed as the task limitation, but results show that keeping time constrain is difficult with screencasting without a prepared script. Filesizes of created screencasts in almost all cases ranged from 5–35MB.

These parameters are similar to parameters of common Internet videos presented on sites like YouTube. Since screencasts have the same parameters as videos shot with cellphones or webcams, there is a chance they may become just as popular. And this would probably push today e-learning methodologies into a new direction.

References

1. Choros, K., Kopel, M.: Telekonsultacje internetowe w systemie StOPKa na Wydziale Informatyki i Zarzadzania Politechniki Wroclawskiej. In: Oficyna Wydawnicza Politechniki Wroclawskiej, Multimedialne i sieciowe systemy informacyjne, Kliczkow, September 19-20, 2002, pp. 613–622 (2002)
2. Richardson, T., Stafford-Fraser, Q., Wood, K.R., Hopper, A.: Virtual network computing. IEEE Internet Computing 2(1), 33–38 (1998)
3. Summers, B., Summers, R.: Official Microsoft Netmeeting 2.1 Book, Microsoft Press (1998)

4. The Associated Press: Need a tutor? YouTube videos await, USATODAY.com (2008),
   `http://www.usatoday.com/news/education/`
   `2008-12-11-youtube-tutoring_N.htm`
5. Udell, J.: What Is Screencasting, O'Reilly Media (2005),
   `http://digitalmedia.oreilly.com/pub/a/oreilly/digitalmedia/`
   `2005/11/16/what-is-screencasting.html?page=1`
6. Udell, J.: Name that genre, blog.jonudell.net (2004),
   `http://jonudell.net/udell/2004-11-15-name-that-genre.html`
7. Wikipedia Contributors: HTML5 video, Wikimedia Foundation (2010),
   `http://en.wikipedia.org/w/index.php?title=HTML5_video&`
   `oldid=366027820`

Chapter 29
An Opened Agent-Oriented System for Collaborative Learning

Mazyad Hanaa and Kerkeni Insaf

Abstract. This paper introduces the specification and the modeling of an opened agent-oriented system. We are interested in the development of a learning platform that offers learners a support for their projects development and tools to support collaboration among distributed users and to manage traces of all interactions related to learners belonging to a given group. We used the Aalaadin methodology in order to model our system.

29.1 Introduction

The evolution of Information and Communication Technology (ICT) helped in the development of systems dedicated to learning. Today, the computer became a simple element of a more complex system involving the cooperation of various human and artificial entities that share a set of teaching resources.

Recently, we are interested, in our laboratory, in the field of collaborative learning. This work resulted in the definition of a teaching method called MAETIC, equipped with ICT, aiming at developing pedagogy for project-based learning [10]. The main purpose of the deployed pedagogy is to provide real support to the management of project-based learning by group of learners.

In the context of e-Learning, it is essential to avoid the failure and/or desertion of learners. In addition, in collaborative learning, both tutors and learners are seen, at the same time, as producers and consumers, gathered in an environment where everyone has something to teach and something to learn. Our objective is to develop a learning platform that offers learners a support for their projects development and monitoring and tools to support collaboration among distributed users who are part

Hanaa Mazyad and Insaf Kerkeni
Université de Lille Nord de France
Laboratoire d'Informatique Signal et Image de la Côte d'Opale, 50 rue Ferdinand Buisson, BP. 719, 62228 Calais Cedex, France
e-mail: mazyad@lisic.univ-littoral.fr,
        kerkeni@lisic.univ-littoral.fr

of the same workgroup and to manage traces of all interactions related to learners belonging to a given group and report the indicators of progression status in this learners group and its durability. Multi-agent systems (MAS) come as a solution for knowledge organization and exploitation problems and also for problems of the coordination and communication in such environments.

This paper presents a web-based peer-to peer system called COLYPAN that supports collaboration among learners, learners groups and tutors. This system allows the implementation of functionalities in order to keep a detailed history of all the actions of the groups when they access to their accounts in order to assess the group's life and its evolution.

In this paper, we start, in section 29.2, by briefly introducing the MAETIC method before discussing some related works. In section 29.3, we describe the multi-agent architecture proposed for a project management collaborative learning system. In section 29.4 we present the group working way. Finally, some directions for future work and conclusions are presented in section 29.5.

29.2 MAETIC Method

29.2.1 Description of the Method

MAETIC [5] (from french "Méthode pédAgogique InstrumentEe par les TIC": a pedagogic method instrumented by the ICT) is a teaching method which, as part of pedagogy for project-based learning, describes a set of formalized and applied procedures according to defined principles. Thus, the objective of MAETIC is to allow a learner to develop requested knowledge and skills by the implementation of a developing process of a "product" and lead to technical project management. For the tutors, MAETIC's objective is to promote the establishment of a process that will facilitate their educational activities.

The learning activity in this device is defined by the tutor. Learners must join groups to accomplish this activity. Learners in each group have the same responsibility: the commitment to finish the work, time management and the respect of deadlines. There are no predefined roles or division of tasks.

29.2.2 Contribution of Multi-agents System

Contrasting with traditional education practices, which view tutors as *producers* and learners as *consumers* of knowledge, in collaborative learning, both tutors and learners are seen, at the same time, as producers and consumers, gathered in an environment where everyone has something to teach and something to learn. This way, instead of playing the role of detaining and transmitting knowledge, the tutor assumes other functions, such as those of *motivator*, *guide* and *collaborator*. Meanwhile, the learners become more active and responsible for their own learning.

There is a critical need, in such context, for tools: (1) supporting collaboration among distributed users with similar interests, or who are part of the same

workgroup; (2) organizing information for facilitating access in various contexts; (3) managing traces of all interactions related to learners belonging to a given group and report the indicators of progression status in this learners group and its durability. Multi-agent systems (MAS) are involved in the modeling of interactions in complex societies of artificial or human individuals. They bring an interesting resolution for knowledge organization and exploitation problems and also for problems of the coordination and communication mechanisms.

29.2.3 Related Works

Before discussing the proposed architecture, we review the approaches presented in the literature that provide multi-agents platforms for collaborative learning. In the field of Artificial Intelligence and education, several approaches have been developed. For example, Guizzardi and al. [4] have developed a Peer-to-peer system called "Help & Learn". This system was modeled using an agent-oriented language called AORML [11]. It is an open system that is designed to support the extra-class interactions between learners and tutors. "Help & Learn" is limited to providing assistance to learners who request it. Other systems have been developed. Fougeres and Ospina [3] have proposed a based-agent mediation system for the project management platform called iPdagogique. This system, modeled in AUML, serves as an interface between the human and the application to enhance their relationship and is used to promote collaboration among users. Recently in [9], the authors presented a model for an adaptative multi-agent system for dynamic routing of the grant's activities from a learning environment. This model allows the assignment of activities taking into account the specialization of learners, their experience and the complexity of activities already taken. None of these three systems cares of monitoring learning and therefore, cannot trace user's activities.

Mbala and al. [8] have developed a multi-agent system called SIGFAD to support users in remote education. SIGFAD is modeled using the MASE methodology and uses the JAM model for building agents. It is interested in monitoring learning. However, it is not sufficiently independent and does not start up alerts to prevent tutors if there's a problem with a learner or group. We propose hereafter, a multi-agent approach for modeling a collaborative learning system aimed to fill this lack.

29.3 A Multi-agent Approach for Modeling a Collaborative Learning System

The agentification of MAETIC method needs the use of a methodology. Several methodologies were proposed for the development of MAS. We will follow Aalaadin methodology. Indeed, Aalaadin supports: Distributed systems where their members are geographically distributed; Opened systems that allow the integration of new members or the departure of a member; Complexity of elaborated systems; Autonomy of agents; Cooperative behavior of agents to achieve their goals.

Aalaadin is an organizational method developed by Gutknecht and Ferber [2]. It is, first, a background for developing multi-agent systems, providing methodological guidance and secondly, a prototyping and running environment for agents based on notions of group and role through the AGR (Agent/Group/Role) model [2].

29.3.1   The System Modeling

The AGR model of Aalaadin methodology is based on three primitive concepts: Agent, Group and Role that are structurally connected and cannot be defined by other primitives.

- Agent: an agent is an active and communicating entity that plays roles within groups. An agent may hold multiple roles, and may be a member of several groups. An important characteristic of the AGR model is that no constraints are placed upon the architecture of an agent or about its mental capabilities. Thus, an agent may be reactive as an ant, or clever as a human.
- Group: a group is a set of agents sharing some common characteristics. A group is used as a context for a pattern of activities, and is used for partitioning organizations. Two agents may communicate if and only if they belong to the same group, but an agent may belong to several groups. This feature will allow the definition of organizational structures.
- Role: the role is the abstract representation of a functional position of an agent in a group. An agent must play a role in a group, but an agent may play several roles. Roles are local to groups, and a role must be requested by an agent. A role may be played by several agents.

29.3.2   The Analysis Phase

This phase identifies the system functions and dependencies within the identified communities. The functionalities of the system have been organized into the following objectives:

- Determine the present / absent learners over a period, in a given group.
- Identify the active learners (who communicate and carry out activities).
- Identify the passive learners (who are connected but do not communicate and do not realize the expected activities) in a given group.
- Determine the dynamic groups (with request to the number of their active people).
- Determine if a group performs or not a given activity.
- Determine if a teacher has started / ended a Weblog: time between the opening of a Weblog and the date of sending the first message; time between the date of the last message and the date of closing the Weblog.
- Provide for each user the ratio: (effective time of connection / expected time) over one week.
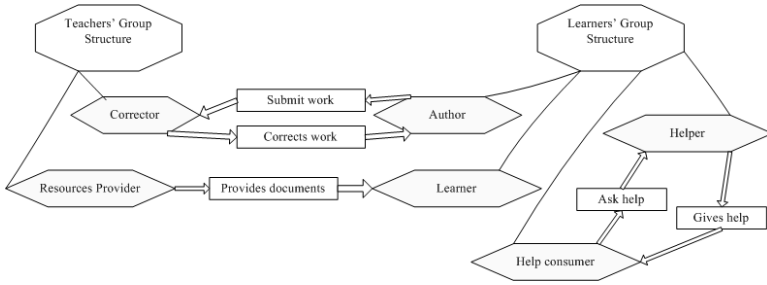
Fig. 29.1 The organizational structure of the groups

29.3.3   The Design Phase

There are two group structures in our system: the teachers group structure and the learners group structure. The teachers group structure has only two roles:Corrector of learners works and Resources provider (documents, courses,..).
The learners group contains the roles below:

- Helper: giving help to learners
- Help consumer: Asking questions to other learners
- Author: upload documents to give it to teachers
- Learner: download documents given by the teachers.

Each teacher agent called a_TEACH plays at least one role of the teachers group roles. Each teacher belongs to at least one learners group. The agent learner called a_Learn plays one role or more of the learners group roles. The Figure 29.1 presents the organizational structure of the groups.
To collect and manage traces, we need to define the following roles:

- Supervisor of activity
- Supervisior of groups
- Supervisior of tools
- Supervisior of interaction with the knowledge base of each group
- Supervisor of interaction with the database.

We attribute each one of these roles to one agent :

- a_ACTIV: supervises actors' activities during a session. It provides statistics concerning the progression of each activity. It reminds learners about deadlines and notifies the late groups by sending alerts.
- a_GROUP: supervises groups' space during a session. It has a list of present and absent agents in a group. The a_GROUP has a list of all the group members and information about each one like skills and especially, the other groups to which each peer belongs. This way, if one member p of the group a needs to communicate with another group b, then the a_Group agent can find which peer x belongs to both of groups a and b and thus the peer x can help the peer p by joining the group b

- a_TOOL: supervises the use of tools. It must provide statistics about the use of groups and tutors space tools (Email, Chat, Forum, Weblog...). With each use of the system tools, interactions are sent to the a_KB;
- a_KB: manages the knowledge base of the group. Each group has its own knowledge base (KB). The KB contains information about the group members and the project. The a_KB agent is also responsible of the interactions with the global database.
- a_DB: manages the database and users accounts.

29.3.4   The Relaization Phase

We have chosen to deploy the system on the multi-agents platform Madkit[1]. This choice is taken owing to the fact that MadKit is intended for the development and the execution of multi-agents systems and more particularly for multi-agents systems based on organizational criteria (groups and roles). However, MadKit does not impose any particular architecture to the agents.

MadKit communication is based on a peer to peer mechanism, and allows developers to quickly develop distributed applications using multiagent principles. In a collaborative learning system, each member must manage and exchange his knowledge and cooperate with others in order to achieve his goals. Compared to these aspects, a Peer-to-peer system is particularly suitable to develop a collaborative learning system since depicted with the following capabilities:

- Supports autonomy: every member of the system is seen as a peer that manages and has control over a set of local technologies, applications and services;
- Is decentralized: the community of peers is able to achieve its goal independently from any specific member or component;
- Is cooperative: in order to join and use the system, each member must provide resources or services to the others;
- Is dynamic: peers and resources can be added or removed at any time.

The multi-agent system (MAS) is also an appropriate framework for realizing a P2P application. The characteristics that they have, especially (a) their capability to allow the sharing or distribution of knowledge, and (b) that they assemble a set of agents and coordinate their actions in an environment to accomplish a common goal, are needed in this P2P application.

29.3.5   The COLYPAN System

In this section, we proposed an agent-based architecture that allowed the implementation of a collaborative learning system dedicated to project management. We called the obtained system COLYPAN (COllaborative Learning sYstem for Project mANagement). It provides users tools to accomplish their project. The users of this

---

[1] http://www.madkit.org

system are learners and tutors. In COLYPAN, the learning activity is defined by the tutor. Learners must join groups to accomplish their activities.

In each group, learners have the same responsibility: the commitment to finish the work, time management and the respect of deadlines. There are no predefined roles or division of tasks.

COLYPAN is also a collaborative learning system where users exchange their information and skills and thus learn from each others. The knowledge resources exchanged in the COLYPAN environment isn't differentiated from those exchanged for other purposes:

- There is a share of physical resources, such as: books, articles, and other educational artifacts;
- With the growing use of information technology and the Internet in these settings, there are plenty of electronic documents, references, and web links;
- There is also tacit knowledge, i.e. knowledge found in people's minds and that is usually informally exchanged among them by different means, for instance for persons through messages, or via Internet communication tools integrated in virtual learning environments.

Another objective of COLYPAN is to provide tutors with tools to enable them to determine the activity level in groups and startup needed mechanisms for re-mediation and assistance [6]. These objectives can be divided into two categories: those related to user support and objectives dedicated to save and extract interaction data. The aim of the second category of objectives is to review automatically the interaction data in order to alert users when the progress of the work session is not satisfactory or that the group risks a burst.

29.4 The Group Working Way

In COLYPAN, groups are independent of each other. Each group works on its own project and has its own resources and its own artificial agents. Each group has an agent for each learner "a_LEARN"; an agent supervisor of activities "a_ACTIV"; an agent supervisor of space group "a_GROUP"; an agent supervisor of tools "a_TOOL"; an agent supervisor of knowledge base "a_KB". Only two agents are common to all the groups: the teachers agent "a_TEACH" and the database super-visor agent "a_DB".

Each peer knows only the peers that belong to his/her group and therefore he/she can communicate with them. Thus, groups can communicate or share information only if they have at least one member in common (Figure 29.2). In fact, there is an overall view of all the groups in the system. Each group knows the number of existing groups, their knowledge and skills. However, there is no direct link between the groups.

Each group has a goal and all the group members must work together in order to achieve it. Collaborative work requires that the members contribute equally for
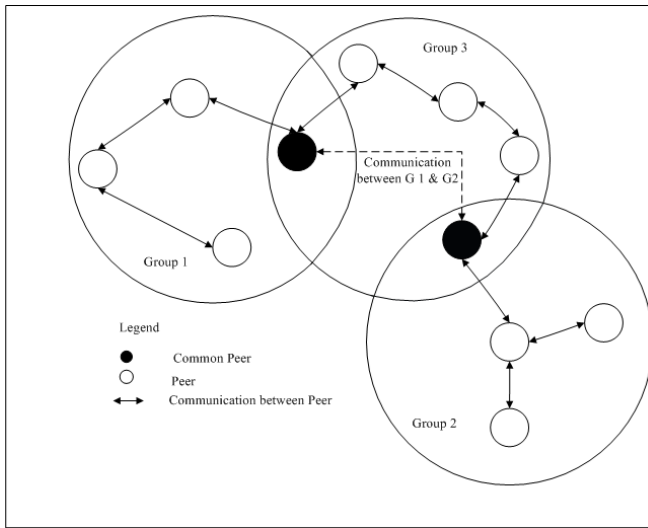
Fig. 29.2 Intra-group and inter-group communication

realizing the work. Thus, all of the members must be active for achieving the group's goal. The inactive members affect the group and delay the objectives achievement. So to ensure the activity of all groups, each inactive group (more than half of its members are inactive) receives first a warning notification sent automatically to all of its members. If after a period determined by the tutor, the group remains inactive, the group must stop its inactive members and send a join request to other members or break up the group and in this case, the members may request to join other groups. At the end of each project, the group may, if its members want, start working on another project or the group will be dissolved.

Initially, a person who wants to form a group sends a join request to all the members. This person becomes the group initiator. Each peer who wants to join a group must submit an admission request to this group. Noting that, a peer can belong to several groups simultaneously. The group initiator has the right to accept the first member of the group. Once the group is composed of more than one member, a vote to accept/reject a new member becomes mandatory. For this, an email is automatically sent to all the group members and voting will be open after sending the email. To vote, each member must access his account and accept/reject the candidature. At the vote deadline, the system automatically records the votes. To consider the vote, at least the half of the group members had to vote. Depending on the vote result, the candidate is admitted or not. In the case of equality of voice, the group initiator decides to accept/reject the candidate. Once the groups are formed, tutors suggest different projects. Each group must choose a project. If several groups choose the same project, they must negotiate to reach a mutual agreement.

The project realization will be step by step according to the project management method MAETIC. Tutors specify the deadlines for each stage and the final deadline.

Note that the tutor's role is to guide and to advise his learners, in addition to his principal role in terms of monitoring and evaluation. Each group accomplishes its project alone. On the other side, it could ask help from other groups. In fact, the "a_group" agent has a list of all its group members and the other groups to which each peer belongs. This group must provide an estimated planning and the final one. By comparing them, inactive members could be determined.

29.5   Conclusion and Future Work

In this paper, we proposed an agent-based architecture for an opened collaborative learning system called COLYPAN. This system has been analyzed, specified and designed according to a methodology for multi-agent system developing called Aalaadin. In the field of collaborative learning, management of traces is important since it allows analyzing the collected information concerning the learner or group of learners and provides the teacher or trainer with accurate and adequate information for his needs on the individual and collective evolution of learning.

Our system is an opened system that supports the interaction between learners, groups of learners and tutors. It provides users with tools for the project management and allows tutor:

- To know the level of productivity of each learner in terms of realization of pedagogic activities, and its level of communication with the other members of the group;
- To know the progression of each group compared to the realization of the activities and also, the present, absent, inactive members;
- To know the levels of realization of the activities by all groups and thus to adjust the calendar if needed.

One of the main assumptions of multi-agent systems is that their underlying network infrastructure is steady, that is the nodes of the network on which agents are deployed are known, and that there is no break in the execution of agents' protocols. Oppositely, P2P networks define a framework particularly characterized by the fact that the nodes that are dedicated to agents deployment may appear or disappear dynamically in the network or even become unavailable, due to many reasons such as the mobility of the user in its environment or his disconnection. These changes in the system configuration create unusual circumstances that make the current mechanisms of agents reasoning inappropriate. For example, how can we manage the protocols interruptions, the disappearance of a set of agents (and thus of competences), the migration of data according to the context [1],[7]. Thus, our purpose is to develop a scalable negotiation-oriented coalition formation method specifically tailored for large-scale distributed systems where nodes may crash and every agent has a partial view of the system and can only communicate with the agents in its own view, i.e., its neighbors.

References

1. Aknine, S., Mir, U., Arantes, L.B.: Multi-agent Coordination in ad'hoc networks based on coalition formation. ICAART (1), 241–246 (2010)
2. Ferber, J., Gutknecht, O.: Aalaadin a meta-model for the analysis and design of organizations in multiagent systems. In: IEEE 3rd International Conference on Multi-Agent Systems, Paris, pp. 128–135 (1998)
3. Fougères, A.J., Canalda, P.: iPédagogique: un environnement intégrant la gestion assistée de projets d'étudiants. Colloque TICE, Lyon (2002)
4. Guizzardi-Silva, R., Aroyo, L.M., Wagner, G.: Help&Learn: A peer-to-peer architecture to support knowledge management in collaborative learning communities. Revista Brasileira de Informatica na Educaçao 12(1), 29–36 (2004)
5. Marchat, H.: Kit de conduite de projet. Organization editions, Paris (2001)
6. Mazyad, H., Kerkeni, I.: A Multi-Agents Approach For Modeling a Collaborative Learning System. To appear in WBE 2010 proceeding, Sharm el-Sheikh (2010)
7. Mazyad, H., Kerkeni, I., Ajroud, H.: Multi-Agent System Architecture for Managing the Coordination in Peer-to-peer Systems. In: ICTTA 2008, DAMAS (2008)
8. Mbala, A., Reffay, C., Chanier, T.: SIGFAD: un système multi-agents pour soutenir les utilisateurs en formation à distance (2003)
9. Simian, D., Simian, C., Moisil, I., Pah, I.: Computer Mediated Communication and Collaboration in a Virtual Learning Environment Based on a Multi-agent System With Wasp-Like Behavior in Large-Scale Scientific Computing book, pp. 618–625 (2008)
10. Talon, B., Leclet, D.: Dispositif pédagogique pour un apprentissage de savoir-faire. Revue Internationale des technologies en pédagogie universitaire 5(2), 58–74 (2008)
11. Wagner, G.: The Agent-Object-Relationship Meta-Model: Towards a Unified View of State and Behavior. Information Systems 28(5), 475–504 (2003)

Author Index