Lecture Notes in Computer Science 5188

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Monica Sebillo Giuliana Vitiello
Gerald Schaefer (Eds.)

Visual Information Systems

Web-Based Visual Information Search and Management

10th International Conference, VISUAL 2008
Salerno, Italy, September 11-12, 2008
Proceedings

Springer

Volume Editors

Monica Sebillo
Giuliana Vitiello
Università di Salerno
Dipartimento di Matematica e Informatica
Via Ponte don Melillo, 84084 Fisciano (SA), Italy
E-mail: {msebillo,gvitiello}@unisa.it

Gerald Schaefer
Aston University
School of Engineering and Applied Science
Aston Triangle, Birmingham B4 7ET, UK
E-mail: G.Schaefer@aston.ac.uk

Library of Congress Control Number: 2008934331

CR Subject Classification (1998): I.4, I.5, I.2.10, I.2.6-8, I.3, H.2.8, H.3, H.5

LNCS Sublibrary: SL 6 – Image Processing, Computer Vision, Pattern Recognition, and Graphics

ISSN 0302-9743
ISBN-10 3-540-85890-3 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-85890-4 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springer.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12522781 06/3180 5 4 3 2 1 0

Preface

The international conference series on Visual Information Systems (VISUAL) originates from the idea that the exchange of views, results and experiences would allow researchers and practitioners working in the area of visual information systems to address common problems more effectively. Following the success of previous editions held in Melbourne (1996), San Diego (1997), Amsterdam (1999), Lyon (2000), Taiwan (2002), Miami (2003), San Francisco (2004), Amsterdam (2005), and Shanghai (2007), the aim of the 2008 edition of the International Conference on Visual Information Systems was to provide a fruitful forum of discussion in research fields related to computer vision, databases, human computer interaction, image processing, information visualization, and knowledge and information management. The University of Salerno was honored to celebrate the 10th edition of VISUAL, choosing a prestigious hotel situated on the enchanting Gulf of Salerno, located between the beautiful Coast of Amalfi and the Cilento National Park (both part of UNESCO patrimony). VISUAL 2008 received 58 submissions, which were reviewed by more than 38 international experts in the related research fields from 19 different countries. The resulting technical program continued the tradition of mixing carefully-reviewed research contributions, experience reports and other activities with plenary presentations. Papers collected in this volume include, but are not limited to, topics on data visualization, geovisualization, image and video databases, mobile visual information systems, visual data mining, machine learning and pattern recognition techniques for visual information management, and applications of visual information systems. In particular, three keynote talks were delivered at VISUAL 2008 and six thematic sessions were organized. The first invited talk, A Visual Analytics Approach to Exploration of Large Amounts of Movement Data, was given by Prof. Gennady Andrienko in cooperation with Natalia Andrienko, from the Fraunhofer Institute Intelligent Analysis and Information Systems. Prof. Andrienko faced the issue of enhancing human interpretation of movement data and described the results of their research experience about adequate analysis methods. The second talk was given by Prof. Franz Leberl from Graz University of Technology. The aim of his presentation, Locational Awareness on the Internet, was “to evangelize the current capabilities of the Virtual Earth system, point to some pieces of new science in the analysis of imagery of the human habitat, and of Visual Computing, and set the stage for an educated speculation about the future of computing.” Finally, Prof. Shi Kuo Chang, from the University of Pittsburgh, concluded the Keynote Sessions by giving a talk on Virtual Spaces: From the Past to the Future, where he first described how a computer scientist with research interests in visual languages may benefit from the theory and practice of architecture, and then introduced the concept of the sentient map, aimed to make an interactive map even more powerful. During the thematic sessions, the authors delivered their contributions and the participants were incited to carry out constructive discussions on research results of great current interest. The first five sessions were concerned with Information and Data Visualization, Advanced Techniques for Visual Information Management, Mobile Visual Information Systems, Image and Video Indexing and Retrieval, and Applications of Visual Information Systems, respectively. The last session was then devoted to some interesting industrial experiences. Finally, let us express our gratitude to the many people who supported the realization of this event with their time and energy. First of all, we would like to thank all the authors for submitting relevant contributions for the publication of this volume. We also gratefully acknowledge the work of the Program Committee Members, as well as that of the external referees, who offered their expertise in the review process. Thanks are due to the University of Salerno, the UniCredit Banca di Roma and the Dipartimento di Matematica e Informatica for their contribution to the success of the conference, as well as to the Unlimited Software, the Engineering and the Pixsta companies for their support. Finally, we would like to mention the work done by Rita Francese, in interacting with our industrial partners; Luca Paolino, in corresponding with the authors; Vincenzo Del Fatto in supporting the scientific organization, and Davide De Chiara, in designing and maintaining the conference website.

September 2008

Guoping Qiu
Genoveffa Tortora
Monica Sebillo
Giuliana Vitiello
Gerald Schaefer

Organization

The 10th edition of the International Conference on Visual Information Systems - VISUAL 2008 was organized by the Dipartimento di Matematica e Informatica, University of Salerno, Italy.

General Chairs

Genoveffa Tortora	Università di Salerno, Italy
Guoping Qiu	University of Nottingham, UK

Program Chairs

Monica Sebillo	University of Salerno, Italy
Giuliana Vitiello	University of Salerno, Italy
Gerald Schaefer	Aston University, UK

Organizing Committee

Rita Francese	University of Salerno, Italy
Luca Paolino	University of Salerno, Italy
Vincenzo Del Fatto	University of Salerno, Italy
Davide De Chiara	University of Salerno, Italy

Program Committee

Fernando Pereira	Instituto Superior Tecnico - Instituto de Telecomunicaciones, Lisboa, Portugal
James Wang	The Pennsylvania State University, USA
Raimondo Schettini	University of Milano-Bicocca, Italy
Alberto Del Bimbo	University of Florence, Italy
Michael Lew	University of Leiden, The Netherlands
Stephane Marchand-Maillet	University of Geneva, Switzerland
Dalibor Mitrovic	Vienna University of Technology, Austria
Maurizio Tucci	University of Salerno, Italy
Robert Laurini	INSA, Lyon, France
Clement Leung	Victoria University, Melbourne, Australia
Stefano Levialdi	Sapienza, Università di Roma, Italy
Jorma Laaksonen	Helsinki University of Technology, Finland
Athula Ginige	University of Western Sydney, Australia

Erland Jungert	Linköping University, Sweden
Markus Koskela	Helsinki University of Technology, Finland
Jean-Luc Dugelay	Eurecom, France
Arjen P. de Vries	CWI, The Netherlands
Xiangyang Xue	Fudan University, China
Tony Pridmore	University of Nottingham, UK
Bertrand Le Saux	Onera, France
Stefano Ricciardi	University of Salerno, Italy
Michela Bertolotto	University College Dublin, Ireland
Matthias Zeppelzauer	University of Technology, Austria
Heidrun Schumann	University of Rostock, Germany
Martin Dzbor	The Open University, UK
William I. Grosky	University of Michigan-Dearborn, USA
Lars Bodum	Aalborg University, Denmark
Denis Laurendeau	Université Laval, Canada
Jose Luis Fuertes Castro	Universidad Politcnica de Madrid, Spain
Amit P. Sheth	University of Georgia, USA
Isabel F. Cruz	The University of Illinois at Chicago, USA
Marinos Kavouras	National Technical University of Athens, Greece
Borko Furht	Florida Atlantic University, USA
Silvia Gordillo	Universidad Nacional de La Plata, Argentina
Gilberto Camara	National Institute for Space Research (INPE), Brazil
Sergei Levashkin	National Polytechnic Institute, Mexico
Paolo Buono	University of Bari, Italy

Referees

Bertolotto, Michela
Bodum, Lars
Buono, Paolo
Castro, Jose Luis Fuertes
Cruz, Isabel
De Chiara, Davide
de Vries, Arjen
Del Bimbo, Alberto
Del Fatto, Vincenzo
Dzbor, Martin
Furht, Borko
Ginige, Athula
Gordillo, Silvia
Grosky, William
Jungert, Erland
Kavouras, Marinos
Koskela, Markus
Laaksonen, Jorma
Laurendeau, Denis
Laurini, Robert
Le Saux, Bertrand
Leung, Clement
Levashkin, Sergei
Levialdi, Stefano
Lew, Michael
Mitrovic, Dalibor
Pereira, Fernando
Polese, Giuseppe
Pridmore, Tony
Ricciardi, Stefano
Schaefer, Gerald
Schettini, Raimondo
Schumann, Heidrun
Sebillo, Monica
Tucci, Maurizio
Wang, James
Xue, Xiangyang
Zeppelzauer, Matthias

Supporting Institutions and Sponsors

University of Salerno, Italy
UniCredit Banca di Roma, Salerno, Italy
Dipartimento di Matematica e Informatica, Salerno, Italy
Unlimited Software srl, Napoli, Italy
Engineering SpA, Pozzuoli (Napoli), Italy
Pixsta ltd, London, UK

Table of Contents

Keynote Papers

Information and Data Visualization

Advanced Techniques for Visual Information Management

Mobile Visual Information Systems

Image and Video Indexing and Retrieval

Applications of Visual Information Systems

Industrial Experiences

A Visual Analytics Approach to Exploration of Large Amounts of Movement Data

Gennady Andrienko and Natalia Andrienko

Fraunhofer Institute Intelligent Analysis and Information Systems, Germany
{gennady.andrienko,natalia.andrienko}@iais.fraunhofer.de

Thanks to the recent progress in positioning and tracking technologies, data about various mobile objects or agents are currently collected in growing amounts. Analysis of such data can yield valuable knowledge about the movement behaviors of the objects and about the environment in which they move. Since movement data by themselves lack semantics (essentially, the records consist of coordinates and time stamps), their analysis requires the involvement of a human expert, who can interpret the data on the basis of his/her knowledge. This includes the general knowledge of the (geographic) space, time, and movement as well as territory- and domain-specific knowledge.

In order to enable human interpretation of movement data, it is necessary to represent the data in a proper, i.e. visual, way. However, traditional approaches to visualization and interactive exploration of movement data cannot cope with the large amounts of movement data. There is a pressing need in adequate analysis methods. A general approach to handling large datasets includes aggregation and summarization. Aggregation means putting together data items that are close and/or similar. Summarization means deriving characteristics of so formed aggregates (i.e. groups of data items) from the characteristics of their members.

To understand how movement data can be appropriately aggregated and summarized, we have introduced a formal model of collective movement of multiple entities as a function $\mu: E \times T \rightarrow S$ where E is the set of moving entities, T (time) is the continuous set of time moments and S (space) is the set of all possible positions. As a function of two independent variables, μ can be viewed in two complementary ways, which are called *traffic-oriented view* and *trajectory-oriented view*. In the traffic-oriented view (the term "traffic" is used in an abstract sense), the movement is seen as a time-ordered sequence of *traffic situations*, where a traffic situation consists of the positions and movement characteristics of all entities at some time moment. In the trajectory-oriented view, the movement is seen as a set of *trajectories* of all entities, where a trajectory describes the movement of a single entity during the time period under analysis.

For each of the two views, different methods of aggregation and summarization are required. In the traffic-oriented view, it is necessary to aggregate and summarize traffic situations. These basically consist of points in space and point-related characteristics. Therefore, the aggregation and summarization methods suitable for point data can be applied here. In particular, the points can be aggregated by spatial compartments (e.g. cells of a regular grid), by time intervals, which may be defined according to the linear or cyclical model of time, and by values of movement attributes such as direction and speed. The resulting aggregated data can be visualized

M. Sebillo, G. Vitiello, and G. Schaefer (Eds.): VISUAL 2008, LNCS 5188, pp. 1–4, 2008.

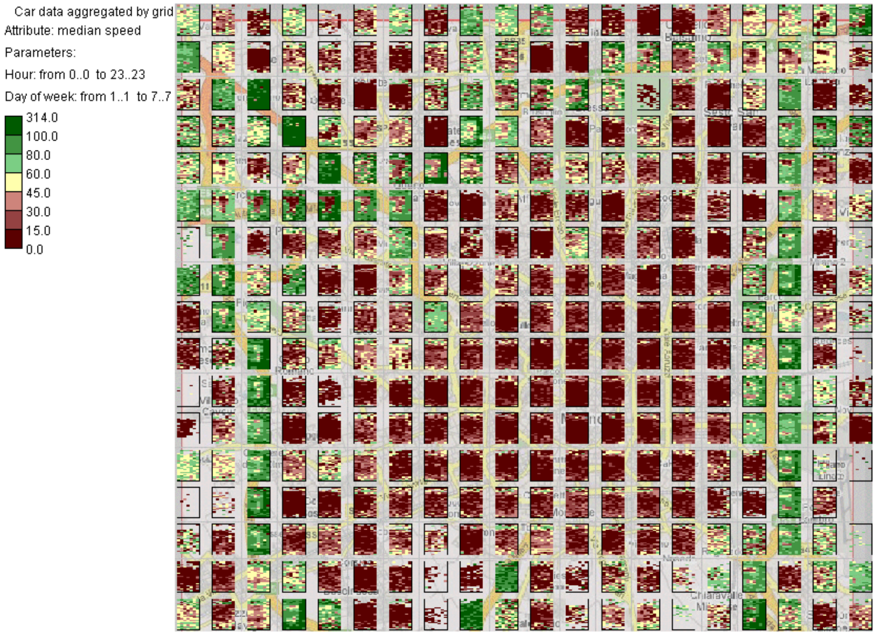


Fig. 1. A map with mosaic diagrams

by means of animated or static maps with the use of coloring or shading, graduated symbols, or diagrams, and non-cartographic displays such as temporal histograms. We particularly suggest two cartographic visualization techniques: mosaic diagrams for the exploration of cyclical patterns in traffic variation (Fig.1) and directional bar diagrams for the exploration of movements in different directions.

In the trajectory-oriented view, it is necessary to aggregate and summarize trajectories, which are much more complex objects than points. We have considered two methods of spatial aggregation of trajectories: (1) by the spatial closeness of their starts and ends; (2) by the spatial closeness and geometrical similarity of the routes. The first aggregation uses a previously defined partitioning of the space into areas. The aggregation is done by putting together the trajectories with the starts and the ends fitting in the same areas. The aggregates can be visualized by means of an origin-destination matrix and by a map with vectors (directed lines) varying in their widths and/or colors or shades according to the characteristics of the aggregates.

The second aggregation is based on clustering of the trajectories according to the similarity of the routes. For the clustering, a special similarity (distance) function is used. For the visualization of the results, we suggest a method based on treating trajectories as sequences of moves between small areas, which are defined automatically using characteristic points of the trajectories, i.e. starts, ends, turns, and stops. The areas are built as circles around clusters of characteristic points from multiple trajectories and around isolated points. The aggregation is done by putting together moves between the same areas. To visualize a cluster of trajectories, only the moves from the trajectories of this cluster are aggregated. The aggregated moves are

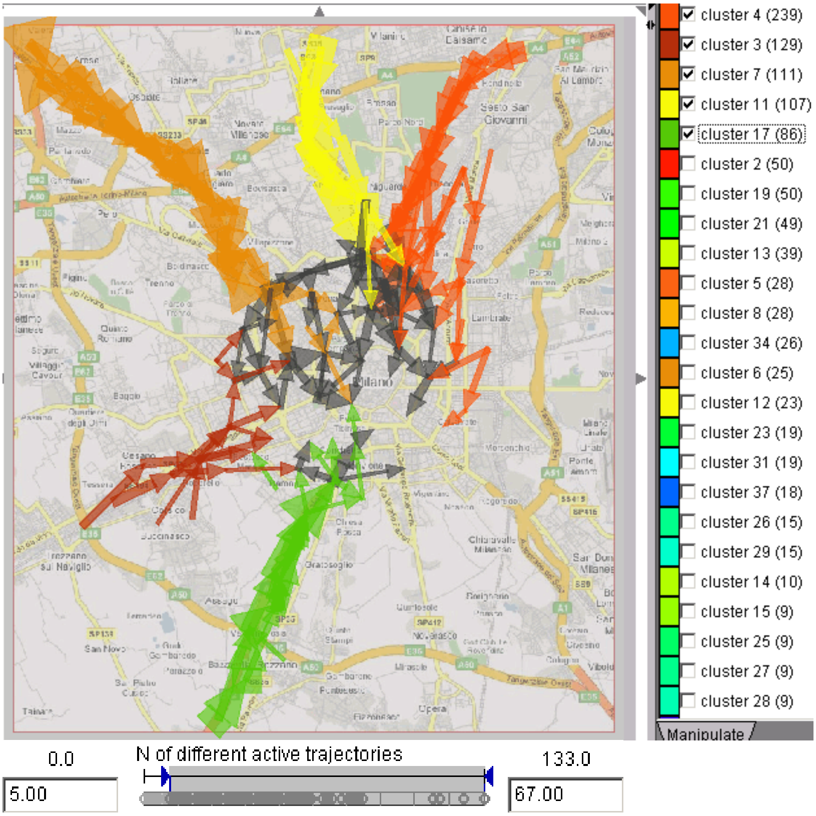


Fig. 2. Visualization of trajectories aggregated according to route similarity by means of clustering. Only the moves occurring in at least 5 trajectories are visible. Five biggest clusters, shown in different colors, are selected for viewing. Dark grey is used for the moves occurring in two or more of the selected clusters.

shown on a map by vectors, as in the case of aggregation by starts and ends. The visualization can be interactively manipulated. Thus, the user may choose to see only the moves occurring in at least k trajectories, where the parameter k can be dynamically changed (Fig.2).

When a dataset is too big for processing in the main computer memory, all the operations on data aggregation and summarization need to be performed in the database, and only aggregated data loaded in the main memory for visualization and interactive exploration. All the suggested aggregation methods except for the route-based aggregation of trajectories can be implemented with the use of standard database operations. For the route-based aggregation, the clustering procedure needs to be implemented so as to work in the database.

However, when the data are not yet too large for in-memory processing but just too large for non-aggregated visualization (due to display clutter), it is possible to do the aggregation in the main memory. This may be done by means of *dynamic aggregators*, which create very interesting possibilities for interactive analysis. A dynamic aggregator keeps references to its members (i.e. the objects it aggregates) and reacts to various interactive operations on the set of the objects such as filtering and classification. In response, it adjusts the values of aggregate attributes and, as a consequence, alters its appearance in visualizations. In particular, the aggregate moves and the generalized places (areas) can be implemented as dynamic aggregators keeping references to trajectories or fragments of trajectories. They react to the temporal filter (selection of a time interval), attribute filter (selection of trajectories by attributes such as duration and length), cluster filter (selection of clusters), and assignment of colors to groups of trajectories resulting from clustering or classification. Being represented on a map as vector symbols, aggregate moves can change their thickness or color and hide from the view when all their members are filtered out. Aggregate moves and places also control the appearance of symbols or diagrams representing values of aggregate attributes on a map or in a matrix display.

References

1. Andrienko, G., Andrienko, N., Wrobel, S.: Visual Analytics Tools for Analysis of Movement Data. ACM SIGKDD Explorations 9(2), 38–46 (2007)
2. Giannotti, F., Pedreschi, D. (eds.): Mobility, Data Mining and Privacy - Geographic Knowledge Discovery. Springer, Berlin (2007)

Locational Awareness on the Internet

Franz Leberl

Graz University of Technology, Graz, Austria
leberl@icg.tugraz.at

In March 2005, at the occasion of his 50th birthday, Bill Gates went public with his "*Virtual Earth Vision*" for local search in the Internet and stated:

"You'll be walking around in downtown London and be able to see the shops, the stores, see what the traffic is like. Walk in a shop and navigate the merchandise. Not in the flat, 2D interface that we have on the web today, but in a virtual reality walkthrough."

This implies an enormous advance in computing power, communications bandwidth, miniaturization of computing, increase of storage capacity and in the ability to model the human habitat (the Earth) in great detail in 3 dimensions, with photographic realism and at very low cost per data unit. Action followed this declaration by Bill Gates, and the transition of a then-10-year old Microsoft business segment called "Map Point" into a new *Virtual Earth Business* Unit was kicked off.

The Microsoft initiative can serve as an example and actually also as a driver for the future of computing and of *computational thinking*. Research in the complete automatic creation of 3D models of urban spaces has become greatly inspired and now is a very active field of research. The level of automation in creating 3D city models has benefited from an increase in the redundancy of the source data in the form of highly overlapping imagery either from the air or from the street.

The creation of a global 3D-model of the human habitat is a very ambitious endeavor. It implies data in the decimeter range of entire countries and cities, typically from the air. This is augmented by an even larger data set taken from the street level, perhaps in the centimeter range to be able to read the signs and understand the details of the human scale objects. Going indoors into shopping malls, museums and historic monuments will need to be at a sub-centimeter-level. The resulting data base will be absolutely unique and useful for a very large range of applications to include the Internet-of-Things, Mixed Reality, navigation, games, and of course the initial killer application – search. The result will a data set consisting of hundreds of Petabytes.

The talk will "*evangelize*" the current capabilities of the Virtual Earth system, point to some pieces of new science in the analysis of imagery of the human habitat, and of "Visual Computing", and set the stage for an educated speculation about the future of computing.

M. Sebillo, G. Vitiello, and G. Schaefer (Eds.): VISUAL 2008, LNCS 5188, p. 5, 2008.

Virtual Spaces: From the Past to the Future

Shi Kuo Chang

University of Pittsburgh, USA
chang@cs.pitt.edu

Space can be seen in many different ways. When an architect and a computer scientist look at Space they see very different things and yet sometimes they make surprisingly similar discoveries. As a computer scientist with strong research interests in visual languages I learned many things from the theory and practice of architecture. This talk on virtual spaces begins by discussing the origins of architectural pleasure and how the space of a dwelling can be divided into refuge and prospect according to Grant Hildebrand. This decomposition of space leads us naturally to consider spatial relations and patterns. On the pragmatic side we illustrate patterns by the works of Frank Lloyd Wright. On the theoretical side we consider Christopher Alexander's theory of patterns and its relationship to the theory of visual languages. After a discussion of William Mitchell's e-topia as an example of the V-topia, the virtual cities of the past, the present and the future are surveyed. An important ingredient of the virtual city is the interactive map. We can make an interactive map even more powerful by introducing the concept of the sentient map.

M. Sebillo, G. Vitiello, and G. Schaefer (Eds.): VISUAL 2008, LNCS 5188, p. 6, 2008.

Better Perception of 3D-Spatial Relations by Viewport Variations

Markus Jobst and Jürgen Döllner

University Potsdam - Hasso Plattner Institute, Research Group 3D Geoinformation
August-Bebel-Str. 88, D-14482 Potsdam
markus@jobstmedia.at, doellner@hpi.uni-potsdam.de

Abstract. Computer-graphics and especially virtual 3D worlds evolve to important tools for digital cartography, where the main aim of efficient spatial communication rules processes of conception, design and dissemination. This paper investigates the enhancement of visual spatial relations in virtual 3D worlds, which's limitation on standard displays is a main drawback of digital cartography. The main limitations of digital cartography concern the extension of viewing plane and its resolution, which have impact on information depth of the map content, transmitting an overview, thus the highlighting of spatial relations and additional request for cognitive load. These drawbacks are not only limited to 2D maps, but are also existing in virtual 3D environments, where additional geometric characteristics, like perspective distortions, multiple scales or overriding, may influence a correct extraction of spatial-related content. On the other hand these specific geometric disadvantages should be formulated as benefit of 3D, especially when infinite numbers of scale can be combined in a "natural way" or spatial content becomes extracted by naïve interaction. One main limitation of digital presentations generally persists: the limitation of the presentation area on standard displays, which leads to a very restricted overview and fewer visible relations of spatial content. View-port variations that modify perspective and/or orthographic projections are one possibility to enhance rendering methods in a way that the main disadvantages of regular perspective views become decreased and the perceptibility for an overview and spatial relations expanded. These variations cover progressive and degressive central-perspectives as well as progressive and degressive "parallel-perspectives", which provide very specific characteristics in use with spatial information transmission. This contribution focuses on enhancing virtual spatial relations in 3D environments by using view-port variations, that modify perspective and orthographic views in a progressive and degressive way. Provided that standard displays deliver significant limitations for effective and expressive geo-communication with virtual 3D environments, offers by 3D cartography are discussed. An exemplary comparison of actual 3D city models allows to identify "dead values" and gives one clue for the requested modification. The description and exemplary visualization of view-port variations lead to their theoretical communication aspects, which will guide pragmatic (user) studies in future.

Keywords: 3D geovisualization, geo-mediatechnique, perception, graphical design, geo-communication.

M. Sebillo, G. Vitiello, and G. Schaefer (Eds.): VISUAL 2008, LNCS 5188, pp. 7–18, 2008.

1 Introduction

Virtual 3D worlds deal as important tool for digital cartography. Especially when the metamorphosis of paradigms becomes accepted in the field of cartography and the "new" task focuses on a successful geospatial communication, virtual 3D worlds play their part for the conception, design and dissemination of spatial information. Virtual 3D worlds seem to be attractive for a wide public community and thus form an important communication tool. In fact virtual 3D worlds support some kind of naïve geography [1] and can enhance the understanding of topographic relations with low experience or spatial knowledge [2]. Spreading web applications, like Google Earth, Virtual Earth or World Wind, support the thesis that virtual 3D worlds are naively understood. At least these kind of geographic applications support cartographic issues and expand the working field of digital cartography in a way that visual spatial relations within the display presentation are more easily perceived, if certain limitations can be identified and further on reduced. Thus this contribution focuses on how visual spatial relations in virtual 3D worlds can be enhanced and certain limitations of their visualization can be reduced.

Virtual 3D worlds as add-on for digital cartography will have to follow semiotic requirements, which allow to identify map elements clearly. According to this requirement, virtual 3D worlds may deal as 3D cartography that generally offers extensions to digital cartography, although massive restrictions exist from another point of view. These specific "geometric disadvantages", such as perspective distortions, multiple scales or overriding, should be considered as benefits of 3D. Then these specific characteristics of 3D propagate the keeping of spatial relations for navigation and tasks of spatial literacy. Although the specific characteristics are generally used by virtual 3D environments, main disadvantages are observable in actual examples of virtual 3D cities. These observations lead to viewport variations with the purpose to extend perspective views in a way that spatial relations become enhanced, landmarks show up directions, prospective routes are not hidden and geometric values are easily comparable. A simple realization with two bending ground plates demonstrates applicability in real-time 3D environments. This technical solution does not consider any impact of this newly created presentation form for virtual 3D environments. This discussion will be subject to future research. But this way of thinking shows up one possibility to overcome some restrictions of digital cartography without changing the techniques of actual transmitting interfaces.

2 Restrictions of Digital Cartography

Digital techniques have fundamentally changed the field of cartography. Not only in the area of reproduction, but especially in terms of dissemination and user participation. In combination with the Internet almost everyone is able to reproduce and disseminate map pictures. In addition large amounts of geospatial data become accessible and can be used by everyone, who knows how to deal with simple webmapping techniques. In most cases these data are not appropriate and result in very poor maps, especially when some processing of data harmonisation and specific coding are left out. Generally professional mapmakers know how to modify geospatial data for digital cartography and displays. These use cases follow expanded cartographic rules due to specific characteristics of transmitting media [3, 4]. In any case digital approaches in cartography lead to a more extensive dissemination of geospatial phenomena, to a global management of various data (e.g. Public data of UN or NASA) and to user participation in recording geospatial data in form of GPS tracks (e.g. www.openstreetmap.org) or leaving spatial tags.

In spite of all the advantages in dissemination, data management and user participation, digital cartography comes along with restrictions due to specifics of the transmitting media "display", which is mostly used. The restrictions of the content concern overview and resolution as source elements that provoke higher cognitive load on one hand and loss of spatial relations on the other. In comparison to analogue media, like paper, digital media are restricted in their extension of the view plane. Whereas paper maps have been printed on sheets up to A0 (~84 x 120 cm), actual displays for digital cartography reach an extension of about 30 x 40 cm. Extraordinary large displays, as these are used for HDTV in home cinemas, will bring a higher extension of the viewing plane, but do not increase resolution in a way it would be useful for digital cartography. In the same way projectors (beamers) result in a large viewing plane. Their resolution is fixed to standard display resolutions (VGA, SVGA, XGA). Thus the size of the projected picture element increases and becomes clearly visible if the viewing distance is not changed. Therefore the information content cannot be increased compared to the standard display. Instead the viewing distance has to be increased in order to perceive a homogeneous presentation. As result we can conclude that the extension and resolution is closely related and is generally not significantly more than 2000 x 1500 pixel (picture elements) on an area of about 30 x 40 cm. The resolution of 2000 x 1500 pixel cannot carry more geospatial information than a printed map of about 17 x 12,5 cm size, when 300 dpi are used for the printing process (!) [5]. Lechthaler speaks of a ratio 1:3 for a dual system that uses print and display media.

The restrictions of information content are not that strong, that digital media are useless for maps. The reason is that appropriate interaction with media can overcome information restrictions with operations like zooming, panning and mouse over interaction. Zooming allows receiving more detail when zooming in or get some overview when zooming out. Both operations come along with changes of information content in order to keep perceptibility and semantics of the presentation. Panning moves the map picture at the same scale with the aim to change the map focus or to explore spatial relations around a map focus, whereas the map focus names an interesting area of a map. Mouse over interaction allows embedding visual hidden information and thus compresses information content. Although all these simple operations help to overcome restrictions of digital media, the cognitive result at the user side is not promising. The reason is that every operation calls for cognitive load and specific attention, which deflects the user from information extraction. Receiving some overview and extracting spatial relations is a challenge due to zooming, panning and mouse over events. We can conclude that digital cartography still lives with the main drawbacks in terms of restricted content (resolution and extension), increasing cognitive load to overcome content restrictions and according to this a cognitive challenge to extract a promising overview and spatial relations.

3 What 3D Cartography Can Offer?

Virtual 3D environments with cartographic pretensions are not an overall solution for digital cartography, but provide some aspects that help for the specific tasks of overview and spatial relations. From a geometric point of view 3D environments come along with geometric characteristics, such as perspective distortions, multiple scales or overriding, that influence a correct extraction of spatial-related content. These geometric disadvantages can be considered as benefits of 3D, especially when infinite numbers of scales can be combined in a "natural" way or spatial content becomes accessible by naïve interaction in the virtual world.

Combinations of multiple scales due to perspective view complicates a single identification of elements throughout the depiction, but reconstruct a natural view for humans, which helps to accept the virtual 3D environment as virtual reality. According to rules of depth perception [6], elements will be identified by cognitive processing as long as similar geometric elements exist and the perspective view is homogeneous. The main advantage of scale combination is the direct and intuitive comparison of large and small scales, which would be possible in neither a large nor small scale alone. Especially when details of a large scale are directly related with far objects, a visual solution seems to be impossible. For example architectural details in a wall that are related with specific buildings on the other side of a city will be visible in nature, but not in single scale maps. The creations of virtual 3D environments with their geometric characteristics help to rebuild reality and these relations again. Especially historic analysis, when buildings do not exist any more, call for reconstructions like these in order to identify spatial architectural relations. If these relations should be discovered in large scale maps, their extension would be to big for an exact angle extraction and a cognitive relation reconstruction (mental map). If the maps are in a small scale, so that the extension is not that large, their details will have to be generalized and are not perceivable any more. With examples like these we can conclude that virtual 3D environments enhance specific spatial relations, which affect various scales in the depiction.

In addition to the spatial relation enhancement, the camera grades of freedom (translation, rotation,..) support the naïve extraction of spatial information. As long as the camera movement and rotation follows "natural" movement metaphors, users are able to identify and interact with the virtual world in a naïve way. Natural movement metaphors rebuild functionality of movement in reality, like walking or moving the head. An interface that allows uses the same actions (walking, head movement) will have the highest impact. With some experience, humans are able to automatically translate mouse interaction to their natural spatial movement, although this translation requires enormous cognitive load, as long as mouse movement follows the natural movement metaphors.

The enhancement of spatial relations and natural movement metaphors are fundamental for a naïve geography as it is incorporated in virtual 3D environments. Following the definition of naïve Physics by Hardt (1992) and the extension by Egenhofer and Mark (1995) the notion naïve geography names "the body of knowledge that people have about the surrounding geographic world" [7, p. 4]. By this meaning naïve describes an instinctive and spontaneous approach by the user, who will search for operational and cognitive structures that match the individual structures when exploring space. Virtual space then forms one part of virtual geographic space that is represented by various geometries and scales beyond the virtual camera. Therefore virtual geographic space cannot be observed from one single viewpoint [8] and relies on natural movement metaphors for exploration within the virtual 3D environment. The cartographic representation, which is an "enhanced" virtual reality in terms of perceptibility of graphical values and the semantic consistency of map elements at various scales, delivers a more real experience than movement in reality. According to Egenhofer and Mark (1995) map-based, map-like or enhanced views of geographic space give a better representation and support a naïve assumption of where one moved through geographic space. This conclusion bases on the statements of users like "..when I get home, I want to look at the route on a map, to see where I went..." [7, p. 8]. Nowadays these people often use Google Earth as easy accessible mapping tool for the visualization and global exchange of their tracks.

Considerations above lead to the conclusion that 3D cartography on one hand enhances users identification with their access to naïve geography and the user's recording of visual spatial relations and on the other hand enable naïve interaction with natural scale combinations. Natural movement metaphors with their direct access to knowledge structures of the user support this conclusion. Naïve geography in our case of virtual 3D environments basically focuses on a movement in virtual space and explorable virtual spatial structures and does not concern GIS functionality or analysis, which may mislead user's knowledge supplement depending on existing experiences and world knowledge [9].

4 Importance of Virtual Spatial Relations

The argument that natural movement metaphors and naïve interaction help to directly access individual knowledge structures at the user side bases on the assumption that reference points, -lines and -areas provide virtual spatial relations, which are at least partly known and help to navigate, explore and extract knowledge in virtual 3D environments. These reference elements serve as virtual landmarks that are key factors for orientation within the virtual world.

Virtual worlds are often misleading and very restricted for the communication of spatial topologies if we look on examples of the gaming industry. A lot of ego-shooter games make use of 2D maps in order to give some rough overview and directions for the movement, although theses little raster maps are very restricted in resolution and extension. Darken and Cevik (1999) describe orientation issues of maps that are used in ego-shooter games. The results showed that for the purpose of navigation in 3D environments the forward-up treatment of the 2D map was the most promising. What Darken and Cevik did not enlighten at this point – it was obviously not the main focus of the paper – was the characteristic of massive overriding within these ego-shooter games. One can hardly see the second row of buildings. This fact restricts the use of landmarks for orientation and of course calls for 2D maps as orientation and wayfinding help.

Landmarks or visual references are used in real and electronic spaces. These elements are mainly used to form the cognitive map of the realistic or electronic/virtual environment and are characterized by their extraordinary features that help to recognize and memorize these features in the environment [10]. Sorrows characterizes landmarks by singularity, prominence, meaning and category. Singularity of the object describes the sharp contrast with its surroundings as regards to visual attributes. Also for Lynch (1960) singularity is the most important characteristic. A landmark is visual flashy when this building/element differs from its immediate surrounding concerning its size, shape, age or cleanliness. Prominence focuses mainly on the spatial location. For instance a building's shape is visible from various locations or it is located significantly at a junction. Meaning qualifies the content of a landmark in terms of cultural importance. Similar attributes of objects may lead to a general used category, which does not relate to a specific location but to a situation. For example a white church in a village center as specific landmark is assigned to other villages as category for town church and the center for the village. In addition to visual attributes, landmarks in virtual 3D environments obtain various kinds of interactivity, which can transform passive visual landmarks to active landmarks. Active landmarks may not be differentiated by visual attributes, but become active by user nearness or interaction, which allows for relieving transmitting interface with its graphical values.

Visual landmarks in a virtual 3D environment are one of three main levels for spatial knowledge. In order to construct a cognitive map, which is the first step in creating spatial knowledge that enables (virtual) spatial movement, the three levels of landmark knowledge, procedural knowledge and survey knowledge have to be acquired [11]. Landmark knowledge represents landmark characteristics for a specific location/area in an environment. The recording of visual landmarks and their topology is done by exploration. Procedural knowledge encodes as navigation actions along a route. It includes perceptual features along the way, which concern distances between locations and turns along the route. Survey knowledge enables a user/traveller to mentally draw an explored region from bird's eye view and to navigate with confidence. According to Elvins (1997) survey knowledge is generally gained from map study or prolonged exploration/movement in an environment.

The importance of visual landmarks and spatial relations in a virtual 3D environment for navigation, wayfinding and spatial knowledge expansion leads to the question, if actual standard perspectives deliver the best perception for 3D-spatial relations in virtual 3D environments. Therefore standard perspectives´ originalities should be evaluated.

5 Evaluating Standard Perspectives

The depiction of a virtual 3D environment on the transmitting interface/media is rendered according to viewport attributes. These attributes define a cameras orientation, distance and field of view (FoV) in case of perspective views, and camera orientation and orthographic height in case of parallel projections. Depending on these camera's attributes and a transmitting interface's resolution the amount of dead values changes. Dead values describe information pixels that are not appropriately used for information transfer. This means that the content of pixel cannot be related to an specific element or the overall visualization content. Due to a restricted amount of pixels for an information transfer, a most efficient use of these pixels is the aim.

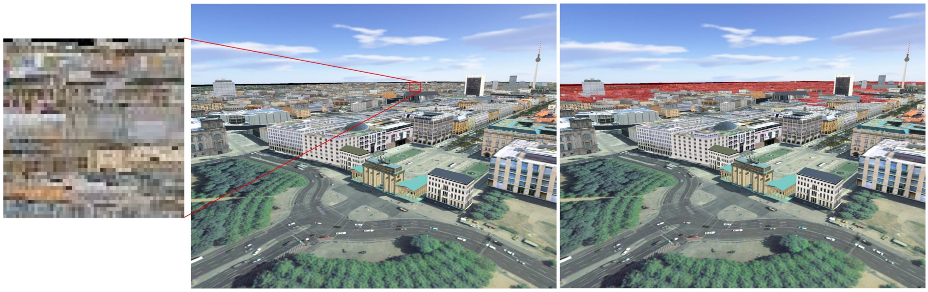


Fig. 1. The area of dead values in a central perspective view is marked red in the right picture. The picture on the left shows this “undefined” pixel area in detail.

Central perspective views combine linear perspective and multiple scales in one view, which leads on one hand to geometric distortions of the elements and on the other hand to a high variance of element sizes (on the presentation plane/viewing plane). Geometric distortions of elements assume that more than one element of the same sort exists in the view in order to retrieve geometric relations (size, primitives, ...). The high variance of scales enables the direct comparison of large and very small scales in one directed view, but also leads to dead values if elements fall below the resolution of transmitting media. If this uniqueness-relation between transmitting media and content element is larger than one, a single picture element of transmitting media has to represent several pixels of the content.

Parallel perspective views offer one single scale throughout the view, which allows a direct comparison of element sizes and orientation. Due to the isometric character of this view, the scene seems to arch upward in the background. Actually the geometric correct characters of parallel perspectives create discrepancies with the human cognitive system, mainly because of the three main rules of composition stability [6, 12, 13]. Thus the main disadvantage is its disturbance of naïve perception. The main advantage lies in a scale dependent illustration.

In case of standard perspectives we can conclude that their use of transmitting media’s viewplane for virtual 3D environments is not perfect. Due to dead value areas for various camera parameters the requirements of cartography-oriented design are mainly not met. In order to reduce these areas, viewport variations show one possible approach.

6 Extended Perspectives by Viewport Variations

A variation of viewport perspectives follows the aim to reduce dead value areas and thus improve the use of transmitting media’s viewplane. This approach is guided by syntactic considerations in terms of expressive geospatial communication. It does not concern effects of semantics and pragmatics, which also influence the composition of virtual 3D environments. Hence the following collection of perspective variations is evaluated by their impact on dead value areas, communication of spatial relations and its syntactic dimension.

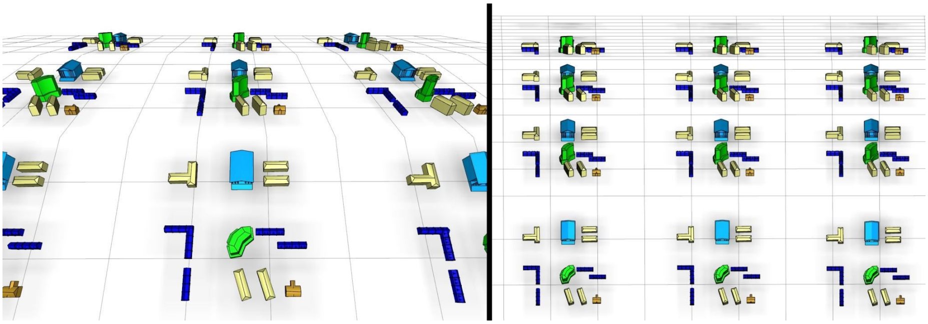


Fig. 2. Conceptual view of a progressive central perspective and progressive parallel perspective

The progressive approach intensifies perspective impression by forcing ground-view zones nearby the camera and front-view zones for far elements. This means that the perspective view becomes arranged depending on the distance to the camera. The grouping to “view angle” zones enables the specific element enhancement in that zone. For example a ground-view generally delivers footprints that are combined with very restricted front-view information. In this situation it is important to keep the footprint information by its semantics, which will lead to highlighting and aggregation for specific scales. On the other hand in an front-view zone the footprint almost disappears. The main transferred information is build up by the upright projection. Therefore the front-view and its semantic/outline has to be highlighted for specific, orientation enabling elements.

The Progressive Central Perspective makes use of “view angle” zoning that allows for element adjustment according to the view angle of viewport camera. In addition its perspective impression becomes intensified. The enhancement of the ground-view in the foreground presents an overview around the current camera position, which mainly relies on buildings footprints and topographic relations. The force for front-views in the background enhances visual landmarks by their front-view. An highlighting of important/outstanding and well known elements directly supports the usage of landmarks for orientation within the virtual 3D environment. Spatial relations between the foreground’s overview and the background’s landmarks are strengthened, especially because standard perspectives will not provide the extension (on the transmitting interface) to show both.

Progressive Parallel Perspective uses the same constellation of “view angle” zoning and results in the same argumentation for using ground-view and front-view situations. Additionally its metric constitution, which disturbs a naïve perception, gets some “perspective” character. Thus the parallel perspective can use its metric and simulates perspective characteristics.

In contrast the degressive approach with a Degressive Central Perspective generally destroys any naïve perspective perception by its use of front-view zones in the foreground and ground-view zones in the back. Its aim is to simulate user’s perspective at the camera’s position in order to generate high identification in the specific area and deliver an overview of the prospective tour at the same time. The request in this case is a high detailed model in the front and a scale-dependent map

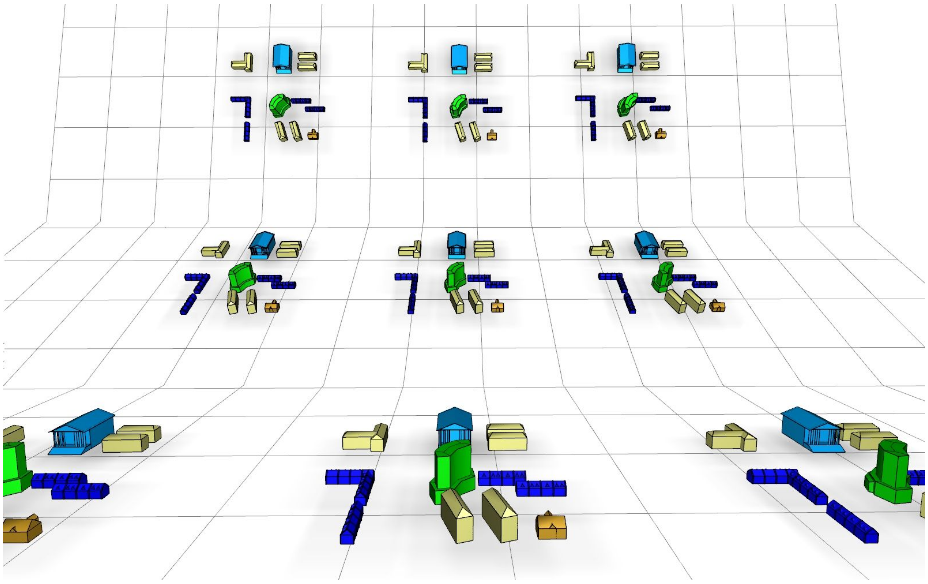


Fig. 3. Conceptual view of a degressive central perspective

design in the back. The rendering style definitely has to change in order to reduce confusion when a map is used instead of clouds and sky.

All viewport variations have in common that dead values on the transmission interface become minimized. The reason is that "view-angle" zones reduce the infinite number of scales to several groups and therefore allow for specific element generalization. The information transmitting area theoretically becomes more appropriate used according to its purpose (generating an overall overview with naïve geography or supporting navigation in virtual 3D environments). As result of this reduction of dead values and a more extensive use of transmission interface, spatial relations are enhanced as well (because more information pixels are available for overview generation).

By reason that theoretical considerations are often not realizable or result in main drawbacks of graphical processing capacities, an exemplary realization of viewport variations for real-time applications is added.

7 Simple Exemplary Realization for Real-Time Applications

A simple example of viewport variation for real-time applications was realized by a bendable ground plate that amplifies ground- and front-view. Both parts of the plate are connected with a transition zone, which shows up as concave molding. Because this concave molding with bended buildings disturbs acceptance at the user, its area should be minimized. The use of this transition zone is a homogeneous change from ground- to front-view zone.

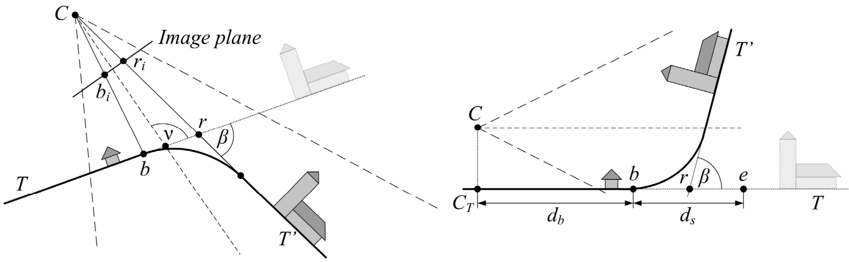


Fig. 4. Simple progressive and degressive concept of a bending ground plate, based on Lorenz et al 2008

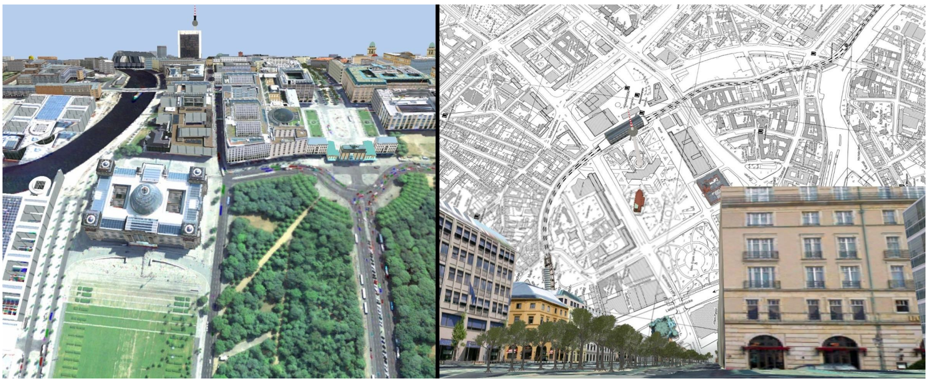


Fig. 5. Simple progressive and degressive views, rendered with LandXPlorer, based on Lorenz 2008

The deformation for the viewport variation is implemented as task for the GPU (Graphical Processing Unit). The processing scheme does not introduce new vertices. The curved transition zone can show up rendering artefacts due to insufficient tessellation. This setting allows a straightforward solution, where a LoD (Level of Detail) algorithm selects more detailed object representations within the transition zone [14].

The rendering frame rates for the real-time environment are mainly resolution independent. The access to textures can be identified as main bottleneck in the test application. An out-of-core algorithm was used to load texture on demand in sufficient resolution from the disk. Progressive perspectives come along with exceptionally high load rates, which are caused by the visible horizon. More terrain and elements become visible. Their low quality does almost not influence the loading process. In degressive perspectives only a part of the terrain and elements are visible compared to standard perspective views. The ground-view texture in the background leads only to a slight reduction in texture load [14].

This exemplary realization demonstrates applicability of preceding theoretical considerations. The implementation bases on global space deformation that is processed by graphics hardware. It permits the seamless combination of different graphical representations.

8 Conclusion

We can conclude that viewport variations enable a better perception of 3D-spatial relations. The variation of perspectives to progressive and degressive as well as central perspective and parallel perspective improves the usage of information pixels of transmission interface. Dead information values can be reduced. Due to "view-angle" zones specific enhancements of elements can be adapted more easily, which supports unambiguous perception of information content. The progressive as well as the degressive character expands the overview of the virtual 3D environment. By this means 3D-spatial relations can be accentuated in an impressive way.

Future work will have to focus on semantic and pragmatic dimension of viewport variations. Whereas degressive perspectives´ implementations for navigation systems are imaginable, their impact on users´ cognitive processing and the resulting mental model has to be proven. It may be that the resulting mental relations do not match with reality. But there also exits the chance that this presentation forms support cognitive processes and improve spatial literacy.

Acknowledgements

We thank Haik Lorenz and Matthias Trapp for their successful implementation of these theoretical syntactic considerations. Our colleagues could show that these ideas can be embedded in an existing virtual 3D city model and do not influence the dynamic character of the virtual 3D city application [14].

References

1. Golledge, R.: Learning about Urban Environments. In: Carlstein, T., Parkes, D., Thrift, N. (eds.) Timing Space and Spacing Time, Edward Arnold, London (1978)
2. Kirschenbauer, S.: Empirisch-kartographische Analyse einer echt-dreidimensionalen Darstellung am Beispiel einer topographischen Hochgebirgskarte; Dissertation; Technische Universität Dresden. Mensch&Buch, Berlin (2004)
3. Neudeck, S.: Zur Gestaltung topografischer Karten für die Bildschirmvisualisierung. Dissertation. Universität der Bundeswehr München, Heft 74 (2001)
4. Brunner, K.: Kartengestaltung für elektronische Bildanzeigen. In: Kartographische Bausteine; Band 19; Technische Universität Dresden (2001)
5. Lechthaler Mirjanka, Stadler Alexandra: Cross Media gerechte Kartengrafik in einem AIS. In: Proceedings CORP 2006 & Geomultimedia06, Manfred SCHRENK (Hrsg.); Wien ISBN 978-3-9502139-0-4
6. Kraak, M.J.: Computer-assisted cartographical three-dimensional imaging techniques. Delft University Press, Delft (1988)
7. Egenhofer, M., Mark, D.: Naïve Geography. In: Frank, A., Kuhn, W. (eds.) COSIT 1995. LNCS, vol. 988, pp. 1–16. Springer, Berlin (1995)
8. Kuipers, B., Levitt, T.: Navigation and Mapping in Large-Scale Space. AI Magazine 9(2), 25–46 (1988)
9. Feltovich, P., Hoffmann, R., Woods, D., Roesler, A.: Keeping it too simple: how reductive tendency affects cognitive engineering. IEEE Intelligent Systems, 90–94 (May/June 2004)

10. Sorrows, M.E., Hirtle, S.C.: The Nature of Landmarks for Real and Electronic Spaces, In: Freksa, C., Mark, D.M. (eds.) COSIT 1999. LNCS, vol. 1661, pp. 37–50. Springer, Heidelberg (1999)
11. Elvins, T.T.: Virtually Lost in Virtual Worlds – Wayfinding Without a Cognitive Map. Computer Graphics (1997)
12. Marr, D.: Vision – A Computational Investigation into the Human Representation and Processing of Visual Information. W.H. Freeman and Company, San Francisco (1982)
13. Rock, I.: Wahrnehmung – Vom visuellen Reiz zum Sehen und Erkennen. Spektrum Akademischer Verlag, Heidelberg (1998)
14. Lorenz, H., Trapp, M., Döllner, J., Jobst, M.: Interactive Multi-Perspective Views of Virtual 3D Landscape and City Models. In: Bernard, L., Friis-Christensen, A., Pundt, H. (eds.) The European Information Society, Taking Geoinformation Science One Step Further. Springer, Berlin (2008)
15. Darken, R.P., Cevik, H.: Map Usage in Virtual Environments: Orientation Issues. Proceedings of IEEE Virtual Reality 99, 133–140 (1999)
16. Hardt, S.: Naïve Physics. In: Shapiro, S. (ed.) Encyclopedia of Artificaial Intelligence, 2nd edn., vol. 2, pp. 1147–1149. John Wiley & Sons, Inc, New York (1992)
17. Lynch, K.: The Image of the City, p. 194. MIT Press, Cambridge (1960)

Effective Visual Scanning of Geographic Information

Olivier Swienty[1], Mathias Jahnke[1], Holger Kumke[1], and Simone Reppermund[2]

[1] Department of Cartography, Technical University of Munich, Arcisstr.21, 80333 Munich, Germany
{olivier.swienty,mathias.jahnke,holger.kumke}@bv.tum.de
[2] Max Planck Institute of Psychiatry, Kraepelinstr. 2-10, 80804 Munich, Germany
reppermund@mpipsykl.mpg.de

Abstract. Whereas a user visually processes geographic information faster and more accurate than a computational system, a system processes a large number of geographic data much faster than a user. To minimise this large capability discrepancy, we propose a cognitively-adequate design of geovisualisation as an external aid to visual information processing. Our main objective is to release the limited cognitive workload of users by adapting the design of geovisualisations to their visual scanning capabilities. For a proof of concept, we design 2D thematic maps and 3D non-photorealistic visualisations, and evaluate the efficiency of visual scanning strategies with a computational attention-model and the eye-movement recording method.

Keywords: cognitive workload, map-design.

1 Introduction

An important aspect of current research in GI science is the accumulation of geographic databases with spatiotemporal information [1, 2]. To process these amounts of geographic data, many technology-driven tools have been developed and implemented to facilitate cognitive-oriented tasks like exploring geographic information [3, 4], route planning and way finding [5]. The delicate task of designing geovisualisations that enables users an effective visual processing of these amounts of data still lags behind the ongoing geographic data acquisition. Although cognitive research of geovisualisation design increases again since the 1990s [6] the technological focus on system engineering dominates the perspective on geographic information processing. GI scientists still code vast amounts of geographic information into complex visualisations and assume that users have sufficient cognitive skills for visually processing geographic information. Such an approach seems to be questionable because it obligates users to adapt their powerful sophisticated skill of visual information processing to the limited ability of systems to design geographic information. This work aims at enhancing a system's acceptability by investigating a user's cognitive skills in a first step and adapting the design of geovisualisation to the user's need in a second step. In doing so, we follow recommendations of prominent research agendas [7, 8] that revealed a lack of cognition-based research for developing scientifically testable representations of geographic information. In particular the recommendations to investigate cognitive design approaches [9] and to study

M. Sebillo, G. Vitiello, and G. Schaefer (Eds.): VISUAL 2008, LNCS 5188, pp. 19–30, 2008.

the potential of geovisualisations to enable decision-making by tackling cognitive and usability aspects [10] are of prime importance to develop cognitive design approaches for geovisualisations. Based on principles of system engineering [11], Swienty et al. [12] consider the cognitively adequate visualisation of geographic information as a usability criterion and the relevance-based filtering of information as a utility criterion of a system's acceptability. As a result, they recently proposed the cognitive design approach of attention-guiding geovisualisation [13]. In this work, we concentrate on adapting the design of geovisualisation to the user's ability of visual scanning. 'Geovisualisation' is here understood as any kind of geographic information visualisation without focussing on exploratory data analysis.

2 Visual Scanning of Geographic Information

To handle the amount of sensory input that exceeds the processing capacity of the human brain, humans make use of their selective visual attention capability that serves as a filter mechanism by highlighting important information and suppressing irrelevant visual information. Selective visual attention is guided by both, bottom-up (sensory stimulation) and top-down factors (e.g. knowledge) [14]. The interaction of these factors controls where, how and to what users pay attention to initiate cognitive control, i.e. the organised and wilful behaviour to execute further operations.

This interaction is involved in attention shifts that are intimately connected with the activation of eye muscles to control the direction of gazes. The skill of shifting visual attention to regions and locations of interest in geovisualisations is referred to as visual scanning. Visual scanning describes current sequences of gaze shifts during visual information processing and involves shifting of attention (accomplished by gaze shifts) as well as detailed information processing (only during gaze fixations). These sequences of gaze shifts and fixations, that occur whenever users visually process geographic information, form the scan path. Figure 1 exemplifies the scan

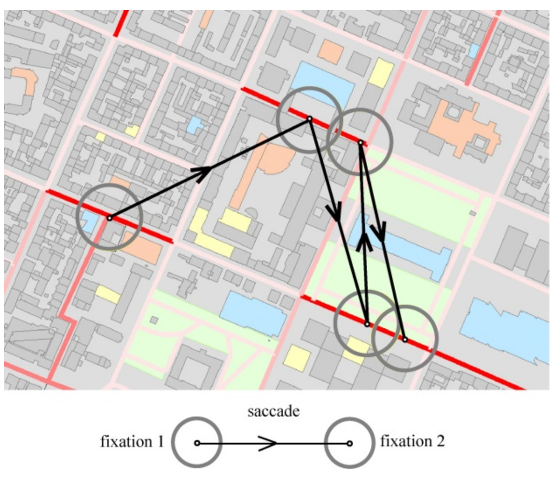


Fig. 1. Sequences of fixations and saccades form the visual scan paths

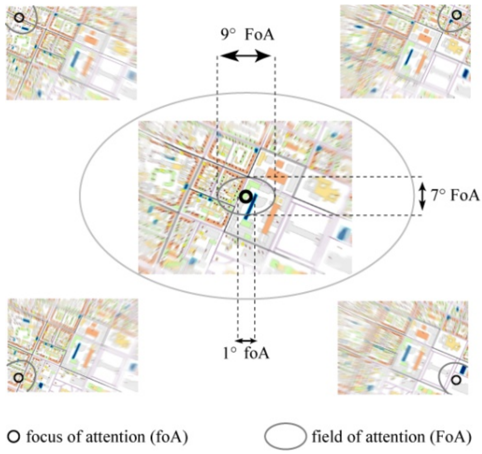


Fig. 2. The focus and the field of visual attention within the visual field when visually scanning geovisualisation

paths that were predicted with an attention model [15]. To visually scan geographic information, users employ distinct types of eye movements to shift their focus of attention from one area or location of interest to another. Saccadic eye movements are probably the most suitable technique to rapidly shift the focus of attention to relevant geographic information. While 'voluntary' saccades (to a location without a visual transient) and 'memory-guided' saccades (to a cued location after delay) are primarily top-down driven, the focus of this work is on guiding 'reflexive' saccades that are triggered by salient stimuli. In contrast to a computerized attention-model that calculates the location of fixations, users' visual scanning capability is more flexible and highly efficient due to the function of the focus and the field of attention.

Users visually process geographic information in a fast and global context-dependent manner before slowing down the scan path to a local mode of information processing. Figure 2 illustrates the dimension of the focus of attention (foA) and the surrounding field of attention (FoA) during a single gaze fixation within the field of view when processing geographic information on a desktop display (27cm x 34cm). According to the extensions and degrees of visual acuity, the geovisualisation is configured with corresponding values of a soft-focus tool to illustrate what and how users see when focusing the centre or shifting their gazes to the corner of a display. The FoA allows users to visually capture geographic context information during visual scanning before 'switching' to the foA where visual acuity and spatial contrast sensitivity amount to 100%. Both visual functions decrease rapidly towards the periphery. This centre-surround mechanism allows users to maintain a crude representation of geographic context information in the visual background for spatial orientation. Subsequently, during detailed local information processing, the scan path is guided to relevant information [16]. In contrast to desktop environments where visual scanning is affected by visual distractors on the display (e.g. irrelevant information), the scan paths of mobile users can additionally be deviated by distractors located in their egocentric space (Figure 3).

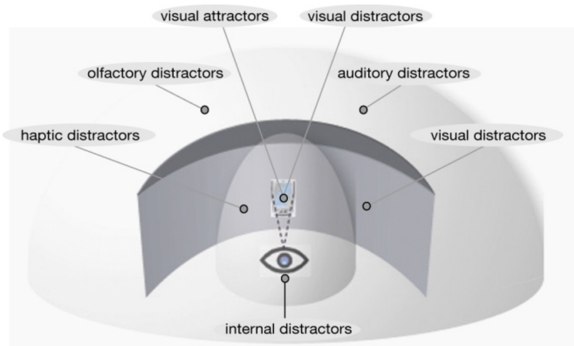


Fig. 3. Internal and external distractors

As a result of the decreased information space (small displays), the dimension of the geographic space (physical environment) increases. The space of distractive sensory input is enlarged to the egocentric space containing distractive stimuli that influence users in visual information processing. Visual attractors (geovisualisation) are in competition with haptic (e.g. rain and wind), olfactory (odours), auditory (e.g. car noise and conversations), and visual distractors (e.g. pedestrians and traffic signs). These external distractive stimuli are more or less extended in dependence of the usage context. While explorative users in offices visually scan for relevant information that is embedded in visual distractors on the display, attentional capacities of mobile users can be exceeded in environments of distractive influences. Moreover, users can be influenced by internal distractors like stress. For instance, fire men, soldiers, and air traffic controllers are often involved in time-critical decision-making tasks. For global as well as for local processing of geographic information, these groups of users need to achieve high visual scanning efficiency while investing small effort for promptly locating relevant geographic information and easily decoding underlying semantics.

3 Visual Scanning Efficiency

Whenever users visually scan for geographic information, they have to invest a certain capacity of attentional resources, regardless of their intention such as data exploration, analysis, synthesis as well as presentation; the basic goals of map-use identified by MacEachren and Kraak [17]. This cognitive load can be regarded as a multidimensional construct representing the load that tasks impose on the user's visual information processing system [18]. Based on the theoretical approach of 'instructional efficiency' [18] and the 'performance-resource function' [19] the model in Figure 4 proposes the axis of 'information complexity' as a determinant for relating basic parameters involved in the efficiency of visual scanning.

Note that information complexity can be determined either by complex visualisations (e.g. information rich geovisualisations) or complex visual scenes (e.g. stimulus rich environments), or by the combination of both. The proposed axis of information complexity derives from psychological visual search tasks where information complexity is a determinant parameter. If the slope of the reaction time (time required to

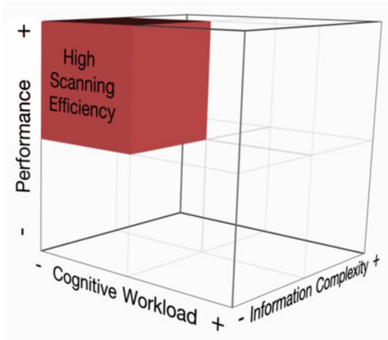


Fig. 4. Visual scanning efficiency

affirm or negate the presence of an item) multiplicated with the set size function (number of items on the display) is near zero, the efficiency of the task can be labelled as high and vice versa. With respect to geovisualisation design, the basic goal is to keep the reaction time (time needed to visually detect the most relevant information) as short as possible and to decrease the number of information to the minimum without neglecting spatial context information needed to accomplish geographic tasks. Information complexity is referred to as 'visual complexity' that is composed of the proportion of visually salient and visually alleviated geographic information.

4 Supporting Users in Visual Scanning

The basic challenge of our work is to keep the cognitive workload as low as possible by supporting users in visually scanning geovisualisations. We tackle this challenge by following an attention-guiding design methodology where relevant information is coded with attention-guiding graphical variables.

1. The attention-guiding design methodology is built on cognitive design principles that are reflected in classical rules of thematic map design. While 'simplicity' aims at reducing the visual complexity and 'visual hierarchy' tends to organise and structure the information into visual layers, 'conciseness' aims at visualising the important information in a salient way. As a result, the design methodology consists of three cognitively adequate geographic information layers, i.e. the design is adapted to basic characteristics of visual scanning. First, a top-down processed layer visualises geographic context information in an alleviated way to support users in global visual information processing. Second, a bottom-up processed layer depicts relevant information in a salient way to support users in local visual information processing. Third, the combination of both layers generates an attention-guiding layer where the relevance-based filtered information can be visually extracted from and related to the spatial context information.
2. To support users in visually detecting relevant information the cartographic symbols are visualised with graphical variables that are appropriate to code relevance classes and to guide visual attention to the location of interest. A collection of these

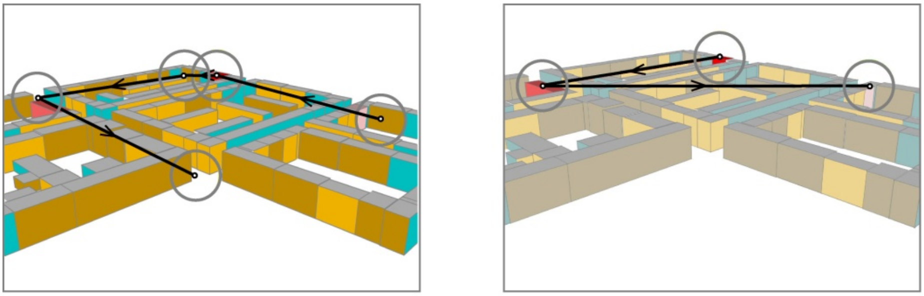
attention-guiding variables was recently proposed by Swienty et al. [20]. They related graphical variables that are postulated to be effective for coding ordinal data to attention-guiding attributes (e.g. colour) and corresponding values (e.g. red) that are suitable to guide visual attention.

5 Guiding Visual Attention in 3D and 2D Geovisualisations

For a validation of the attention-guiding design we present design examples that were applied to 3D geovisualisations for mobile devices and to 2D geovisualisations for desktop displays. The 2D geovisualisations were designed in the context of a recent research project that has developed, implemented and evaluated the attention-guiding approach. As introduced, we consider the relevance-based filtering as an utility criterion and the cognitively adequate visualisation as an usability criterion of a systems acceptability. Accordingly, we designed three test cases: Case 1 visualises unfiltered information in a cognitively adequate way. Case 2 visualises filtered information in a cognitively inadequate way and case 3 visualises filtered information in a cognitively adequate way, i.e. the design of attention-guiding geovisualisation is adapted to the skill of visual scanning. The evaluation of the 2D examples was conducted with the eye-movement recording method. The 3D examples were pre-evaluated with a computational attention-model to gain first knowledge of the design suitability when being applied to a 3D geovisualisation.

Non-photorealistic 3D geovisualisation. To improve visual scanning efficiency, the overall objective of attention-guiding geovisualisation is to visualise as much information as needed and as little as possible. Correspondingly, two basic challenges arise in the field of mobile devices: (1) How can we filter relevant information to reduce information complexity on the display? (2) How can we design the filtered information to support users in visually scanning the display? The first challenge is tackled by adapting the system to the current situation of a user by determining and calculating the components of geographic relevance (e.g. spatial, temporal, topical) [21]. The second challenge may be tackled by the high level of abstraction of non-photorealism that applies artistic techniques (e.g. ink painting) to computer graphics to visually highlight information of interest. A basic research challenge is to design 'meaningful' abstraction [22]. We assume that the 'meaning' of geographic information is determined by its location (where) and its relevance (what) in dependence of a users query to the system like visualising three relevant buildings in a specific mobile usage context (e.g. looking for three relevant restaurants). The relevance of these buildings may vary in dependence of the combination of different relevance types such as topical, temporal or spatial relevance [21]. To approach effective visual scanning, we have pre-evaluated a simplified non-photorealistic geovisualisation, i.e. we do not consider buildings with different heights and additional spatial context information like road networks. To adapt the non-photorealistic design to the users' skill of visual scanning, the relevance classes (what) and the location of buildings (where) are coded with attention-guiding graphical variables that are able to encode ordinal data and to guide a users attention to relevant information [20].

The pre-evaluation was conducted with a computational attention model that has been successfully validated against experimental evidence for visual search tasks [23]



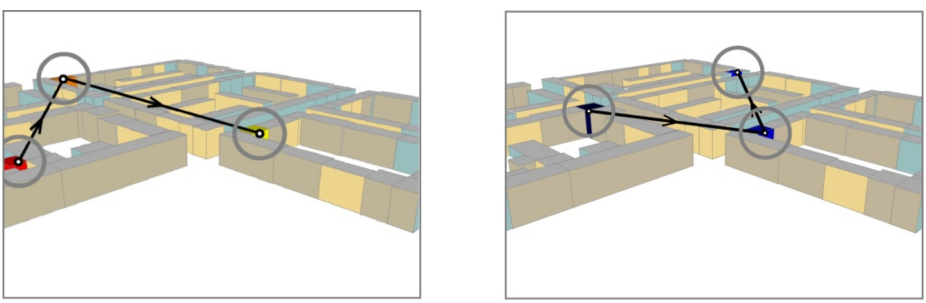
a b

Fig. 5. Guiding attention to relevant information with the variable ‘saturation’. a) case 2, b) case 3.

and has served as a promising pre-evaluation method for eye movement recording studies [24]. The model is based on the centre-surround mechanism and aims at predicting human gaze paths and fixations by extracting three attentively processed stimuli (colour, orientation, intensity contrasts) through multi-scale feature extraction. Figure 5 shows the outcome of predicted scan paths when visually scanning for relevant information coded with ‘saturation’ in case 2 and case 3.

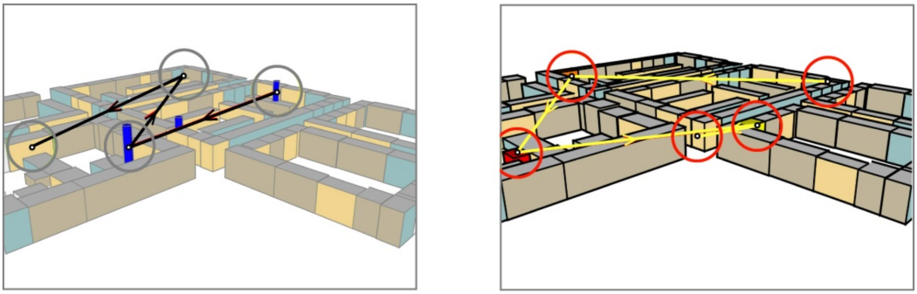
Figure 5a shows the predicted scan path when irrelevant information (irrelevant buildings) is saliently designed. Only the second and fourth fixation is guided to relevant information while the other fixations are deviated due to high colour and intensity contrasts. Figure 5b illustrates that the model systematically directs the focus of attention to relevant information in the order of decreasing saliency. Although the irrelevant information in the visual background is not alleviated to a maximum, the variable ‘saturation’ is suitable to overcome these visual distractors. The predicted scan path is not deviated from the target but systematically guided to the information of interest. Figure 6 shows the predicted scan paths when visually scanning for information coded with the variable ‘hue’ and ‘value’ in case 3.

Like ‘saturation’, relevant information in the visual foreground coded with ‘hue’ and ‘value’ seems to be efficient to guide visual attention. Another method to code relevance classes is symbolisation. Figure 7a illustrates the predicted scan path when



a b

Fig. 6. Guiding attention with the variable a) ‘hue’ and b) ‘value’ in case 3



a b

Fig. 7. Guiding attention with a) bar charts and b) 'hue' (contoured)

processing relevant information that is coded with the variable 'size' in terms of different heights of bars representing ordinal data. Due to high colour and orientation contrasts, the first fixation is guided to the second relevance class before capturing the most relevant information. Instead of fixating the third bar, the scan path is then deviated to irrelevant information due to high intensity (third fixation) and colour contrast (fourth fixation).

Figure 7b shows the 'hue' example and illustrates the deviating effect of homogeneously too strong contoured information. The nearer contours approach to each other and the wider the extension of the third dimension, the higher is the intensity contrast due to converging and overlapping contours.

2D geovisualisation. For a proof of the attention-guiding concept in 2D geovisualisation, we evaluated more or less complex visualisations with the eye movement recording method at the eye movement laboratory, Max Planck Institute of Psychiatry, Munich. A gaze fixation was recorded within a spatial area of 1° with a minimum duration of 100 ms. 15 participants (5 male, 10 female) with a mean age of 28 years (range: 22-38 years) took part in the study. Because of impaired visual acuity, one subject was excluded. The relevant information was coded with the variable 'contour' by visualising three contoured point symbols with three different line widths. The task on demand was: *Please search for the relevant information that is coded in three point symbols and say 'o.k.' when you have found all three point symbols.* To compare multiple means, an analysis of variance (ANOVA with Greenhouse-Geisser correction) was used.

All parameters show significant differences between the test cases (Table1). When visually scanning for relevant information in case 1, the scanning pattern illustrates that subjects needed the most time (5.49 sec.) to accomplish the task due to the highest degree of scan paths (77.75°) and the highest number of fixations (10.33). They employed 1.33 fixations to redirect their focus of attention and fixated the information in a mean time of 0.21 seconds. When visually scanning case 2 where relevant information was filtered but the visual background was not alleviated, the subjects scan paths revealed a mean degree of 32.77°. They needed 5.33 fixations and 3.01 seconds to accomplish the task. The last point symbol in the lower right corner was captured with the field of attention, i.e. the participant was not required to process the information in detail to identify the third relevant point symbol. To visually scan

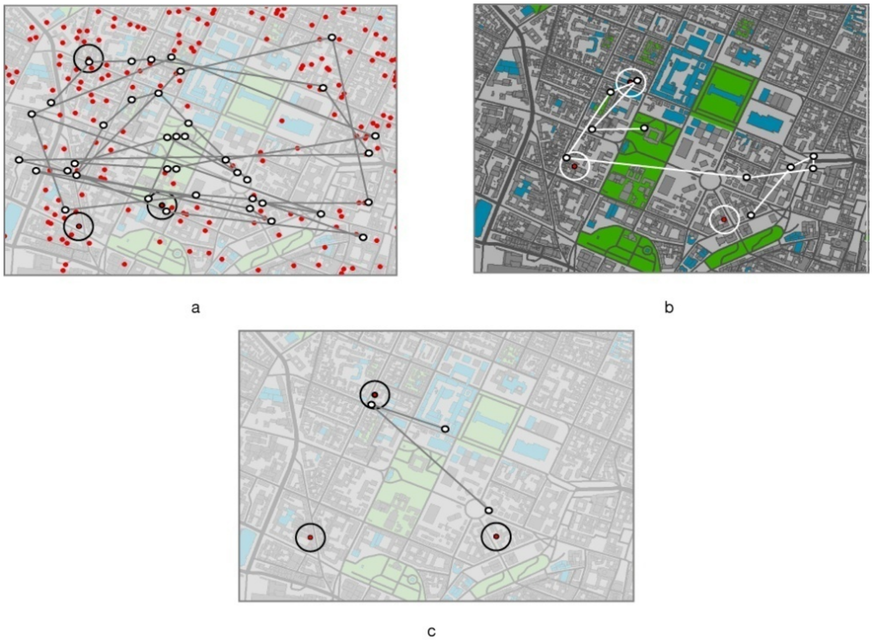


Fig. 8. Guiding attention with the variable ‘contour’. a) case 1, b) case 2, c) case 3. (base data are from Basis DLM of BKG, Germany; address data are from Städtisches Vermessungsamt Munich, Germany).

for relevant information in the attention-guiding design (case 3), participants needed a mean time of 1.95 seconds by employing 3.33 fixations. Their scan paths had a mean degree of 21.51°. The participants didn’t employ any re-fixation to find the information of interest and fixated the information in a mean time of 0.18 seconds. To visually scan for relevant information in case 3, participants even accomplished the task without the need to position the gaze fixation on the relevant information itself. The primarily global mode of visual scanning was sufficient enough to identify the three point symbols due to the attention-guiding approach, i.e. participants did not invest more attentional resources to accomplish the task due to the well structured and organised visual information.

Table 1. Visual scanning parameters. σ (standard deviation), p (significance).

	contour			
	case 1	**case 2**	**case 3**	**p**
time (σ)	5.49 (1.90)	3.01 (1.88)	1.95 (0.75)	<.001
degree (σ)	77.75 (35.28)	32.77 (15.42)	21.51 (9.12)	<.001
number of fixations (σ)	10.33 (5.50)	5.33 (2.89)	3.33 (1.40)	<.001
repetition of fixations (σ)	1.33 (1.84)	0.33 (0.62)	0.00 (0.00)	.019
duration of fixations (σ)	0.21 (0.04)	0.20 (0.04)	0.18 (0.02)	.031

6 Conclusions

The basic goal of this work is to significantly improve visual scanning efficiency when processing geovisualisations. Based on characteristics of visual scanning and performance-oriented cognitive workload models we proposed the visual scanning efficiency model that tends to release the cognitive workload by reducing the information complexity. Regardless of the goal of map use, we believe that this approach optimises the performance to reach high visual scanning efficiency. Therefore, 'visual hierarchy', 'simplicity' and 'conciseness' are considered as fundamental design principles. For a proof of concept we designed 3D and 2D geovisualisations based on an attention-guiding approach. The analysis of the visual scanning parameters revealed a high visual scanning efficiency of users when processing 2D geovisualisations, i.e. users needed the shortest time, employed the least number of fixations with the smallest degree of scan paths. Here, we solely focused on effectively visualising the location of relevant information. This does not necessarily implicate that the underlying semantics of the relevance classes can be easily decoded. A user is probably able to promptly locate the most important information and to relate this information to spatial dimensions. However, if the symbolisation is not appropriate to encode the semantics of the information, users have to employ more mental effort, which will decrease the efficiency of visual information processing. Further research in the field of semiotics will help to investigate the potentials of symbolisation to achieve high visual scanning efficiency. The outcome of these studies will help to optimise the speed and accuracy of visual geographic information processing. The attention-model revealed that visually alleviating irrelevant information and highlighting relevant information in 3D non-photorealistic visualisations might be a suitable method to support users in visually detecting information of interest. However, we have presented a simplified geovisualisation. Before augmenting the visual complexity by visualising e.g. shadows or tall buildings, we first concentrated on basic research to investigate users potentials and restrictions of locating information of interest and decoding relevance classes in 3D environments with low information complexity. Therefore, we do not consider additional means like annotations due to the doubtful status of numbers and letters to guide visual attention [25] that probably slow down the visual scanning process. Based on the efficiency of the variable 'size' to code numeric values in 2D geovisualisations, another method to visualise relevant information in 3D environments might be the design of different sized symbols like bars. The use of symbolisation to code numeric values in 3D environments must be handled with care. A recent study revealed that the task of comparing the values of two bars in 2D environments was completed less successfully in a 3D setting than in 2D [26]. The decoding of numeric values coded with varying heights and constant widths of bars seems to require some training to improve visual geographic information processing.

To summarise, we found that visual scanning efficiency can be improved in 2D geovisualisations by following cognitive design principles reflected in classical rules of thematic map design. Further research in the field of non-photorealism will elucidate the potential of this approach to support users in visual scanning 3D geovisualisations.

References

1. Hornsby, K., Egenhofer, M.: Identity-based change: a foundation for spatio-temporal knowledge representation. International Journal of Geographical Information Science 14, 207–224 (2000)
2. Worboys, M.F.: Event-oriented approaches to geographic phenomena. International Journal of Geographic Information Science 19, 1–28 (2005)
3. Gahegan, M., Takatsuka, M., Wheeler, M., Hardisty, F.: Introducing Geo-VISTA Studio: an integrated suite of visualizsation and computational methods for exploration and knowledge construction in geography. Computers, Environment and Urban Systems 26, 267–292 (2002)
4. Demšar, U.: Knowledge discovery in environmental sciences: visual and automatic data mining for radon problems in groundwater. Transactions in GIS 11, 255–281 (2007)
5. Coors, V., Kray, C., Elting, C., Laakso, K.: Presenting Route Instructions on Mobile Devices. In: Dykes, J., MacEachren, A., Kraak, M.-J. (eds.) Exploring Geovisualization, pp. 529–550. Elsevier, Amsterdam (2005)
6. Montello, D.R.: Cognitive map-design research in the twentieth century: Theoretical and Empirical Approaches. Cartography and Geographic Information Science 29, 283–304 (2002)
7. Chinchor, N., Hanrahan, P., Robertson, G., Rose, R.: Illuminating the path. In: Thomas, J.J., Cook, K.A. (eds.) Creating the R&D Agenda for Visual Analytics. IEEE, Los Alamitos (2005)
8. MacEachren, A.M., Kraak, J.M.: Research challenges in geovisualization. Cartography and Geographic Information Science 28, 3–12 (2001)
9. Slocum, T.A., Blok, C., Jiang, B., Koussoulakou, A., Montello, D.R., Fuhrmann, S., Nicholas, R.H.: Cognitive and Usability Issues in Geovisualization. Cartography and Geographic Information Science 28, 61–75 (2001)
10. Dykes, J., MacEachren, A.M., Kraak, M.-J.: Exploring Geovisualization. In Exploring Geovisualization. In: Dykes, J., MacEachren, A.M., Kraak, M.-J. (eds.) International Cartographic Association (ICA), pp. 3–19. Elsevier Oxford, Amsterdam (2005)
11. Nielsen, J.: Usability Engineering. Morgan Kaufmann Academic Press, London (1993)
12. Swienty, O., Zhang, M., Reichenbacher, T.: Attention Guiding Visualization of Geospatial Information. In: Proceedings of SPIE, Geoinformatics 2006 – Geospatial Information Technology, vol. 6421 (2006)
13. Reichenbacher, T., Swienty, O.: Attention-guiding geovisualisation. In: Cd-Proceedings of the 10th AGILE International Conference on Geographic Information Science (2007)
14. Corbetta, M., Shulman, G.L.: Control of goal-directed and stimulus-driven attention in the brain. Nature, Reviews-Neuroscience 3, 201–215 (2002)
15. Itti, L., Koch, C., Niebur, E.: A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. Transaction on Pattern Analysis and Machine Intelligence 20, 1254–1259 (1998)
16. Torralba, A., Oliva, A., Castelhano, M.S., Henderson, J.M.: Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. Psychological Review 113, 766–786 (2006)
17. MacEachren, A.M., Kraak, J.-M.: Exploratory Cartographic Visualization: Advancing the Agenda. Computer and Geoscience 23, 335–343 (1997)
18. Paas, F., Renkl, A., Sweller, J.: Cognitive load theory and instructional design: Recent developments. Educational Psychologist 38, 1–4 (2003)
19. Norman, D.A., Bobrow, D.G.: On data-limited and resource limited processes. Cognitive Psychology 7, 44–64 (1975)

20. Swienty, O., Zhang, M., Reichenbacher, T., Meng, L.: Establishing a neurocognition-based taxonomy of graphical variables for attention-guiding geovisualisation. In: Proceedings of SPIE-Geoinformatics, vol. 6751 (2007)
21. Reichenbacher, T.: The concept of relevance in mobile maps. In: Gartner, G., Cartwright, W., Peterson, M. (eds.) Location Based Services and TeleCartography, pp. 231–246. Springer, Heidelberg (2007)
22. Santella, A., de Carlo, D.: Visual Interest and NPR: an Evaluation and Manifesto. In: Proceedings of the Third International Symposium on Non-Photorealistic Animation and Rendering (NPAR), pp. 71–78 (2004)
23. Treisman, A., Gelade, G.: A feature integration theory of attention. Cognitive Psychology 12, 97–136 (1980)
24. Fabrikant, S.I., Montello, D.R., Rebich, S.: The Look of Weather Maps. In: Proceedings of the GIScience 2006, vol. 28, pp. 269–271 (2006)
25. Wolfe, J.M., Horowitz, T.S.: What attributes guide the deployment of visual attention and how do they do it? Nature reviews - Neuroscience 5, 1–7 (2004)
26. Bleisch, S., Dykes, J.: Evaluating the effectiveness of representing numeric information through abstract graphics in desktop 3D environments, Draft-Paper, From Geovisualization to geovisual analytics - Conference of ICA-Commission on Visualization and Virtual Environments, Helsinki (2007)

Navigation and Search in 3D Visualizations of Large Unstructured Photo Collections: An Empirical Study

Olivier Christmann and Noëlle Carbonell

Nancy-Université, LORIA, CNRS & INRIA
BP 239, 54506 Vandœuvre-lès-Nancy Cedex, France
{Olivier.Christmann,Noelle.Carbonell}@loria.fr

Abstract. We present an empirical study which aims at assessing the effects of dynamic 3D visualizations of randomly ordered photo collections on visual search effectiveness, efficiency and comfort. 20 participants performed visual search tasks in collections of about 1000 colour photos using 2 perspective views of a vertical cylinder: thumbnails were displayed on its lateral surface either on the inside (IV) or on the outside (OV). Scrolling IV suggests locomotion in an immersive virtual space while scrolling OV suggests manipulation of a 3D virtual object. Perspective distortions in OV "channel" gaze towards the centre of the screen while IV permits of freer gaze movements.

A majority of participants (12) performed noticeably better with one view (IV or OV) than with the other. Animation and perspective distortions influenced visual exploration strategies (16 participants). Preferences, which varied across participants, were mainly motivated by individual visual capabilities; the influence of interaction metaphors was marginal. Qualitative analyses of participants' behaviours suggest that IV has the potential to support spatial memory. These results indicate that adaptable perspective views may facilitate and improve visual search in unstructured picture collections.

Keywords: 3D visualizations, Interaction metaphors, Picture browsers, Visual search.

1 Introduction

The size of personal and community photo collections is rapidly growing, due to the spread of cheap digital cameras and their integration into everyday objects (e.g., cell phones). Retrieving specific photographs from photo collections is thus becoming increasingly difficult. Assisting search in large collections of photos is a critical issue which has been addressed by a number of researchers in human-computer interaction.

A current research direction is to provide users in the general public with advanced annotation facilities so as to encourage query-based photo retrieval. First attempts date back to the nineties; see [21], [14] and the ontology-based annotation tool proposed in [26]. At present, photo management and sharing applications on the Web, such as Flickr or Picasa, offer annotation facilities. However, people have neither the motivation nor the time for annotating hundreds or thousands of photos [15], [17]. They just save their personal photos in thematically labelled folders which may include loads of photo files.

M. Sebillo, G. Vitiello, and G. Schaefer (Eds.): VISUAL 2008, LNCS 5188, pp. 31–43, 2008.

An alternative approach focuses on facilitating visual search in unstructured or loosely structured collections by providing users with multi-scale interactive visualizations (MIVTs) where photographs are represented as thumbnails.

The empirical study presented here is a contribution to the second approach. It aims at assessing the potential of dynamic (i.e., scrollable) 3D MIVTs for supporting visual search in large unstructured photo collections. We compare the performances, behaviours and attitudes of participants who performed visual search tasks using two dynamic 3D representations of photo collections. Each 3D view suggests a specific interaction metaphor: locomotion in a virtual space vs manipulation of the visualization as a 3D virtual object. In addition, perspective distortions vary from one view to the other. Comparisons focus on eliciting the possible influence of visual distortions and interaction metaphors on visual search effectiveness, efficiency and comfort. Results and conclusions may prove useful for improving design of standard picture browsers which display 2D scrollable arrays of thumbnails, and picture or video retrieval from large structured banks [6]. Queries to such banks often generate unstructured loads of results, due to coarse or inappropriate picture indexing.

2 Related Work

Numerous multi-scale interactive visualization techniques have been proposed for facilitating visual exploration of very large data sets [29].

Focus+context visualizations [19], such as fisheye views [10, 2] or perspective walls [20], enhance items that users are currently paying attention to, while gradually reducing the visual prominence of other items as their distance from the user's current object of interest increases. Another class of MIVTs aims at providing users with overviews of hierarchically structured data spaces, while relying on multi-scaling for facilitating search and navigation. Tree-maps are well-known instances of this type of visualizations [28]. An untypical popular technique in this class, hyperbolic trees [18], displays focus+context overviews of hierarchical data structures. Visualization of large graphs (millions of nodes) has motivated the design of specific visual representations, for instance zoomable adjacency matrices [12].

The majority of MIVTs represent data sets with static 2D views; "static" here means that display changes are side effects of the user's actions such as zooming in on a treemap [5] or changing focus on a hyperbolic tree. Most 3D visualizations are also static, such as fisheye views, perspective walls and "Data Mountains" [24]. A few only are dynamic, for instance, cone trees [25] and some 3D treemaps [4] which users can rotate, or perspective views of documents which they can "fly" over [11].

These techniques have been applied to large structured sets of digital [12], textual [23] and multimedia [24] information. Static 2D visualizations, especially treemaps [3] and hyperbolic trees [15], have been used for visualizing small structured picture collections as thumbnails. Few specific techniques have been proposed save for a focus+context visualization [17] and an intuitive 2D visualization for a community library of 3,000 photos [16]. Fewer 3D MIVTs have been used for visualizing structured picture collections, and most of them are static [13].

Applications of MIVTs to large structured collections of pictures are scarce, due mainly to the lack of large banks of indexed pictures. Present image processing techniques can only detect coarse, visual or thematic, similarities [9]. In addition, consistent and versatile manual indexing of pictures is much more difficult and time consuming than textual and multimedia document annotation, due to the richness of pictorial information. A few attempts only have been made to apply MIVTs to large collections of pictures. For instance, [13] reports the implementation of two semantic fisheye views for a collection of 56,500 images annotated with a rich vocabulary of 28,000 words.

Although scrollable MIVTs are required for exploring unstructured photo collections, only two dynamic 3D MIVTs have been implemented for visualizing unstructured picture collections: a landscape of “bill-boarded” thumbnails (similar to the “Data Mountain”) which users can “walk” through or “fly” over, and a kind of carousel composed of photos positioned vertically between two horizontal disks which users can rotate [8]. The potential of dynamic 3D MIVTs to support visual search in unstructured picture collections has been scarcely investigated. How the specific features of these visualizations may influence and support visual search is an issue which has been ignored by researchers despite its possible impact on numerous widespread applications.

The study reported here addresses this issue based on two main assumptions. First, perspective distortions have the potential to “channel” the user’s gaze towards larger and/or less distorted items. Such gaze “channelling” effects may reduce the amplitude of eye movements and improve their coordination, especially when the positions of enhanced items on the screen do not vary during scrolling. Therefore, visual search in dynamic 3D visualizations where the shapes of enlarged photos are undistorted should prove to be quicker and more accurate than search in 3D visualizations where shape and size distortions affect the same photos. Secondly, depending on the virtual camera viewpoint, users may feel either as if they are immersed in the 3D visualization (locomotion metaphor) or as if they are manipulating it. Whether locomotion as an interaction metaphor better supports spatial memory tasks than manipulation is an open issue. Inconclusive results only have been obtained for a related issue, the potential of static 3D vs 2D MIVTs to support spatial memory [7].

The purpose of the study reported here is to assess the actual influence of the main specific features of dynamic 3D MIVTs, perspective distortions and interaction metaphors, on the effectiveness, efficiency and comfort of visual search in unstructured photo collections.

The methodology is described in the next section. Results of quantitative and qualitative analyses of participants’ behaviours and attitudes are presented and discussed in section 4. Main findings are summarized in section 5.

3 Method

We first present the two interactive perspective views designed for this study. Then, we describe experimental tasks, participants’ profiles, procedure and setup, working hypotheses and measures.

3.1 Visual Material

3.1.1 The Two Interactive Visualizations

To achieve meaningful comparisons, we designed two simple dynamic 3D visualizations which differ from each other in two features only: interaction metaphors (locomotion vs manipulation) and perspective distortions. Unstructured photo collections of about one thousand items (869) are represented using two views of a vertical cylinder which can be rotated from left to right and vice versa. Users can also move the cylinder forward or backward according to their visual acuteness. Thumbnails arranged in rows (11) and columns (79) are "plastered" either on the inside lateral surface of the cylinder (inner view, IV) or on its outside lateral surface (outer view, OV). 176 thumbnails are visible simultaneously (11 lines, 16 columns) as shown in Figure 1. Participants in a pilot study (7 out of 8) reported that, when using IV, they felt immersed in a virtual 3D space and had the impression of moving along a picture wall (locomotion metaphor). Contrastingly, they perceived OV as a kind of virtual photo holder they could rotate (manipulation metaphor). Perspective distortions vary across the two views. OV enlarges items in the central columns on the screen without distorting their shapes while IV enlarges items in the lateral columns and distorts their shapes. Therefore, greater gaze channelling effects are likely to be observed with OV than with IV.

The same curvature was chosen for both views; it was determined empirically through user testing so as to achieve an appropriate trade-off between visibility of perspective effects and easy interpretation of distorted items. IV and OV were projected on a wall screen (1.00 x 0.75 m) with 1280 x 1024 pixel resolution; participants were seated at a distance of 2.20 m from the screen.

We preferred 2.1/2 views to full 3D visualizations such as the "Data Mountain" so as to avoid occlusions which reduce the number of simultaneously visible items. OV has some similarities with the static "perspective wall" [20], while IV has common features with the static background view of bookcases in the virtual library described in [1]. IV and OV can be rotated like the "PhotoGoRound carousel" [8] and the "tripods" in the library [1]; both of which, however, do not prevent occlusions. A vertical slit on the cylinder surface (one empty column) helps users to keep track of navigation progress; they can use it as an indicator of full turn completion (see Figure 1). Replacing the

Fig. 1. The two visualizations: inside view (IV) and outside view (OV) of a vertical cylinder

circular base with an elliptical surface makes it possible to significantly increase collection size without reducing thumbnail dimensions.

3.1.2 Interaction Facilities

IV and OV were endowed with identical interaction facilities so as to avoid biasing comparisons between participants' attitudes and performances for each view.

The cylinder can be rotated (horizontal scrolling) by clicking on the left/right yellow arrow at the bottom of the display; see Figure 2. Rotation speed can be increased/decreased using the mouse wheel (12 speed levels). Clicking on one of the two other yellow arrows moves the visualization forward or backward. Pressing down the mouse right button stops the cylinder motion. The four control arrows are placed on the surface of a miniature overview of the current visualization. This miniature reproduces the visualization's rotation motions. A red vertical bar on its lateral surface indicates the current position, in the visualization, of the column in the center of the display. Thus, whenever users feel "lost" while navigating the current collection, they can look at the miniature overview in order to "take their bearings".

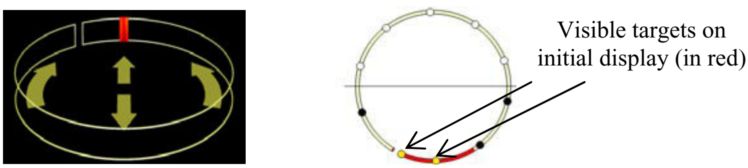


Fig. 2. Miniature cylinder with control arrows (IV) – Targets' positions on the cylinder

Clicking on a thumbnail with the mouse left button zooms in on it. The zoomed in item pops up in the centre of the screen (320 x 246 pixels). Clicking anywhere on the display with the mouse right button zooms it out. Fixed size zooming in on a particular item was preferred to continuous zooming in on the whole visualization which is useless for search in unstructured photo collections.

3.1.3 Photo Collections

Photo collections were built from a database of copyright free colour photos collected from popular websites and classified thematically (sport, animals, cars, etc). Each collection which reflects this thematic diversity includes equal numbers of landscapes (natural/urban landscapes, interior scenes) and objects, based on experimental results supporting the "Coarse-to-Fine" model of visual perception [27]. According to this model, pictures where high spatial frequencies predominate take more time to interpret/recognize than pictures where low spatial frequencies are dominant. Distributions of higher and lower spatial frequencies in photos of landscapes differ from those in photos of objects, as landscapes usually include more details than objects. Portraits and photos of animals were classified as "objects" in accordance with this model. Half targets are landscapes, and the other half are objects. Photos were placed randomly on the cylinder surface and targets were evenly distributed over it (see Figure 2).

3.2 Tasks, Participants and Procedure

3.2.1 Tasks

Participants performed two types of tasks with IV and OV (within-participant design):

- T1: Searching for a picture matching specific, thematic and visual, criteria; the target picture, its position in the visual representation of the collection and the representation itself are unfamiliar to the user.
- T2: Searching for a familiar picture in a familiar collection; the target and its position in the visual representation are familiar to the user who has had the opportunity to memorize them during previous interactions with the representation.

T1 and T2 are basic visual tasks which users perform repeatedly when searching in unstructured picture collections. So, results of this study will be useful for the design of photo browsers or visualizations of picture sets retrieved from large picture banks.

Implementation of T1 Tasks. Each T1 task consists in retrieving a specific photo from the visualization of a collection using a written description of this target photo. The target is always present in the collection and can easily be identified from its verbal description. Descriptions are short (from 4 up to 6 sentences/phrases), and their information content and structure are stereotyped. Each description includes four types of information: the topic of the photo and its salient visual features, details meant to help participants elaborate an accurate mental representation of the target (shooting angle and centring, scene components and organization, additional discriminatory details). Candidate targets that might be confused with other photos in the collection or could not be described meaningfully and unambiguously were eliminated following a pilot study (8 participants).

When participants were finished with reading a target description, they began searching for it in the current visualization. They were instructed to focus their efforts on accuracy rather than speed although search time was limited to 3 minutes so as to limit eyestrain. Selection of a photo by clicking on it with the mouse left button finished off the task. Then, an animation showed the quickest way to reach the target from the initial view and zoomed on it so that all participants had the opportunity to become familiar with its visual features and position in the representation.

Implementation of T2 Tasks. T2 tasks were carried out following T1 tasks. They consist in retrieving the same target photos as those used for T1 tasks in the same collections. Each target (in zoomed format) is displayed during 3 seconds in the centre of the screen. Then the visual representation of the collection is displayed and visual search can start. Unlike T1 tasks, each T2 task ends with the selection of a thumbnail. Participants were requested to perform T2 tasks as fast as they could.

3.2.2 Participants' Profiles

20 experienced computer users with ages between 21 and 34 years (average: 25.15, SD: 3.13) participated in the study. All of them had normal sight according to the Bioptor test of monocular and binocular vision (Stereo Optical Co., Inc.). These constraints on participants' profiles were meant to limit inter-individual differences in mouse control, motor response speed and visual acuteness. All participants were French native

speakers, since T1 tasks required flawless understanding of target written descriptions. The group was gender-balanced to take account of possible gender differences in performances and attitudes.

3.2.3 Procedure and Setup

After the vision test, participants carried out 2 training T1 tasks with one of the two views to get familiar with the user interface. Then, they performed a block of 10 T1 tasks with this view. The same procedure was repeated for the other view. The order of interaction with IV and OV was counterbalanced among participants. All targets and collections used for T1 tasks were different. Following T1, participants performed T2 tasks using each view in the same order as during T1. The same collections and targets as during T1 were used but they were presented in a different order to limit task learning effects. Two random task orders were defined per subject. After completing T2, participants expressed their subjective judgments and preferences in a verbal questionnaire and during a short debriefing interview. All participants' interactions with the two views were recorded.

3.3 Working Hypotheses and Measures

Analyses and comparisons have been focused on assessing (i) the actual contribution of dynamic perspective visualizations to the efficiency and comfort of visual search, and (ii) the possible effects of interaction metaphors, locomotion and manipulation, on spatial memory (i.e., memorization of T1 targets' locations in the visualizations). We expect that:

H1. Better performances will be achieved with OV than with IV as gaze channelling effects due to perspective distortions are more salient and helpful in OV.

H2. Assuming that both views have the potential to support spatial memory tasks, the locomotion metaphor and feeling of immersion induced by IV will better support memorization of targets' locations on the cylinder wall, as studies of human locomotion indicate that people form enduring representations of object-to-object spatial relations when discovering the layout of a novel real environment [22].

For each type of task (T1, T2) and each of the 2 views (IV, OV), the following measures have been computed over all participants (20):

- T, mean visual search time.
- S, mean number of successful target selections.
- R, mean rotation speed of the cylinder (columns/sec.).
- C, mean number of scrolled columns during target search.

4 Results: Presentation and Discussion

We first present participants' performances, subjective judgments and visual exploration strategies, and discuss them in relation to H1. Then, findings stemming from interaction log analysis are summed up and their support to H2 considered.

4.1 Participants' Performances

View processing order had no significant effect on participants' performances (Mann-Whitney test), which made it possible to merge data from all participants. Nonparametric Wilcoxon tests were used, as the distributions of some data samples did not fit with the normal distribution (Kolmogorov-Smirnov test). Results of comparisons between the two views are presented, for T1 and T2 tasks, in Table 1.

IV appears to support T1 tasks better than OV: errors are significantly fewer (p=.011), and search times seem shorter (tendency, p=.093). Results of other comparisons are not statistically significant. Qualitative analyses indicate that participants missed 27 targets out of 200 with IV against 43 with OV (time limit exceeded). Large differences in search times (20% or more) were observed between IV and OV for 12 participants, 9 of them carrying out T1 tasks much faster with IV. These results suggest that searching for unfamiliar targets matching verbal descriptions is easier and more effective with IV than with OV.

Table 1. Participants' performances (means over T1/T2 tasks): search time (T, sec.); number of successful target selections (S); number of scrolled columns (C); speed rotation (col./sec.)

20 participants 10 T1/T2 tasks		**T1 tasks** (10)			**T2 tasks** (10)			
		T	**S**	**C**	**T**	**S**	**R**	**C**
Mean (*SD*)	**IV**	821 *(200)*	7.8 *(1.9)*	587 *(220)*	441 *(172)*	9.3 *(0.8)*	1.79 *(.53)*	654 *(330)*
	OV	939 *(193)*	6.7 *(1.6)*	622 *(233)*	485 *(164)*	9.3 *(0.8)*	1.43 *(.41)*	561 *(219)*
Wilcoxon test		z=1,680 p=.093	z=2.536 p=.011	z=0.784 p=.433	z=1.195 p=.232	z=0.000 p=1.000	z=3.771 p<.0005	z=1.344 p=.179

Contrastingly, participants' performances with IV and OV are similar for T2 tasks. Differences between search times are not statistically significant, and errors are too few to be considered. Rotation speed is much higher with IV (p<.0005) while navigation paths (or number of scrolled columns) are shorter with OV (tendency, p=.179). Effects of rotation speed and navigation path length on search times may have neutralized each other. This may explain why search times with OV and IV were not significantly different during T2.

To sum up, IV appears to better support visual exploration of picture collections than OV (in T1 tasks), while the effects of both views on search for familiar pictures (in T2 tasks) are similar. Differences in perspective distortions (between IV and OV) together with differences in visual activities (between T1 and T2 tasks) may explain these results. A larger number of thumbnails can be surveyed at a glance with IV, while OV better supports close observation of thumbnails. Participants reported that, for T1 tasks, they navigated the collections in search of photos matching major visual features in the current target description and, whenever they detected a likely candidate, they checked whether it fitted in with details in the description. By facilitating global visual surveys, IV may have helped participants to implement this strategy efficiently, thus contributing to limit target misses during T1. Significantly higher rotation speed with IV during T2 confirms the superiority of IV over OV for supporting visual surveys. On the other hand, recognition of visually familiar photos requires closer examination of thumbnails. In keeping with H1, shorter search times should have been obtained with OV for T2 tasks, since OV enlarges undistorted thumbnails in the centre of the screen. However, the influence of perspective distortions may have been thwarted by other factors the influence of which could also account for the high diversity of individual performances (see SDs in Table 1).

The next section describes participants' preferences and visual exploration strategies from analyses of questionnaires and debriefings which also suggest plausible explanations for the high diversity of participants' performances and the apparently weak influence of perspective distortions on them.

4.2 Participants' Preferences and Visual Exploration Strategies

Table 2 shows that the majority of participants (14) expressed preference for one view over the other. 10 preferred IV (GI group), and 4 liked OV better (GO group). The 6 remaining participants either changed opinions between questionnaires and debriefings or liked/disliked both views (GH group). GO includes only female participants while GI is gender balanced. However, the influence of gender on preferences cannot be inferred from this distribution, due to the small size of samples.

Table 2. Participants' preferences and visual exploration strategies

<table>
<tr><th colspan="2">Group</th><th colspan="10">GI group</th><th colspan="4">GO group</th><th colspan="6">GH group</th></tr>
<tr><th colspan="2">Participant</th><th>1</th><th>5</th><th>6</th><th>8</th><th>11</th><th>14</th><th>15</th><th>17</th><th>19</th><th>20</th><th>7</th><th>9</th><th>10</th><th>16</th><th>2</th><th>3</th><th>4</th><th>12</th><th>13</th><th>18</th></tr>
<tr><th rowspan="2">Strategy</th><th>IV</th><td rowspan="2">I</td><td rowspan="2">I</td><td rowspan="2">I</td><td rowspan="2">I</td><td rowspan="2">I</td><td rowspan="2">C</td><td rowspan="2">C</td><td rowspan="2">I</td><td>I</td><td>I</td><td>I</td><td>?</td><td rowspan="2">I</td><td>C</td><td>C</td><td>I</td><td>?</td><td>?</td><td>I</td><td rowspan="2">I</td></tr>
<tr><th>OV</th><td>C</td><td>?</td><td>C</td><td>C</td><td>I</td><td>I</td><td>C</td><td>?</td><td>?</td><td>C</td></tr>
</table>

Preferences were mainly motivated by perspective distortions and their effects on individual visual strategies. Enlargement of items in the central columns of OV was highly-rated by 6 participants. On the other hand, 5 participants appreciated the wide overviews of items provided by IV; compared to OV, IV actually displays a larger number of undistorted items in the centre of the screen; central columns are also shorter, hence easier to scan vertically. Larger items on the sides of IV were also much appreciated by 7 participants. Visual features of the views were most often mentioned as the main or even sole explanation for preferences. 6 participants only referred to the feeling of immersion induced by IV as an incidental justification for their choice. Thus, preferences were mainly influenced by perspective distortions; effects of interaction metaphors were only marginal.

Visual exploration strategies were obtained from debriefings and interaction logs. Analyses were focused on visual search during rotation of the views, since participants, when searching for a target, just glanced over the initial display before launching rotation, and only stopped it to scan likely target candidates. Strategy identification was successful for 16 participants and failed for the remaining 4 partly or totally, as shown in Table 2. Participants, while navigating collections, focused their gaze either on items in the central columns (*C* strategy), or on items in the lateral columns coming into view (*I* strategy). 6 participants used flexible strategies. 4 of them took advantage of perspective distortions, especially enlargement of central/lateral items, to improve their visual comfort: they used *C* for OV and *I* for IV, consistently with H1. The other 2 did the opposite, maybe to reduce the amplitude of vertical eye movements. As for the 10 remaining participants, they used the same strategy for the two views.

Additional comments from participants suggest that other visual factors than perspective distortions may exert an influence on gaze search strategies when interacting with dynamic 3D views. Several participants who used *I* for both views explained that scanning incoming columns was easier and more reliable than exploring central columns; the lateral screen edge along incoming columns helped them to control vertical scanning; on the other hand, they had difficulty in keeping track of central columns while exploring them. Better support of vertical scanning joined to better visibility of items in incoming columns may explain why 13 participants out of 17 used *I* with IV and only 7 used *C* with OV. These features might also explain why the majority of participants who used *I* for both views (6/8) preferred IV to OV while preferences of those who applied both strategies were varied. Perspective distortions seem to have exerted less influence on participants' visual strategies than animation. The fact that a majority of participants used the same strategy with both views suggests this conclusion, since IV and OV implement different static features but the same dynamic feature, namely horizontal scrolling.

To sum up, the visual features of IV and OV had a greater influence on participants' preferences than interaction metaphors meant to induce or not a feeling of immersion. These features include, in addition to perspective distortions affecting item visibility, horizontal scrolling which interferes with vertical scanning of items. The 4 visual exploration strategies which were identified illustrate how participants took advantage of static and dynamic features of IV and OV for improving item visibility and gaze control, thus increasing visual search efficiency and comfort. Inter-individual diversity in visual capabilities and strategies seems responsible for the high differences between individual performances reported in section 4.1. However, visual exploration strategies used with dynamic 3D visualizations need to be further investigated using sophisticated observation techniques such as eye tracking, due to their critical influence on visual search efficiency and comfort. Such empirical data are needed to revise H1 and generalize its scope by taking into account effects of scrolling in addition to static perspective distortions on performances and strategies.

4.3 Participants' Behaviours: Spatial Memory Activity

To determine whether participants had memorized T1 targets' actual positions on the cylinder, we analyzed the rotation directions they had chosen for reaching 3 T2 targets. These 3 targets are those for which the difference in length between the two possible navigation paths towards them is the greatest (see the black targets in Figure 2). Thus, if these targets' positions have been memorized during T1, it is likely that the shortest paths will be chosen to reach them during T2. Participants were just instructed to perform T2 tasks as fast as they could; they were not requested to choose the shortest ways to T2 targets since it might have affected task realism. This may explain why the majority of participants used the same rotation direction for all T2 tasks; 6 of them only alternated rotation directions. Interaction logs suggest that these 6 participants memorized target locations better with IV than with OV: for the 3 selected targets, the shorter paths were spontaneously chosen 2.5 times on average with IV against 1.75 with OV. These qualitative results confirm H2. They are at variance with the conclusions of the studies described in [7] which consider simple artificial search tasks in static 3D visualizations while our study focuses on realistic search tasks in dynamic 3D visualizations.

These findings make it worth while to further investigate whether dynamic immersive 3D visualizations implementing locomotion as interaction metaphor actually help users to memorize navigation paths more effectively than non immersive dynamic 3D or 2D visualizations. To address this issue appropriately, it is first required to design a methodology that provides objective reliable measures for assessing spontaneous memorization of navigation paths and item locations during realistic visual search activities in virtual spaces like T2 tasks. These measures should not interfere with task realism and spatial memory spontaneous activity.

5 Conclusion

We have presented an empirical study of the efficiency and comfort of visual search in dynamic 3D visualizations of large unstructured photo collections. 20 participants performed two types of visual search tasks (T1, T2) using two scrollable 3D views of collections including about a thousand colour photos. We designed 2 perspective views of a vertical cylinder: thumbnails are displayed on the cylinder lateral surface either on the inside (IV) or on the outside (OV). Interaction with IV suggests locomotion and a feeling of immersion while interaction with OV implements manipulation. Perspective distortions in OV “channel” gaze towards the center of the screen while IV permits of freer gaze movements.

Contrary to our expectations, participants’ performances with the two views differed only slightly, due to large inter-individual differences. Error rates for T1 tasks were lower using IV than OV, and rotation speed for T2 tasks was higher. Preferences were mainly motivated by visual features of the views; the feeling of immersion induced by IV was taken little account of. Half participants preferred IV, 4 chose OV and 6 expressed no clear opinion. 4 visual exploration strategies were identified: focusing gaze on central (*C*) or lateral incoming (*I*) columns for both views, using *I* for IV and *C* for OV or vice-versa. All participants took advantage of static and dynamic visual features of both views for improving item visibility and gaze control. The diversity of individual visual capabilities may account for the high inter-individual differences in performance observed for T1 tasks. Perspective distortions seem to have exerted less influence on participants’ visual strategies than dynamic features. The fact that IV and OV implement the same dynamic feature, horizontal scrolling, may explain why performances with IV and OV did not differ much, although participants took advantage of their static and dynamic visual features to improve search efficiency. Contrary to [7] and in line with [24], qualitative analyses of participants’ behaviours suggest that IV has the potential to better support spatial memory than OV.

In order to meet users’ needs and expectations, a possible short-term solution is to offer them an appropriate selection of dynamic 3D visualizations of photo collections such that positions of enlarged items should vary between visualizations. Further research on dynamic 3D visualizations is needed to elaborate advanced design recommendations. Visual search strategies in scrollable perspective views need to be better understood and modelled as their influence on performance seems decisive. We are currently replicating the study reported here using more elaborate observation techniques (e.g., eye tracking). Participants experiment a scrollable 2D visualization in addition to IV and OV. Our main aim is to investigate the potential of 3D views to improve visual search efficiency and comfort compared to 2D views.

References

1. Andrade, B., Almeida, R., Cubaud, P., Dupire, J., Natkin, S., Topol, A.: Experiments towards 3D Immersive Interaction for Digital Libraries. In: Pan, Z., Aylett, R.S., Diener, H., Jin, X., Göbel, S., Li, L. (eds.) Edutainment 2006. LNCS, vol. 3942, pp. 1348–1357. Springer, Heidelberg (2006)
2. Bederson, B.B.: Fisheye menus. CHI Letters 2(2), 217–225 (2000)
3. Bederson, B.B.: PhotoMesa: a Zoomable Image Browser Using Quantum Treemaps and Bubblemaps. CHI Letters 3(2), 71–80 (2001)
4. Bladh, T., Carr, D., Scholl, J.: Extending tree-maps to three dimensions: a comparative study. In: Masoodian, M., Jones, S., Rogers, B. (eds.) APCHI 2004. LNCS, vol. 3101, pp. 50–59. Springer, Heidelberg (2004)
5. Blanch, R., Lecolinet, E.: Browsing Zoomable Treemaps: Structure-Aware Multi-Navigation Techniques. IEEE Trans. on Visualization and Computer Graphics 13(6), 1248–1253 (2007)
6. Christel, M.: Establishing the Utility of Non-text Search for News Video Retrieval with Real World Users. In: Multimedia 2007, pp. 706–717. ACM Press, New York (2007)
7. Cockburn, A., McKenzie, B.: Evaluating spatial memory in two or three dimensions. Int. J. of Human-Computer Studies 61, 359–373 (2004)
8. Combs, T.T., Bederson, B.B.: Does Zooming Improve Image Browsing. In: DL 1999, pp. 130–137. ACM Press, New York (1999)
9. Frome, A., Singer, Y., Malik, J.: Image Retrieval and Classification using Local Distance Functions. In: NIPS 2006. pp. 417–424 (2006)
10. Furnas, G.W.: Generalized Fisheye Views. ACM SIGCHI Bulletin 17(4), 16–23 (1986)
11. Guiard, Y., Du, Y., Chapuis, O.: Quantifying degree of goal directedness in document navigation: Application to the evaluation of the perspective-drag technique. In: CHI 2007, pp. 327–336. ACM Press, New York (2007)
12. Henry, N., Fekete, J.-D.: Matrix Explorer: a Dual Representation System to Explore social Networks. IEEE Trans. on Visualization and Computer Graphics 12(5), 677–684 (2006)
13. Janecek, P., Pu, P.: An Evaluation of Semantic Fisheye Views for Opportunistic Search in an Annotated Image Collection. Int. J. on Digital Libraries 5(1), 42–56 (2005)
14. Kang, H., Shneiderman, B.: Visualization Methods for Personal Photo Collections: Browsing and Searching in the PhotoFinder. In: ICME2000, pp. 1539-1542 (2000)
15. Kuchinsky, A., Pering, C., Creech, M., Freeze, D., Serra, B., Gwizdka, J.: FotoFile: A Consumer Multimedia Organization and Retrieval System. In: CHI 1999, pp. 496–503. ACM Press, New York (1999)
16. Kules, B., Kang, H., Plaisant, C., Rose, A., Shneiderman, B.: Immediate usability: a case study of public access design for a community photo library. Interacting with Computers 16(6), 1171–1193 (2004)

17. Kustanowitz, J., Shneiderman, B.: Meaningful Presentations of Photo Libraries: Rationale and Applications of Bi-level Radial Quantum Layouts. In: DL 2005, pp. 188–196. ACM Press, New York (2005)
18. Lamping, J., Rao, R., Pirolli, P.: A Focus+Context Technique Based on Hyperbolic Geometry for Visualizing Large Hierarchies. In: CHI 1995, pp. 401–408. ACM Press, New York (1995)
19. Leung, Y.K., Apperley, M.D.: A review and taxonomy of distortion-oriented presentation techniques. ACM Trans. on Computer-Human Interaction 1(2), 126–160 (1994)
20. Mackinlay, J.D., Robertson, G.G., Card, S.K.: The perspective wall: detail and context smoothly integrated. In: CHI 1991, pp. 173–179. ACM Press, New York (1991)
21. Mills, T.J., Pye, D., Sinclair, D., Wood, K.R.: Managing photos with AT&T Shoebox. In: SIGIR 2000, p. 390. ACM Press, New York (2000)
22. Mou, W., McNamara, T.P., Rump, B.: Roles of Egocentric and Allocentric Spatial Representations in Locomotion and Reorientation. Int. J. Experimental Psychology: Learning, Memory, and Cognition 32(6), 1274–1290 (2006)
23. Pirolli, P., Card, S.K., Van Der Wege, M.: The effect of information scent on searching information: visualizations of large tree structures. In: AVI 2000, pp. 161–172. ACM Press, New York (2000)
24. Robertson, G., Czerwinski, M., Larson, K., Robbins, D., Thiel, D., van Dantzich, M.: Data mountain: Using Spatial Memory for Document Management. In: UIST 1998, pp. 153–162. ACM, New York (1998)
25. Robertson, G., Mackinlay, J.D., Card, S.K.: Cone trees: animated 3D visualizations of hierarchical information. In: CHI 1991, pp. 189–194. ACM Press, New York (1991)
26. Schreiber, A.T., Dubbeldam, B., Wielmaker, J., Wielinga, B.: Ontology-Based Annotation. Intelligent Systems 16(3), 66–74 (2001)
27. Schyns, P.G., Oliva, A.: From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition. Psychological Science 5(4), 195–200 (1994)
28. Shneiderman, B.: Tree visualization with tree-maps: 2D space-filling approach. ACM Trans. on Graphics 11(1), 92–99 (1992)
29. Ware, C. (ed.): Information Visualization. Perception for design, North-Holland. Elsevier, Amsterdam (2004)

Evaluation of Information Visualization Tools Using the NFR Approach

Pushpa Kumar[1], Nary Subramanian[2], and Kang Zhang[1]

[1] University of Texas at Dallas, 2601 N. Floyd Road,
Richardson, TX 75080, USA
{Pkumar,kzhang}@utdallas.edu
[2] University of Texas at Tyler, 3900 University Blvd.,
Tyler, TX 75799, USA
nsubramanian@uttyler.edu

Abstract. Determination of quality of information visualization (infovis) tools will assist practitioners in industry and academia to choose the appropriate tools for their needs. Infovis tools are widely used to visualize data-intensive applications so that quick decisions may be taken. In this paper we use a novel technique, the NFR Approach, to determine quality of infovis tools. The NFR Approach considers visualizability as an important non-functional requirement (NFR) for infovis tools and evaluates visualizability to determine the quality of these tools. The chief artifact that the NFR Approach employs for evaluation is the softgoal interdependency graph (SIG) wherein factors affecting visualizability are related to tool features by means of contributions that come in four flavors. Well-defined propagation rules help determine a qualitative score for visualizability based on the contributions. The chief advantages of the NFR Approach include rationale for the scores, flexibility to accommodate different definitions for visualizability, and historical record-keeping. The NFR Approach is validated by applying the Approach to three infovis tools – Prefuse, Tom Sawyer, and GnuPlot.

Keywords: visualization, quality, evaluation, tools.

1 Introduction

Quality has been defined as one or more desirable characteristics that a product should possess [1]. An important characteristic of information visualization tools is visualizability which refers to their ability to visualize data-intensive applications. Determination of quality of information visualization tools (also called infovis tools [4]) is important to practitioners in industry and academia so that they may select the most appropriate tool for their requirements. For example, some tools support rich interactive user interfaces while others support command-line interfaces; some tools offer customizable algorithms while others use predetermined algorithms; and some tools provide rich graphic capabilities while others offer limited graphic options.

This paper uses a novel technique called the NFR Approach [11] to qualitatively evaluate visualizability characteristic of infovis tools. The NFR Approach considers

M. Sebillo, G. Vitiello, and G. Schaefer (Eds.): VISUAL 2008, LNCS 5188, pp. 44–55, 2008.

visualizability as a non-functional requirement (NFR) to be achieved by a tool and employs a goal- directed graph called softgoal interdependency graph (SIG) that treats visualizability to be a goal to be achieved by any infovis tool. The SIG decomposes visualizability into its constituent factors based upon a definition of visualizability. The features of a tool are then taken into account when determining the contributions made by the tool to the visualizability factors. The contributions determine the extent to which features satisfy a factor and are of four types; strongly positive, positive, negative, strongly negative. The well-defined propagation rules of the NFR Approach help to propagate the contributions up the SIG to finally obtain the qualitative score for visualizability. Other techniques used for evaluating infovis tools are discussed below.

Usability study of information visualization techniques [2] separates usability issues into three main categories: visual representation usability, interface usability and data usability, devoted mainly to the quality. This technique links interface usability knowledge, concepts and methods with evaluation of the expressiveness, semantic content, and interaction facilities of visualization techniques. Usability inspection methods and user testing are used for evaluating usability.

An empirical comparison of visualization tools [3] studies the effects of different visualization techniques on user performance on information searching tasks and the effects of different sizes of the web spaces. Visualizations of subspaces on the World Wide Web can provide users the ability to identify relevant information from a set of web pages, while gaining new insights or understanding of the space. This paper experimentally tests the three classes of visualization techniques, distortion, zoom, and expanding outline, to better understand which classes of visualization techniques may better represent the underlying structure.

Evaluation of information visualization [7] proposes initial steps such as the development of benchmarks and repositories for information visualization, and refined evaluation methodologies and toolkits. This paper investigates three examples of transformations from prototype to product. A common evaluation measure for any technology is adoption by others, and the move into commercial products. It compares three examples of transformations from prototypes to products and applications; Film Finder to Spotfire, Treemap to the Map of the Market, a prototype to an interface developed for disseminating Census data the DataMap.

However, the NFR Approach gives a systematic framework for qualitatively evaluating visualizability that not only results in a score but also gives us the justifications for it. This paper validates the NFR Approach by applying it to three infovis tools, namely, Prefuse [4, 5, 6], Tom Sawyer [8, 9], and GnuPlot [10].

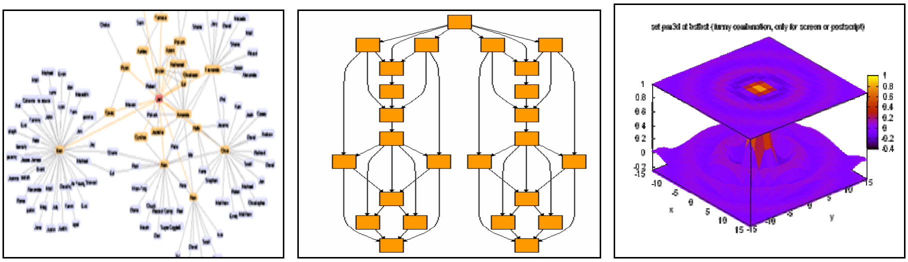
This paper is organized as follows: Section 2 discusses the three tools used for validation, Section 3 describes the NFR Approach, Section 4 applies the NFR Approach to the three tools used for validation, Section 5 discusses our observations in applying the NFR Approach, while Section 6 presents our conclusions.

2 Examples of InfoVis Tools

Prefuse [4, 5, 6] is an open source software framework that helps to create interactive information visualization applications using the Java programming language. It can be used to build standalone applications, visual components embedded in larger applications, and web applets. Animated force directed layouts can be generated easily using Prefuse and it has a physical force simulation engine. A library of interaction controls provides the user interactive operations. Output views have flexibility and view transformation support panning and zooming. A built in SQL-like expression language for writing queries to prefuse data structures and support for issuing queries to SQL databases and mapping query results into prefuse data structures is present. Prefuse has simple and developer-friendly APIs for creating; rendering components and source data is mapped into data tables that back visualization. The visual abstraction is then used to create interactive views of the data. The main advantage of this tool is it provides well for user interaction.

The *Tom Sawyer Visualization* product family [8, 9] enables users to develop graph visualization applications quickly and efficiently. This licensed software supports multiple editions like ActiveX, ASP.NET, Java, JSP, and MFC Editions that quickly and easily integrate into applications. Various industries that currently use these products can be classified as Academic Research, Computer Graphics, Defense and Intelligence, Engineering Design, Enterprise Business, Financial Services, Life Sciences, Networking, and Software Engineering. Software engineering applications address the analysis and manipulation of program structure. Reverse engineering functions usually generate extremely large control flow diagrams. Forward engineering functions focus on the inverse process where software architecture designers define class structure, inheritance relations and associations diagrammatically with state charts and use-case diagrams. The hierarchical layout technology can scale to display thousands of nodes and edges in each diagram, visualizing the structure of very large software programs. It also supports incremental layout technology so that users can maintain perspective as large diagrams are explored. The products allow users to create an intuitive visual representation of complex network systems, financial flows, social network structures, system infrastructure hierarchies, organizational systems, logistical and operational plans, and other relational data structures. Users can dynamically update their diagrams to incorporate new information immediately. A real-time view of the data is generated and the products provide a complete diagramming framework that enables users to interactively create visual models.

Gnuplot [10] is a portable command-line driven interactive data and function plotting utility for UNIX, IBM OS/2, MS Windows, DOS, Macintosh, VMS, Atari and many other platforms. The software is open source and was originally intended as to allow scientists and students to visualize mathematical functions and data. It has grown to support many non-interactive uses, including web scripting and integration as a plotting engine for third-party applications like Octave. Gnuplot has been supported and under development since 1986. Gnuplot supports many types of plots in either 2D and 3D and can draw using lines, points, boxes, contours, vector fields, surfaces, and various associated text. It also supports various specialized plot types. Gnuplot supports many different types of output, interactive screen terminals direct output to pen plotters or modern printers, and output to many types of files. Gnuplot is easily extensible to include new devices. Figures 1 a), b) and c) represent outputs for each of the above tools.



(a) (b) (c)

Fig. 1. Sample Outputs of three Infovis tools: (a) Prefuse (b) Tom Sawyer (c) GnuPlot

3 NFR Approach for Evaluating InfoVis Tools

In order to evaluate infovis tools we apply a qualitative reasoning approach called the NFR Approach [11] which is a goal-oriented approach and the goal here is to achieve good visualization for infovis tool – the ability which we refer to as *visualizability* of the tool. The NFR Approach is based on the NFR Framework [9] and views visualizability as a non-functional requirement (NFR) that needs to be satisfied by the infovis tools. In order to apply NFR Approach, a structure called the Softgoal Interdependency Graph (SIG) is developed which will be used to qualitatively evaluate visualizability [11]. The SIG development consists of four steps:

1. Decomposition of the factors that affect infovis visualizability into an AND-OR-EQUAL graph: two factors (also referred to as softgoals) are related by an AND relation if both the factors are important to satisfy the parent factor; two factors are related by an OR relation if either of the factors is important to satisfy the parent factor; a factor is related to another by EQUAL relation if the child factor is important to satisfy the parent. Generally, factors are named using the convention Type[Topic1, Topic2,…] where Type is the factor and Topic is the field of application of Type; Topic is optional.
2. Determine the various features for each tool – could be the graphical interfaces, graphing algorithms, multi-user ability, and the like; we refer to these as the infovis tool features.
3. Determine the extent to which the tool features affect the leaf visualizability factors identified in the first step by calculating the contributions the tool features make to the visualizability factors; the contributions can be one of four types: strongly positive or MAKE, positive or HELP, negative or HURT, and strongly negative or BREAK.
4. Capture the justifications for the contributions in step 3 so that a historical record of rationale for contribution changes is maintained.

It should be noted that the factor decomposition in step 1 may also assign priorities to some of the factors. Another point to note is that the contributions to the visualizability factors can either satisfice or deny that factor – satisficing is a term derived from economics and refers to relative satisfaction as opposed to absolute satisfaction. By using the satisficing attitude, the propagation rules permit collections of contributions to be given one level of satisficing. Once the SIG is computed, the propagation rules of the NFR Framework are applied to propagate the contributions in step 3 above to the top of the SIG. The propagation rules are (in the following TYPE is used to stand for one of MAKE, HELP, HURT, or BREAK):

R1. If all the contributions received by a leaf visualizability factor are TYPE then that leaf visualizability factor is considered TYPE-satisficed.
R2. If a leaf visualizability factor receives at least one HELP contribution then that leaf visualizability factor is HELP-satisficed.
R3. If a leaf visualizability factor receives at least one HURT contribution then that leaf visualizability factor is HURT-satisficed.
R4. If a leaf visualizability factor receives at least one BREAK contribution then that leaf visualizability factor is BREAK-satisficed.
R5. If R2, R3, and R4 apply, then the tie is broken in the order R4, R3, and then R2.
R6. If a leaf visualizability factor does not receive a contribution then it is considered MAKE-satisficed if it is involved in an AND or EQUAL relations, and BREAK-satisficed if it is involved in an OR relation.
R7. In the case of AND-related factors, if all child factors are TYPE-satisficed then the parent visualizability factor is TYPE-satisficed.
R8. In the case of AND-related factors, if even one of the child factors is TYPE-satisficed then the parent visualizability factor is TYPE-satisficed; the priority decreasing in the order: BREAK > HURT > HELP > MAKE.
R9. In the case of OR-related factors, if all child factors are TYPE-satisficed then the parent visualizability factor is TYPE-satisficed.
R10. In the case of OR-related factors, if even one of the child factors is TYPE-satisficed then the parent visualizability factor is TYPE-satisficed; the priority decreasing in the order: MAKE > HELP > HURT > BREAK
R11. In the case of EQUAL-related factors (only one child) the parent is TYPE-satisficed if the child is TYPE-satisficed.

In order to evaluate the candidate visualization tools the following actions are iteratively performed for each infovis tool:

1. Step 3 of SIG development – determines the extent to which the infovis tool satisfices the visualizability factors,
2. Step 4 of SIG development – captures the reasons for contributions in Step 3, and
3. Application of propagation rules R1 through R5 to evaluate visualizability of the infovis tool.

The result of this evaluation is that visualizability score for each tool will be qualitatively categorized into one of MAKE, HELP, HURT, or BREAK, with their ranking being:

MAKE > HELP > HURT > BREAK.

This determination of visualizability score will help us conclude the visualization capability of the tool as follows:

C1. If the root visualizability factor (the factor at the top of the SIG) is MAKE-satisficed, then that tool provides excellent visualization capability.

C2. If the root visualizability factor (the factor at the top of the SIG) is HELP-satisficed, then that tool provides good visualization capability.
C3. If the root visualizability factor (the factor at the top of the SIG) is HURT-satisficed, then that tool provides poor visualization capability.
C4. If the root visualizability factor (the factor at the top of the SIG) is BREAK-satisficed, then that tool is not good for visualization.

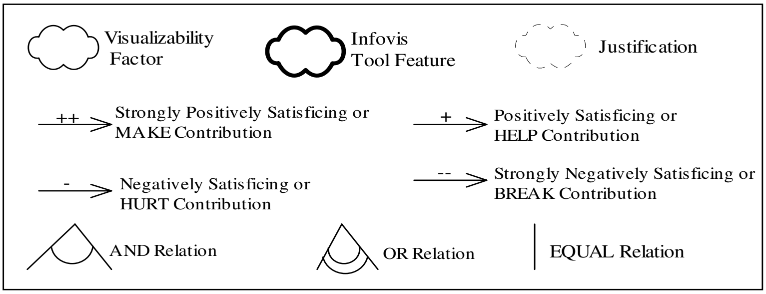


Fig. 2. Ontology for Softgoal Interdependency Graph

4 Application of NFR Approach for Evaluation of Info Vis Tools

SIG for evaluating infovis tools is given in Figure 3. The upper part of Figure 3 shows the decomposition of the visualizability softgoal. The softgoal Visualizability[InfoVis Tool] refers to the visualization ability of information visualization tools. Based on [1, 3], this root softgoal is decomposed into child softgoals Technique[Visualization], Utility[Features], Versatility[Tool], and Usability [Users]. Technique [Visualization] softgoal refers to the technique (or algorithm) used by the tool for visualization; Utility[Features] softgoal refers to the features that provide utility to the tool; Versatility[Tool] softgoal refers to the ability to use the tool for difference purposes; and Usability[Users] softgoal refers to the ease with which users could use the tool. That all these softgoals are necessary for the visualizability of an infovis tool is indicated by the AND-decomposition of the parent Visualizability[Infovis Tool] softgoal – the single arc stands for AND-decomposition. The softgoal Technique [Visualization] has been can be either static or customizable [2] – therefore, the softgoal has been EQUAL-decomposed (indicated by the line) into softgoal Customizability[Technique] – if this softgoal is strongly satisficed then the tool is customizable and if this softgoal is strongly not satisficed (or strongly denied) then the tool is static (non-customizable); static tools do not allow user selection of algorithms while dynamic technique tools do. The softgoal Utility[Features] is AND-decomposed (indicated by the single arc) into three child softgoals based on [3]: Availability[Graphics], Importability[Data] and Publishability[Results]. Availability [Graphics] softgoal refers to the ability to prepare different graphical views for large number of end-users; Importability[Data] softgoal refers to the ability to import data generated by other applications; and Publishability[Results] refers to the ability to export data to other applications. The AND-decomposition of the parent softgoal

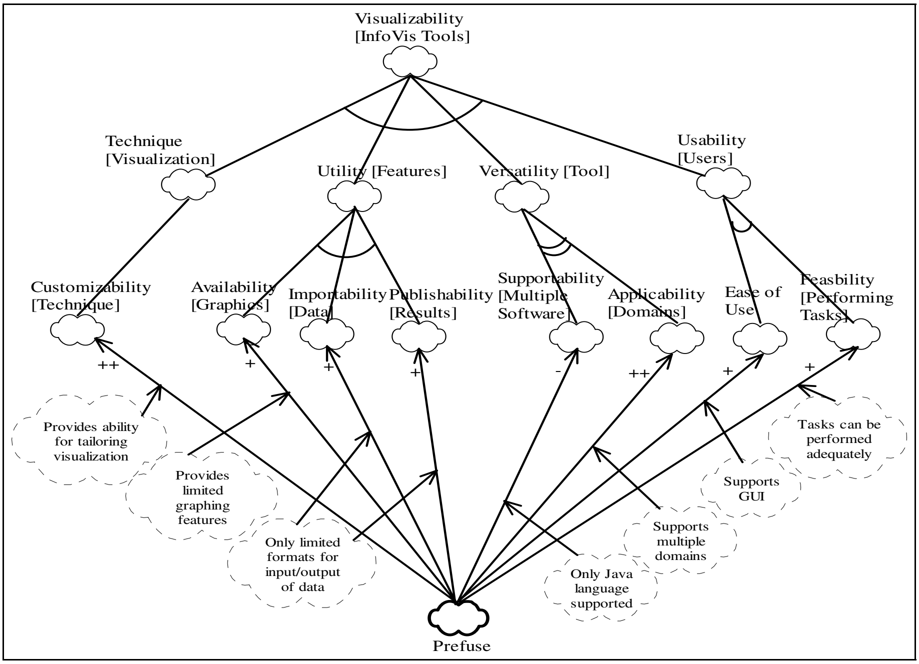


Fig. 3. SIG for Evaluating Visualizability of Prefuse

Utility[Features] means that all the child softgoals, namely, Availability[Graphics], Importability[Data] and Publishability [Results], should be satisficed for the parent to the satisficed. The softgoal Versatility[Tool] has been OR-decomposed into two child softgoals: Supportability [Multiple Software] and Applicability[Domains]. The softgoal Supportability[Multiple Software] refers to the ability for supporting multiple languages and applications, while the softgoal Applicability [Domains] refers to the ability to apply to different domains. The OR-decomposition of the child softgoals indicates that the parent (Versatility[Tool]) is satisficed if either or both of the child softgoals (Supportability[Multiple Software] and Applicability [Domains]) is satisficed. The softgoal Usability[Features] is AND-decomposed as indicated by the single arc into softgoals Ease of Use and Feasibility[Performing Tasks]. The softgoal Ease of Use refers to the ability of being able to easily use the tool, while the softgoal Feasibility[Performing Tasks] refers to the ability of being able to perform tasks expected of that application [3].

In Figure 3 the Prefuse visualization tool is being considered. The contributions that Prefuse makes to the various softgoals are represented by the arrows annotated with '++', '+', '-', or '--'. The reasons for the contributions are captured by the justification softgoals – thus, the contribution of Prefuse to the softgoal Customizability[Technique] is MAKE or '++' because "Prefuse provides ability for tailoring visualization". Likewise the justifications for other contributions are also shown in Figure 3. In order to evaluate visualizability of Prefuse the propagation rules R1 through R11 are now applied to the SIG of Figure 3 as illustrated by the following steps:

1. Application of rule R1:
   a. the leaf visualizability factor Customizability[Technique] is MAKE-satisficed.
   b. the leaf visualizability factor Availability[Graphics] is HELP-satisficed.
   c. the leaf visualizability factor Importability[Data] is HELP-satisficed.
   d. the leaf visualizability factor Publishability[Results] is HELP-satisficed.
   e. the leaf visualizability factor Supportability[Multiple Software] is HURT-satisficed.
   f. the leaf visualizability factor Applicability[Domains] is MAKE-satisficed.
   g. the leaf visualizability factor Ease of Use is HELP-satisficed.
   h. the leaf visualizability factor Feasibility[Performing Tasks] is HELP-satisficed.
2. By R11, the parent factor Technique[Visualization] is also MAKE-satisficed since its EQUAL-related child Customizability[Technique] is MAKE-satisficed.
3. By R7, the parent factor Utility[Features] is HELP-satisficed since all its children are HELP-satisficed.
4. By R10, the parent factor Versatility[Tool] is MAKE-satisficed.
5. By R7, the parent factor Usability[Users] is HELP-satisficed since both its children are HELP-satisficed.
6. By R8, since two child factors of the parent (root) factor Visualizability[Infovis Tool] are HELP-satisficed and the other two child factors are MAKE-satisficed, the root factor is also HELP-satisficed.
7. By C2, the infovis tool Prefuse is a good visualization tool.

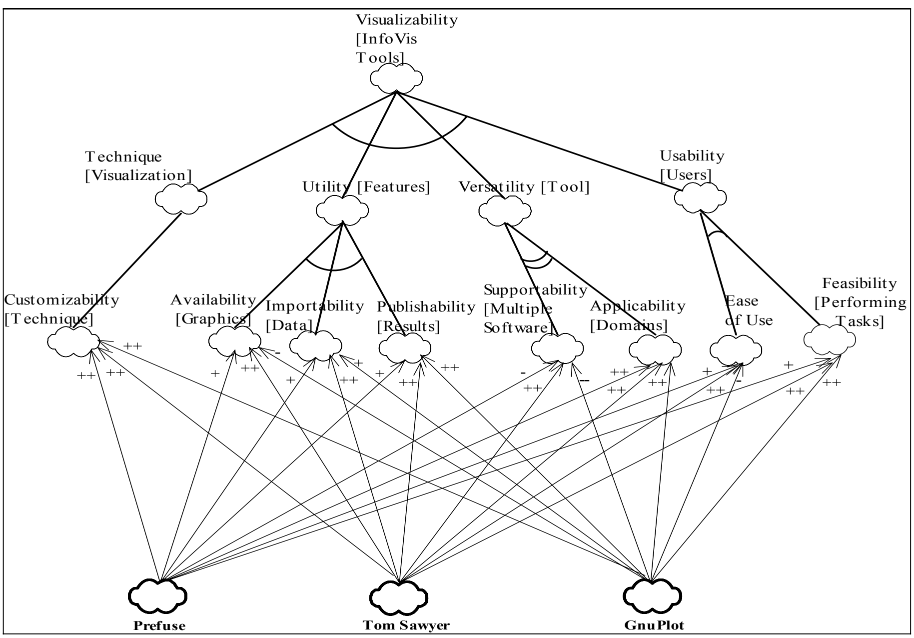


Fig. 4. SIG for Evaluating Visualizability of Three Infovis Tools

Table 1. Justifications for the Contributions in the SIG of Fig. 4

Visualizability Factor \ InfoVis Tools	Prefuse	Tom Sawyer	GnuPlot
Customizability [Technique]	MAKE - Provides ability for tailoring visualization	MAKE - Provides ability for tailoring visualization	MAKE - Provides ability for tailoring visualization
Availability [Graphics]	HELP - Provides limited graphing features	MAKE - Provides excellent graphing suite	HURT - Provides no graphing suite
Importability [Data]	HELP - Only limited formats for input of data	MAKE - Loads from various sources and formats	HELP - Only limited formats for input of data
Publishability [Results]	HELP - Only limited formats for output of data	MAKE - Provides varied formats for output of data	MAKE - Output to many types of file
Supportability [Multiple Software]	HURT - Only Java language supported	MAKE - Multiple languages support	BREAK - No language support
Applicability [Domains]	MAKE - Supports multiple domains	MAKE - Supports multiple domains	MAKE - Supports multiple domains
Ease of Use	HELP - Supports GUI	MAKE - Excellent GUI support	HURT - Command Line Interface
Feasibility [Performing Tasks]	HELP - Tasks can be performed adequately	MAKE - Tasks can be performed excellently	MAKE - Tasks can be performed adequately

In the SIG of Figure 4 we have considered the other two tools as well. The justifications for the contributions in Figure 4 are shown in Table 1 for easy reading. Application of the propagations rules to the SIG of Figure 4 is shown in Table 2.

5 Observations

We applied the NFR Approach to evaluate the visualizability characteristic for three different infovis tools, namely, Prefuse, Tom Sawyer, and GnuPlot. Based on the qualitative score for visualizability, we could rank in a straightforward manner the visualization ability of these tools. The advantages of the NFR Approach for evaluation are discussed below.

5.1 No Fixed Definition for Visualizability

The upper part of the SIG captures the decomposition of the visualizability aspect and this decomposition also captures the definition of the NFR visualizability. Therefore, if the definition had to be changed, it would only require a different decomposition. The overall process of evaluation however will remain the same which provides the

Table 2. Application of Propagation Rules to the SIG of Fig. 4

Visualizability Factor \ InfoVis Tools	**Prefuse**	**Tom Sawyer**	**GnuPlot**
Customizability [Technique]	MAKE (from user)	MAKE (from user)	MAKE (from user)
Availability [Graphics]	HELP (from user)	MAKE (from user)	HURT (from user)
Importability [Data]	HELP (from user)	MAKE (from user)	HELP (from user)
Publishability [Results]	HELP (from user)	MAKE (from user)	MAKE (from user)
Supportability [Multiple Software]	HURT (from user)	MAKE (from user)	BREAK (from user)
Applicability [Domains]	MAKE (from user)	MAKE (from user)	MAKE (from user)
Ease of Use	HELP (from user)	MAKE (from user)	HURT (from user)
Feasibility [Performing Tasks]	HELP (from user)	MAKE (from user)	MAKE (from user)
Technique [Visualization]	By R11, MAKE	By R11, MAKE	By R11, MAKE
Utility [Features]	By R7, HELP	By R8, MAKE	By R8, HURT
Versatility [Tool]	By R10, MAKE	By R9, MAKE	By R10, MAKE
Usability [Users]	By R7, HELP	By R7, MAKE	By R8, HURT
Visualizability [InfoVis Tools]	By R8, HELP	By R8, HELP	By R8, HURT
Final Evaluation	**By C2, good visualization capability**	**By C1, excellent visualization capability**	**By C3, poor visualization capability**

advantage of being flexible to accommodate almost any definition of visualizability. For example if we were to include "cost" as an additional factor in evaluation of visualizability, then open source products such as Prefuse gain an advantage over licensed products like Tom Sawyer and therefore Prefuse may score higher than Tom Sawyer.

5.2 Capture of Justifications

One of the neat aspects of the NFR Approach is that justifications for every contribution are captured as shown in Table 1. This permits the evaluator to record her reasons whenever she determines the type of contribution a tool feature makes to a visualizability factor. If the reasons were to change, a modification in the type of contribution may be called for. Therefore, the determination of contribution types is led by their justifications.

5.3 Reasons for Evaluations Captured by SIG

Evaluations are based on the contributions received by the leaf visualizability factors and progress upward in the SIG based on the propagation rules. Therefore, the qualitative score for each factor in the SIG can be justified. Since the scores start with the contributions which are themselves justified, every score has a firm basis. For example, if a factor were to be MAKE-satisficed we know exactly why it is so and it can be traced back to the user's reasons for the contributions.

5.4 Can Help Improve Design of Infovis Tools

As shown in Table 2, the GnuPlot receives a lower than average score because of its command-line user interface. If the visualizability score for GnuPlot tool needed to be improved, one way to achieve this would be to design a more interactive user interface. Therefore, the NFR Approach helps determine how to improve infovis tools by identifying components in each tool that actually helps improve or worsen a visualizability factor.

5.5 Historical Record Maintained in the Form of SIGs

With the passage of time, new factors and tool features may need to be included. However, each change is recorded on the SIG together with its justifications. Therefore, SIGs help maintain historical records of changes. If any change were to result in an improved or worsened score it would be easy to locate the reasons for the new score from the historical records.

5.6 Automatable

In Table 2 we show how the qualitative scores were evaluated manually. However, it is relatively easy to automate this process by using, for example a spreadsheet program such as MS-Excel. It is possible to programmatically apply the propagation rules once the contributions are provided [12]. An example of such an implementation is provided in [14].

6 Conclusion

Information visualization (infovis) tools are extensively used to gain cognitive knowledge of huge repositories of information in industry and academia. For example, Tom Sawyer is an infovis tool utilized by several companies for viewing business processes and by universities for educational purposes. It will therefore be of considerable interest to practitioners to choose an appropriate infovis tool for their requirements. In this paper we present a novel approach to evaluate the quality of visualization of infovis tools by qualitatively computing their visualizability. The NFR Approach employed by this paper considers visualizability as a non-functional requirement (NFR) that has to be achieved by an infovis tool. The NFR Approach develops a softgoal interdependency graph (SIG) to evaluate qualitatively the visualizability of an infovis tool. Based on this evaluation, the visualization ability of the tool can be directly inferred. This paper validates this approach by applying it to evaluate the visualizability of three infovis tools, namely, Prefuse, Tom Sawyer, and GnuPlot. The chief advantages of this approach include ability to accommodate different definitions of visualizability, ability to rationalize the scores, historical maintenance of records, and ability to be automated.

For the future, we plan to develop software to automatically compute visualizability for different infovis tools so that practitioners may quickly evaluate tools at hand. Yet another line of research would be to develop catalogs of definitions for visualizability so that practitioners may choose the most appropriate definition for their use. A third line of potential research would be to prioritize different factors and study their impact on the final evaluation. Yet another possibility is to convert qualitative scores into quantitative metrics. However, we believe that the NFR Approach provides a systematic method to qualitatively evaluate visualizability scores for infovis tools.

References

1. Montgomery, D.C.: Introduction to Statistical Quality Control. John Wiley & Sons, Canada (2001)
2. Freitas, C.M.D.S., Luzzardi, P.R.G., Cava, R.A., Winckler, M.A.A., Pimenta, M.S., Nedel, L.P.: Evaluating Usability of Information Visualization Techniques. In: IHC 2002 - 5th Workshop on Human Factors in Computer Systems. Brazilian Computer Society Press (2002)
3. Heo, M., Hirtle, S.C.: An Empirical Comparison of Visualization Tools to Assist Information Retrieval on the Web. J. American Society for Information Science and Technology 52, 666–675 (2001)
4. Prefuse Information Visualization Toolkit, `http://prefuse.org/doc/manual/introduction`
5. Heer, J., Card, S.K., Landay, J.A.: Prefuse: a toolkit for interactive information visualization. In: Proceedings of ACM Conference on Human Factors in Computing Systems, Portland, Oregon, pp. 421–430 (2005)
6. Kumar, P., Zhang, K.: Social Network Analysis of Online Marketplaces. In: Proceedings IEEE International Conference on E-Business Engineering, Hong Kong, China, pp. 363–367 (2007)
7. Plaisant, C.: The Challenge of Information Visualization Evaluation. In: Proceedings of the 7th International Working Conference on Advanced Visual Interfaces, Italy, pp. 109–116 (2004)
8. Tom Sawyer Software Industries, `http://www.tomsawyer.com/industries/`
9. Tom Sawyer Software Products, `http://www.tomsawyer.com/products/index.php`
10. GnuPlot Homepage, `http://www.gnuplot.info/`
11. Chung, L., Nixon, B.A., Yu, E., Mylopoulos, J.: Non-Functional Requirements in Software Engineering. Kluwer Academic Publishers, Boston (2000)
12. Subramanian, N., Whitson, G.: Augmented WebHelix: Practical Process for Web Engineering. In: Software Engineering for Modern Web Applications: Methodologies and Technologies. IDEA Publishers (to appear)
13. Chung, L., Subramanian, N.: Process-Oriented Metrics for Software Architecture Adaptability. In: Proceedings of the International Symposium on Requirements Engineering, pp. 310–311. IEEE Computer Press, Los Alamitos (2001)
14. Subramanian, N.: Visualizing Safety Score for the Digital Home. In: Proceedings of the IEEE International Conference on Software Engineering Theory and Practice, pp. 128–134. IEEE Computer Press, Orlando (to appear, 2008)

Interactive Exploration of Large Dynamic Networks

Mathias Pohl[1] and Peter Birke[2]

[1] Dept. for Software Engineering, University of Trier, Germany
[2] Dept. for Databases and Information Systems, University of Trier, Germany
{pohlm,birke}@uni-trier.de

Abstract. A variety of applications especially in the area of Web 2.0 produce frequently altering hierarchical networks. Thus application operators, members of community websites but also media scientists often are interested in gaining deeper insights in the complex structures of their project and its development over time. In order to enable a suitable visualization of such networks it is not only required to implement an intelligent data management but also a suitable network drawing engine that cares for the user's needs – especially for visualization of network dynamics. This paper presents the XLDN visualization tool that enables visual exploration of evolving hierarchical networks. It includes a layout generation algorithm that allows for the preservation of the mental map – a crucial property when visualizing dynamic networks. By utilization of an efficient data management and a parallel implementation of the graph drawing algorithm XLDN provides a reasonably fast and interactive network visualization. This way XLDN facilitates an in-depth evolution analysis of large hierarchic networks.

1 Introduction

Many datasets especially those created in the recent years contain large hierarchical structures. One example for such structures are virtual social networks established by applications such as Wikipedia, FaceBook, or XING. The larger the network the harder to analyse and to visualize it. Furthermore, one is also interested in the evolution of a network. Thus, a network visualization tool is required that enables an easy and fast exploration of large dynamic networks. Therefore, we introduce the XLDN tool (*eXplorer for Large Dynamic Networks*).

Hereby, the term dynamic refers to the fact that XLDN is able to show structural changes (i.e. adding and removing nodes and edges) in the networks as well as changes in their hierarchies. This is a difference to other dynamic network visualization tools. These tools usually allow structural changes only.

When creating XLDN we faced several problems. Our first challenge was to decide for a suitable visualization metaphor. Matrix views allow the visualization of huge datasets at once but make it hard to find interesting structures while node-link diagrams tend to produce visual clutter. However, by using an interactive exploration environment (as presented by Archambault et al. in

M. Sebillo, G. Vitiello, and G. Schaefer (Eds.): VISUAL 2008, LNCS 5188, pp. 56–67, 2008.

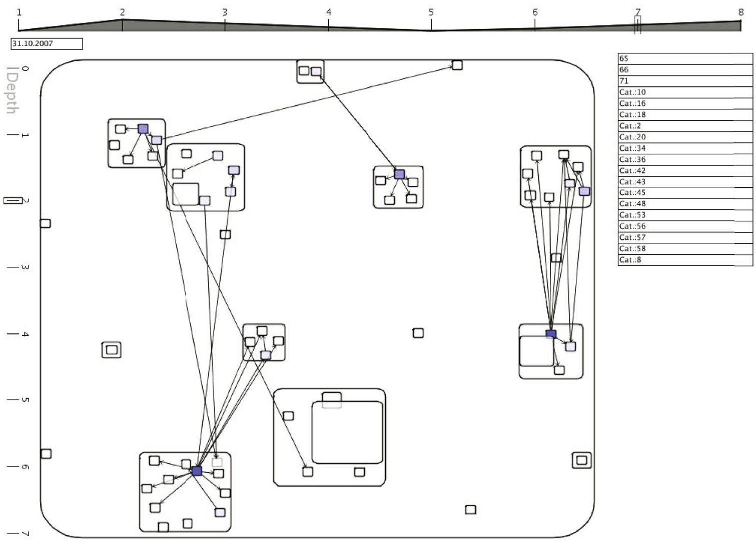


Fig. 1. The main view of XLDN

GROUSE [1] and Storey et al. in SHriMP [2]) the disadvantages of node-link diagrams can be reduced. Due to this consideration we decided to use the node-link metaphor.

In general, a dynamic network is considered to be a sequence of static networks (so-called *snapshots*). Hence, layout algorithms for dynamic networks should be able to deal with such sequences in a user-friendly way – a problem that is also known as the *dynamic graph drawing problem*. This means that the used layout algorithm should generate layouts that are readable but that do not bother the user with *unnecessary* layout changes. As XLDN is intended to be a post-mortem visualization tool it can make use from the fact that the complete data is available in advance. This allows more promising results since the used network visualization algorithms can anticipate the dynamic network behaviour and does not have to perform their computations in real time. In order to handle large networks and to offer fast interactions on the precomputed visualizations XLDN needs an efficient data management system.

To illustrate the usefulness of XLDN we applied it to an easy to extract network structure from a local installation of the MediaWiki software. The network is built from the articles contained in the wiki and the (directed) links between them. The category system of the software inherits a suitable hierarchy structure for the link network. Using the link- and category structure XLDN enables the analysis of the arrangement of articles in categories and the density in certain areas of interest. The benchmark results presented in section 3 are based on dumps of separate language versions of Wikipedia.

2 Overview: The Three Stages of XLDN

The main goal of XLDN is to provide an interactive network exploration system for large dynamic networks. To achieve this goal XLDN works in three consecutive stages where the third stage actually creates the visualization view. At the beginning, the network information is reshaped and reorganized in a backend database. In the second stage a relative network layout is precomputed. This relative layout then serves as a basis for the visualization stage. Due to this division into separate phases XLDN is able to offer a fast interactive control to the visualization as only a few computational issues have to be performed in this stage.

2.1 Stage 1: Loading Data into the Database Backend

Even modern computer hardware does not always offer the possibility to store the network data fully in RAM. To get rid of this problem XLDN uses a *PostgreSQL 8.2* database that consists of five tables (network structures, available nodes, available edges, the computed layout of nodes and information about the edges). Although storing this information in a database is slightly slower than working in memory this approach prevents the operating system from swapping to virtual memory and hence keeps the whole system alive. This prerequisite is crucial for a system that is supposed to be interactive.

2.2 Stage 2: Performing Graph Layout

After data processing is finished XLDN computes a relative network layout that is stored in the database as well. The layout is generated by a parallel offline-dynamic graph layout algorithm. Offline-dynamic in this context means that the layout algorithm requires the knowledge of a complete network sequence in order to work correctly. As all data is completely preprocessed this is no limitation for XLDN. Due to the parallel layout computation this step is reasonably fast on multicore processors available today.

2.3 Stage 3: Interactive Visualization

In the last phase XLDN presents an interactive visualization view to the user. As all necessary data is precomputed and stored in a suitably shaped database this visualization view reacts sufficiently fast to any user request. Furthermore, the database approach allows stage 3 to be executed independently. Thus, it is possible to continue exploration and analysis without rerunning stages 1 and 2.

3 Generating Dynamic Layouts

Dynamic networks, i.e. sequences of static networks, do not only require layouts with a high aesthetic quality but also with a good layout stability. An animation between two consecutive networks should not bother users with too many layout changes to simplify the navigation through the network sequence. Therefore, dynamic layout algorithms have to strike a balance between aesthetic quality and global stability. The latter is also known as the *preservation of the mental map* which is one of the key issues of dynamic layout generators [3].

In case of a dynamic hierarchical network, the layouts additionally have to accentuate variations of the hierarchy over time since these variations generally indicate a primarily insufficient and revised structure. An animation between two consecutive networks should enable users to recognize these structural changes of the hierarchy easily. Furthermore, hierarchically structured networks also raise the navigation through the sequence because of focusing on partial information.

The layout algorithm of XLDN takes both aspects, the network's dynamic and hierarchy, into account while computing layouts. The mental map will be generally preserved unless XLDN advises the user of hierarchical changes. Focusing on selected hierarchies will also be permitted by the visualization.

In contrast to so-called online dynamic algorithms as the one by Frishman and Tal [4] the architecture of XLDN allows the use of offline-dynamic algorithms, i.e. the complete sequence is known prior to the layout process.

3.1 FLT

Foresighted Graph Layout with Tolerance (FLT) has originally been presented by Diehl et al. [5]. It is an offline-dynamic graph drawing method that allows for any classical graph drawing algorithm to be applied to a sequence of graphs such that the user's mental map [6] can be preserved.

The main idea behind FLT is to compute a global layout template for the sequence from the layout of the so-called supergraph. Starting from the layout of this template FLT recommended to optimize all layouts within tolerance limits (hence the name). However, it was still possible that FLT produced layouts of poor quality. In particular several graphdrawing paradigms like layered or orthogonal graph drawing led to a severe loss of layout quality. The usage of revised versions of graph drawing algorithms in combination with an improved version of FLT [7] was required. This improvement has also been subject of an evaluative user study that proved the usefulness of the computed layouts [8].

Unfortunately, FLT is limited to sequences of networks that do not contain any hierarchical information. Its key concept – the supergraph – is not suitable for this kind of networks.

3.2 FLT and Hierarchy: The Supertree

A dynamic hierarchical network is an ordered sequence of clustered networks $H_0, \dots H_{n-1}$. A clustered network H is an ordered pair (G, T) of an undirected graph $G = (V_G, E_G)$ and a rooted tree $T = (V_T, E_T)$ such that $V_G = leaves(T)$.

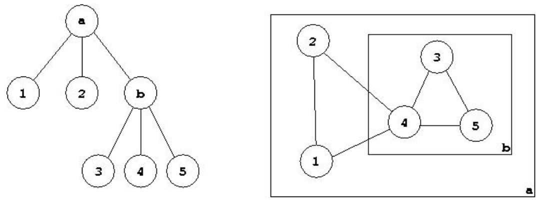


Fig. 2. A clustered network: The hierarchy tree (left) induces the clusters in the graph (right)

In other words, the hierarchy is given by the rooted tree T whereas the nodes and edges of the network are represented in the graph G.

In order to deal with dynamic hierarchical networks in FLT we developed the concept of the supertree. This concept is similar to that of the supergraph. It can be regarded as a global hierarchy that contains the hierarchical structure of the sequence and allows to compute a global template for the layouts of every network in the sequence. More formally, for a given sequence $T_0, \ldots, T_{n-1}$ of hierarchies with $T_i = (V_i, E_i)$ the *supertree* $\widetilde{T} = \left(\widetilde{V}, \widetilde{E}\right)$ is a tree with the following properties:

1. $\forall\ 0 \leq i < n:\ \exists\ \sigma_i : V_i \rightarrow \widetilde{V}$ injective
2. $\forall\ 0 \leq i < n:\ (u, v) \in E_i \Rightarrow (\sigma_i(u), \sigma_i(v)) \in \widetilde{E}$

A *minimal supertree* for a sequence $T_0, \ldots, T_{n-1}$ is a supertree with the property that $\widetilde{V}$ has minimal size.

Due to the first property for every node in an arbitrary network of the sequence exists a unique representing node object in the supertree. Thus, the global hierarchical structure of the whole dynamic network is encoded in the supertree. Therefore a layout of the supertree in simple terms provides a sort of look-up table for an initial positioning of the sequence's nodes. We simply adjust these initial layouts iteratively by using a classical hierarchical graph drawing algorithm and we achieve the final layout of the dynamic network. Hence, the layout algorithm preserves the mental map because the starting point for the adjustment is defined by the supertree layout.

The main drawback of this approach is the number of representing node objects in the supertree. A huge number of representatives induces a layout which uses a large area. If a node is contained in every network of the sequence we will create, in worst case, n representatives. But in general it is possible to represent different nodes by just one common node object in the supertree. Generally speaking there are two different premises which indicate a common representative.

The first premise deals with the hierarchy. It is possible that a certain node n is contained in different clusters over time. In this case, the paths from this given node n to the corresponding root nodes in the unequal networks of the sequence differ. Varying arrangements in the hierarchy of the dynamic network are indicated. More formally, let $T_0, \ldots, T_{n-1}$ with $T_i = (V_i, E_i)$ be a sequence of trees and $r_0, \ldots, r_{n-1}$ with $r_i \in V_i$ the corresponding root nodes. Then a node $v \in V_i \cap V_j$ is called *hierarchically stable* in T_i and T_j with $0 \leq i, j < n$ if $path(v, r_i) = path(v, r_j)$ with $path(v, r_i) \subseteq V_i$ and $path(v, r_j) \subseteq V_j$.

To decide whether a given node v is hierarchically stable or not is easy. Nevertheless, the number of supertree nodes will be reduced due to the fact that we create only representing node objects for every hierarchy of v and not for every network which contains v. In addition, the layout algorithm accentuate variations of the hierarchy over time due to the different representatives.

The second premise concerns the dynamics of the network. As mentioned above, a node is potentially contained in several different networks of the sequence. We therefore associate with every node v a set of indices from the set $\{0, \ldots, n-1\}$ which indicates the v containing networks of the sequence. We call these sets lifetimes. The layouts of two nodes with disjoint lifetimes do not interact in any network of the sequence. That is why this independence generally enables the representation by a single node object in the supertree. In addition, we especially take into account the hierarchical structure of the nodes in the network sequence. A single supertree node for two different nodes with disjoint lifetimes and different hierarchies results in an inconsistent supertree which potentially violates the properties of a tree. In order to guarantee that the supertree is a tree again, we have to deal with the hierarchy and the dynamics of the network at the same time. Therefore, the second premise can only be used for hierarchically stable nodes.

The necessary combination of the specified premises allows to compute a supertree with a small number of nodes. Unfortunately, it is not minimal because the computation of a minimal set of nodes with disjoint lifetimes is NP-hard [5]. The algorithm which computes a supertree works in two phases. The result of the first phase guarantees the generation of hierarchically stable node objects in the intermediate supertree. In the second phase the supertree nodes with disjoint lifetimes were merged so that they were represented by a single common node object.

More precisely during the first phase we create an empty supertree by generating just a new root node object r. The node r represents all the roots in the network sequence because of the disjoint lifetimes and the fact that they are obviously hierarchically stable. After that we iteratively add new supertree nodes for the children of the root nodes of the networks in the sequence. If there already exists a representative for a child, we simply find a hierarchically stable node and do not create an additional supertree node. As a side-effect the algorithm additionally computes the lifetimes of the supertree nodes. A new node object in the supertree will either be created and the lifetime is simply the position of the network in the sequence, or a hierarchically stable node was found and the lifetime will be updated by simply adding the networks position.

During the second phase, the algorithm minimizes the number of children by merging them with disjoint lifetimes together. Because this problem as mentioned above is NP–hard we use a greedy approach which merges pairwise the nodes with disjoint lifetimes. By doing so, the algorithm computes a small number of representing children in the supertree.

The children in the supertree are kept in a queue and the algorithm proceeds by recursively inserting the next node levels. By doing so we compute top-down in BFS-style a supertree with a small number of nodes.

3.3 Layout Generation in Parallel

To make the most of the hierarchical information in the processed networks we implemented a parallel layout algorithm to work with extended FLT. On today's multicore processors an algorithm working in several threads needs reasonably less computation time. Our algorithm is based on the NUAGE-algorithm [9] by Bertault and Miller that works as follows: The set of tree nodes is partitioned into clusters prior to the layout process. The mapping between the nodes and clusters is induced from the network's hierarchy tree; all nodes sharing the same parent in the tree are grouped together. After partitioning the tree nodes all edges between nodes of different clusters are modified such that there are only edges between two nodes of the same cluster. The modification is done by replacing an edge $e = (v, w)$ with the edge $e = (v', w')$ such that v' and w' are two ancestors of v and w that have the same parent in the tree. This replacement results in a set of independent clusters still containing the full network structure. NUAGE then layouts the clusters recursively according to the hierarchy tree and computes a final layout by putting together the relative layouts.

In contrast to NUAGE the algorithm in XLDN layouts the clusters in parallel. Each cluster is processed independently to obtain a relative layout. This approach decreases the overall running time of the layout algorithm (see Table 1. The relative layout for each cluster is obtained by applying a standard layout generation algorithm – XLDN uses an implementation of the force-directed placement method by Fruchterman and Reingold [10]. Due to the independence among each cluster this layout generation can be done in separate threads. Furthermore, XLDN performs the final generation of the absolute layout information during the visualization phase.

Table 1. Average time for layout computation using the sequential and the parallel approach. The examples were derived from the article-link structure of the different language versions of Wikipedia. Each dynamic network consists of eight slices. Hence, the overall running time refers to the time for the complete sequence of networks.

Example	Avg. num. nodes	Avg. num. edges	Running time (ms) 1 Thread	32 Threads
`pdc.wikipedia.org`	972	1701	12101	4880
`zh-classical.wikipedia.org`	1257	3009	37954	32368
`fr.wikiquote.org`	287	100	1414	723

4 The Final Step: Visual Exploration

4.1 The Visualization View

The main view of XLDN (see Fig. 1) consists of five components:

- The node-link diagram shows the currently selected part of the network. Depending on the selected depth it displays nested graphs within their respective cluster. In this view the user can navigate through the network's hierarchy tree (see Sect. 4.3) by activating a selected cluster. Furthermore, the user can fold and unfold nodes and use graphical pan and zoom to have a finer control of what is being drawn. Nodes in the visualization are not annotated by labels unless the mouse pointer moves upon a node. To enable the connection of nodes to scalar data they can be displayed in different colors to encode further information. In our example we encoded the outdegree centrality such that high centrality values lead to more bluish nodes.
- The label box on the right side displays the names of all nodes in the uppermost visible hierarchy level. To increase the readability of long lists this box displays the labels in alphabetical order and provides a fisheye effect.
- The time-slider on the upper side of the window allows for navigation through the network sequence. In advance to the standard slider concept this slider is also able to display a function curve representing some time-series data. For the examples in this paper the slider displays the overall number of changes, i.e. at position i the number of differences between network $i-1$ and network i is displayed. Depending on the intended task of XLDN it is possible to use different values in the slider.
- Labels below the time-slider show the name of the network (which is likely to be a timestamp) and the list of nodes that show the path from the network's hierarchy root to the currently visible network.
- The depth-slider on the left side enables the user to control the maximal visible depth. This feature is discussed in the following section.

According to Shneiderman's *Visualization Mantra* [11], XLDN does not immediately present all available information but it shows just an overview. Afterwards, the user can navigate through the hierarchy to obtain more detailed information.

4.2 Overview first: Fixed Visible Depth

The first network of the sequence is also the first network being displayed by XLDN. It is shown with a fold level of one, i.e. all visible nodes have at most a distance of one to the root node with respect to the hierarchy tree. Edges of the network only appear if the connected nodes are both visible. In order to increase or decrease the number of visible objects the fixed depth can be switched to a different value using the depth slider. This causes the visualization to perform an animation towards more or less visible objects (see Figure 3).

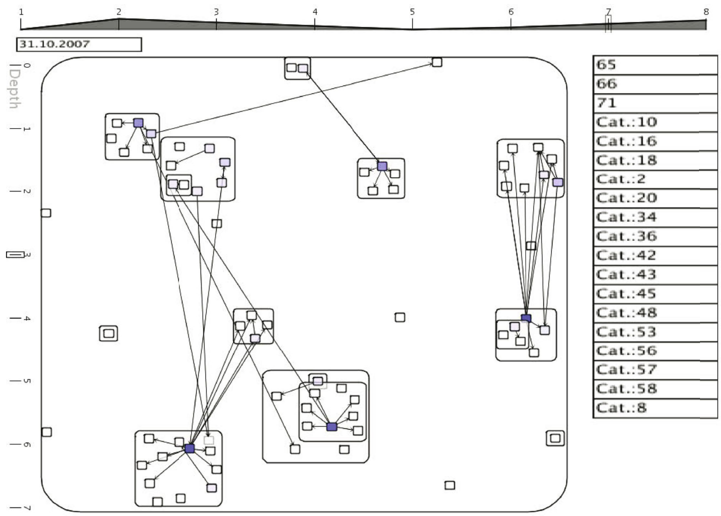


Fig. 3. The network of Figure 1 with a depth of 3.

In addition to the depth slider control it is also possible to control the number of visible objects in more detail. Nodes can be folded which results in disappearing of all nodes beneath folded nodes. The folding operation is triggered by a right click with the mouse on a node. Unfolding a node is triggered by another click.

4.3 Details on Demand: Semantic Zoom

The initial visualization only offers some levels of the hierarchy tree. To descent in a subtree of the hierarchy its representative cluster node can be made the new visible root. Clicking on a node triggers an animation that updates the visualization to a new state (see Figure 4). While nodes and edges in different hierarchy subtrees disappear the selected node is enlarged. The fixed visible depth still remains the same and hence nodes beneath the selected node may become visible. Furthermore, the hierarchy navigator is updated: The selected node's label becomes visible between the animation slider and the network view to indicate that the visible network is the subnetwork beneath the selected node.

4.4 Dynamics: Animation

The main difference of XLDN in advance of different tools is its support for network sequences. Since the layout algorithm already provides suitable layout information it is sufficient to have a linear interpolation between two different networks of the sequence. The animation is triggered by the time-slider. During the animation XLDN is aware of the currently visualization state and does only

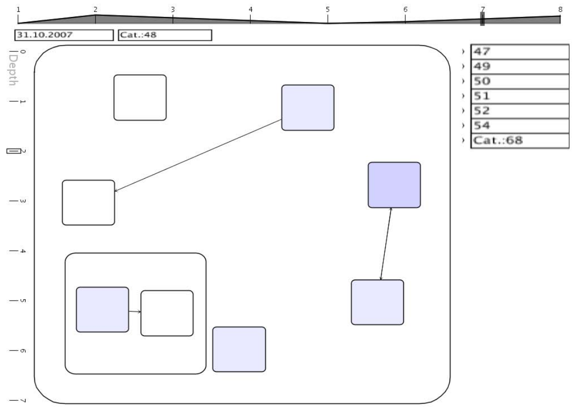
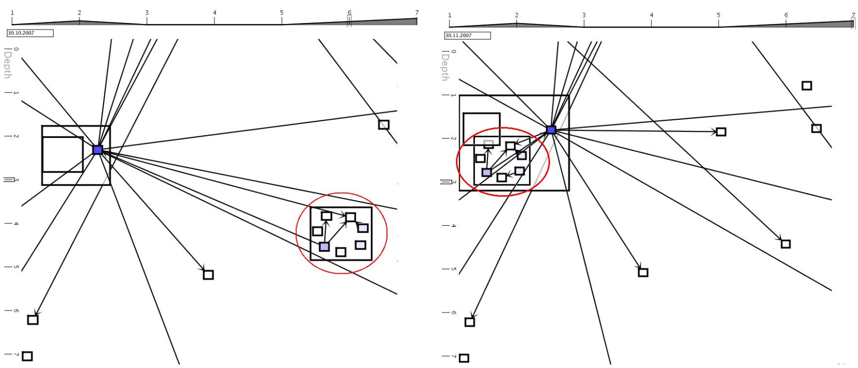


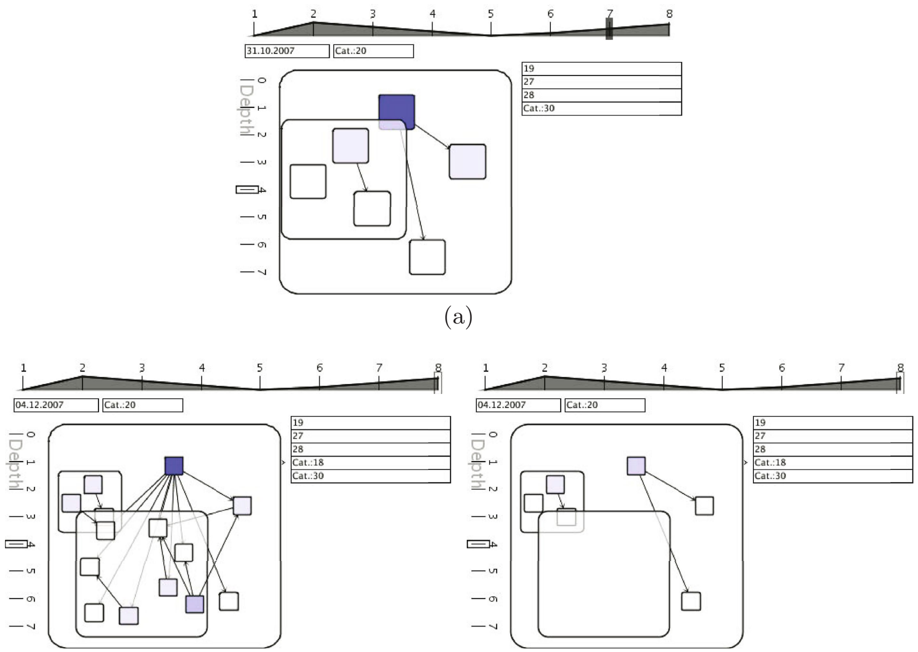
Fig. 4. Semantic Zoom: Based from the network in Figure 1 node "Cat.: 48" has been selected. The transition is performed in a smooth animation and the layout within the cluster is left unchanged.



(a) (b)

Fig. 5. Two consecutive time slices from the example network. The former toplevel cluster originated on the right has moved its hierarchy position and is a subcluster of the large cluster on the left. The motion of the cluster during animation emphasizes this change in hierarchy.

show changes between the subnetworks induced by the currently visible root. If the currently examined subnetwork is not part of the network at the new sequence index this will be indicated by disappearance of all visible objects. Then the user has to navigate back to an available position in the hierarchy tree. An animation between two different network indices does not change the folding state of nodes as well. This way it is possible to quickly navigate through the time without bothering the user with uninteresting information. This property is illustrated in Figure 6.



(a)

(b) Successor to (a) without folding

(c) Successor to (a) with folding

Fig. 6. Comparison of time-navigation with and without folding of nodes. The absence of structural changes between the nodes in (a) compared to (b) is hardly visible.

4.5 Labelling

Labelling is a very important aspect in network visualization. Since XLDN is designed to work with large networks it would not be reasonable to place labels to all nodes. Therefore we used the tool-tip labelling approach where nodes are only highlighted if the mouse points to them. Highlighted nodes are annotated by a label and are slightly enlarged. In addition to directly working on the diagram the label box on the right side is also connected to this mechanism; hovering an element in the label box highlights it in the visualization. Finally due to the implemented fisheye effect the label box also works for many labels.

5 Conclusion

This paper presented XLDN a visualization tool for large dynamic networks. It provides the visual exploration through the evolution of hierarchical structures. XLDN's usefulness is based on an efficient database backend and a fast implementation of FLT– a layout algorithm that is aware of quality aspects concerning dynamic layouts, i.e. preservation of the user's mental map. To make FLT work with dynamic hierarchical networks it has been extended with the concept of the supertree. All in all, XLDN provides a user-friendly network visualization environment that allows an interactive exploration of large hierarchical structures. With its special design it supports networks with several thousands of nodes.

Acknowledgements

Mathias Pohl is partially supported by Cluster of Excellence "Social Networks" at the University of Trier.

References

1. Archambault, D., Munzner, T., Auber, D.: Grouse: Feature-based, steerable graph hierarchy exploration. In: EuroVis, pp. 67–74. Eurographics Association (2007)
2. Storey, M.A.D., Best, C., Michaud, J.: SHriMP Views: An Interactive Environment for Exploring Java Programs. In: Proc. of 9th Int. Workshop on Program Comprehension, IWPC, pp. 111–112. IEEE Computer Society Press, Los Alamitos (2001)
3. Purchase, H.C.: Which aesthetic has the greatest effect on human understanding? In: DiBattista, G. (ed.) GD 1997. LNCS, vol. 1353, pp. 248–261. Springer, Heidelberg (1997)
4. Frishman, Y., Tal, A.: Online dynamic graph drawing. In: EuroVis, pp. 75–82. Eurographics Association (2007)
5. Diehl, S., Görg, C., Kerren, A.: Preserving the Mental Map using Foresighted Layout. In: Proc. of Joint Eurographics - IEEE TCVG Symp. on Visualization, VisSym (2001)
6. Bridgeman, S.S., Tamassia, R.: Difference Metrics for Interactive Orthogonal Graph Drawing Algorithms. In: Whitesides, S.H. (ed.) GD 1998. LNCS, vol. 1547, pp. 57–71. Springer, Heidelberg (1999)
7. Görg, C., Pohl, M., Birke, P., Diehl, S.: Dynamic Graph Drawing of Sequences of Orthogonal and Hierarchical Graphs. In: Pach, J. (ed.) GD 2004. LNCS, vol. 3383, pp. 228–238. Springer, Heidelberg (2005)
8. Purchase, H.C., Hoggan, E., Görg, C.: How Important is the Mental Map. In: Kaufmann, M., Wagner, D. (eds.) GD 2006. LNCS, vol. 4372. Springer, Heidelberg (2007)
9. Bertault, F., Miller, M.: An Algorithm for Drawing Compound Graphs. In: Kratochvíl, J. (ed.) GD 1999. LNCS, vol. 1731, pp. 197–204. Springer, Heidelberg (1999)
10. Fruchterman, T.M.J., Reingold, E.M.: Graph Drawing by Force-directed Placement. Softw., Pract. Exper. 21(11), 1129–1164 (1991)
11. Shneiderman, B.: The Eyes Have It: A Task by Data Type Taxonomy for Information Visualization. In: Proc. of the IEEE Symp. on Vis. Languages, pp. 336–343 (1996)

Flood Emergency Interaction and Visualization System

Rui Nóbrega, André Sabino, Armanda Rodrigues, and Nuno Correia

IMG (Interactive Multimedia Group), DI and CITI/FCT, New University of Lisbon
{rui.nobrega,amgs,arodrigues,nmc}@di.fct.unl.pt

Abstract. In this paper we describe a visualization system for an emergency simulation. We start by presenting a flooding emergency case scenario and all the elements that are involved in it. Then we describe the design decisions that were made in order to create a credible representation of the scene. This includes using a game engine to render the scenario using a three-dimensional terrain, with objects and information retrieved from geographic information systems. Additionally we describe experiments with new touch and tangible devices that support a war table like interaction with the simulation. We further describe our ideas for the emergency interface and conclude describing the directions for future work.

Keywords: Emergency, simulation, interface, interaction.

1 Introduction

Visualization and interaction are key aspects in Emergency Management. Timely decisions depend on having the right tools, available in a way easily perceived by the users. In this paper we describe the user interface of the LIFE-SAVER project, and discuss our choices in the scope of the ongoing work.

Project LIFE-SAVER was created in the Portuguese context of Dam Break Emergency Management [1]. In this country, the management of flooding emergencies, resulting from problems related with Dams, is planned through the guidance of existing special emergency plans. This project aims to develop a system that can effectively validate these plans, through simulation of the flooding natural phenomena and of all the actors intervening in the situation. Another important aspect is the definition of the plan in the simulation in a visual environment.

Creating a visual interface for the simulation presents several challenges. It is important that the users interact with a visually rich environment that allows a fast perception of the emergency that is being simulated. It should also be easy to use and be targeted for people with no special skills in computer science.

Having this in mind we created a prototype application that uses a graphical game engine to render a scenario where the user can observe and interact with natural and manmade disasters. Additionally the user can edit and define the agent behavior and test several scenarios.

LIFE-SAVER includes the development of a prototype of the Validation System, tailored to work with the Alqueva Dam Emergency Plan, which also involves another

M. Sebillo, G. Vitiello, and G. Schaefer (Eds.): VISUAL 2008, LNCS 5188, pp. 68–79, 2008.

smaller dam, Pedrogão [2]. The Portuguese National Civil Engineering Laboratory, responsible for the development of the Plan, is a partner in the project.

The study of new interaction devices is another important aspect of this work, it is our intention to run away from the traditional mouse and keyboard configuration.

This paper focuses mostly on the interface and interaction aspects of the project and is organized as follows. The next section presents related approaches in user interfaces for physical phenomena and novel interaction approaches. Then the study scenario and the visualization components are presented. The following section describes the user interface and the interaction options. The paper concludes with preliminary conclusions and directions for future work.

2 Related Work

There are many emergency related projects with distinct features, system support and user interfaces. Here special focus was placed on projects that provide a relevant background to our work and have a multimedia interactive interface with interesting features.

The work reported in [3] explains thoroughly the multimedia system implemented to manage the Valencia's metro emergency plans. Just like the river dams, the metro needs to have an emergency plan. Their solution was turning the emergency plan into a multimedia software system that would integrate text, audio, video, 3D models, and animation to handle emergencies in underground metropolitan transportation.

The Tangible Media Group [4] from MIT has made some interesting proposals in the disaster simulation area. The Tangible DSS (Disaster Simulation System) is a collaborative tool for planning disaster measures, based on disaster and evacuation simulation using Geographic Information Systems. They use a tool called Sensetable[5] which is a table where an image is projected enabling interaction with viewed objects.

Another important interface is the Multi-Touch board developed by Jeff Han [6]. In some examples he uses his multi-touch interface with geographic systems.

While the use of the ArcGIS [7] set of tools (such as ArcMap) has proven useful, through the years of its existence, in the manipulation of geographically-referenced information, these are not primarily designed to perform real time simulations. Their use involves difficulty in keeping the data synchronized, while being displayed in ArcMap and subject to intense update at the same time. Moreover, typical GIS user interfaces, are also not appropriate. Another approach, more similar to our own, is to use a graphical game engine as a platform for the interface.

The HazMat: Hotzone [8] from Carnegie Mellon University is a simulation that uses computer game technology to train fire fighters against chemical and hazardous emergencies. The application looks and feels just like a 3D multiplayer first-person shooter and takes advantage of a game engine to render the emergency scenario.

In the LIFE-SAVER project we are using the real time strategy (RTS) game approach, similar to games like Age of Empires or Comand&Conquer. For that we use an open source game engine called Ogre3D [9]. Before the Ogre3D was chosen, other alternatives were considered. The other main candidate was Virtools [10] a product from Dassault Systèmes'3DVIA. Virtools is a framework solution to create 3D applications visually and not programmatically. Beyond these two solutions several other game engines exist. Some open-source projects such as Crystal Space, Irrlicht or The Nebula Device 2. There are also the Doom, Doom2, Quake and Quake2 engines which have been open sourced by Id Software. The best [11] game engines available are commercial closed-source libraries used especially in the FPS game genre. These are Doom3 and Quake 3 engines from Id Software, Half-Life 2 from Valve Software and Unreal Tournament and Unreal Tournament 2004 from Epic Games.

3 Case Study Scenario

It is very difficult to evaluate one specific Dam Break Emergency Plan (DBEP) as it requires testing very specific natural conditions, associated with flexible access to updateable information and the control of specialized rescue teams. One way to solve this problem is to computationally "generate" natural or induced flooding, as well as the response to the emergency, through simulation using one spatial-temporal representation.

The simulation is supported by the available spatial data from the downstream valley of the Dam. This information is handled by the agent-based simulation engine, which automatically feeds relevant spatial information to the user interface as needed. The simulator will be able to define emergency scenarios which will include available DBEP resources, actors and roles. The system dynamics is visualized and manipulated with a graphical interface representing the emergency scenario, while parameters characterizing this dynamics are registered during the simulation period, for later analysis.

The Alqueva dam is placed in the Guadiana River, about 150 km north of its base level, at Vila Real de Santo António. It is the largest dam in Europe, holding Europe's largest artificial lake.

In a catastrophe scenario, i.e., the collapse of the dam wall, there is a set of events that can be predicted. First of all, there is a formation of a wave, starting from the dam's position. The wave is expected to reduce in strength, while it advances through the valley, eventually becoming a flood hazard. According to studies on the wave, the civil protection infrastructure can only initiate evacuation operations after the first 30 minutes of flooding. Every person placed in the territory that can be reached by the wave before that time, has to escape to pre-determined safe spots. Sirens are placed in the valley to alert the population in the case of emergency.

After this first period of time, the civil protection teams will conduct rescue operations in order to evacuate all the potentially dangerous areas.

4 Visual Features

In order to visualize the scenario case, the multimedia interface must fulfill certain requirements. It must show a full-fledged three-dimensional terrain with a complete and accurate representation of the valley. Important structures and geographic features like the dam, the river or the retained water should be represented. Additionally, there is the need to represent vehicles, people, buildings and other objects. All these

Fig. 1. Emergency Visualization Application using the Interactive White Board

elements must communicate through the network with the simulator, which defines their actions. The interface must be intuitive, easily manageable by people with no special skills in computer interaction (fig.1).

The Computer Games [11] area provides good insights on matching some of these requirements. The idea is to make the emergency visualization interface to look as close as possible as a RTS game (Real-Time Strategy game). In a RTS game the world is seen from above, the units work as agents with limited artificial intelligence and information about each unit or building can be retrieved by clicking on it. Many RTS games have sound alerts and provide several shortcuts that enable a quick response by the user to any emergency.

In the LIFE-SAVER project we are using the RTS game approach, with an open source game engine called Ogre3D [9]. This graphic C++ library supports most of the features and effects currently available in games. Additionally a plug-in to create the terrain was necessary and we used the PLSM2 (Paging Landscape Scene Manager 2) [12].

4.1 Requirements

The LIFE-SAVER Visualization application features several components, most of them inspired by RTS games or by geographic information systems.

Visualization supports multiple maps. The user should be able to see the same map or region in different ways. This includes seeing the map in false colors view, in satellite view, military charts view or black and white. The map can be scrolled in any direction, rotated, zoomed and tilted (to better observe the three-dimensional display).

Almost all RTS games have an overview window where the entire world is represented. The LIFE-SAVER will also have a RTS fashion navigation mini-map. The mini-map should enable fast moving to any map position, and fast situation highlight for any emergency situation.

Special situations should also be highlighted with sound warnings and colored arrows pointing in the direction of the events. Danger situations include, for example, a building being hit by the wave, a water-surrounded group of people or a falling bridge.

Roads are important because this is where most of the simulation will take place. Besides its visual representation, it is important to have an internal graph representation. This graph is used to determinate the actions of each agent, vehicle or person. The graph must be stored as a suitable data structure, to be able to take advantage of search-path algorithms, (e.g., Shortest-Path). Since the terrain is three-dimensional, so are the roads and all the objects in a road take in account the inclination of the terrain.

One of the features is the capability to load information layers from external systems, including shapefiles from ArcGIS [5]. This facilitates loading information like building locations, flooding areas, road blocks, bridges and all the map information that can be represented by points, lines or areas. It is also important that the user is able to switch on and off each layer.

Selecting one object shows important information in a console near the mini-map. It also shows all the enabled actions for the object. Important information, like the identification of the object, is shown directly above the object. When selecting a group of objects, all the common data should be viewed in the graphic console. Agent information can also be edited, including the agent position and its range of action.

Other features include saving the simulation. In disaster simulation, it is always important to be able replay the simulation to allow for detail observation. The evolution of the simulation is recorded and it is possible to move back and forward with it and save it in a file for later viewing.

Most of these features are already implemented and will be described in detail in the next chapters.

5 Simulation Component

The visual environment is supported by a simulation component, which is controlled by the interface and controls all the agents and their behaviour.

The agent-based simulation [13] system integrates GIS data with an Object Oriented, Individual-Based Modelling approach, providing a successful simulation platform for the emergency plan validation process. The simulation model is mainly concerned with the location-based and location-motivated actions of the involved agents, providing data describing the likely effects of a specific emergency situation response, upon which the acceptance of the plan can be supported or changes suggested.

The design of the simulation engine follows closely the Dam Break Emergency Plan (DBEP) requirements. The definition of the DBEP requires prior identification of the entities involved. These entities may be divided into two categories: geographically bound entities, whose role is the execution of a set of actions in the scenario; and organizational entities, responsible for the coordination of emergency response actions. The entities responsible for the coordination of the emergency response operate outside the scenario, and are not physically interacting with it. The entities responsible for the execution of the plan's actions are directly interacting with the scenario, and their mobility characteristics (moving by foot or using a specific vehicle) are one of the two main defining aspects from an emergency response point of view, the other being their expertise characteristics (medical personnel, firefighting personnel, etc.).

Given the special relevance of the entities locomotion characteristics, for the emergency plan, a previously determined type of entity is directly linked to its corresponding agent. Therefore, agents will be defined based on possible travel characteristics.

The geography of the scenario is a major factor for its characterization (as seen with the definition of the agents). The simulation model must take into account those characteristics, and be parameterized by it. The geographic information involved in the project is currently managed in a Geographic Information System (GIS). The simulator will be fed GIS data and this information will be used as a constraint to the various agents contemplated by the DBEP. A geographic data stack (Feature Space) is used to manage each geographic feature (e.g., roads, water streams, buildings, etc.), allowing the confrontation of a specific feature (or set of features) to an agent as a movement constraint.

Each agent has a set of actions (either predetermined as standard response or imposed by the entities responsible for the response organization) and while the scenario evolves, the emergency response will be the result of the coordination of several agents, each with specific behaviour characteristics and movement constraints.

To accomplish the specific evolution of one agent's actions, a priority system of actions is proposed. Each agent will initiate the scenario response with a predetermined set of actions, and, if seen fit by emergency response coordinators, new actions with higher priority may be issued, or even a reorganization of the previous actions can be required to face a specific scenario configuration.

The prioritization of actions enables the coexistence of routine operations and quick response actions by the agents, and allows the definition of a generic interface to deal with the agent's actions.

Demographic information is also important, and a database with information gathered in the location, contemplating the characterization of the people involved, health conditions and location, will be available.

6 The Visual User Interface

Presently we have created a prototype of the emergency testing application. Currently the program loads a three dimensional terrain. It loads geographic information from GIS data, namely ShapeFile data from ArcGis. There is a navigation system implemented, with sliders and a mini-map. The user can turn on and off several layers of information. The interface allows the selection of agents and the definition of actions for them. Currently it supports and presents a simulation of a flooding wave (fig. 2).

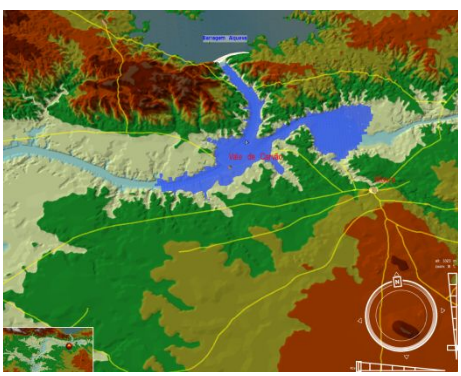


Fig. 2. Simulation of the flooding wave in the Guadiana River created by the failure of the Alqueva Dam

6.1 Design Architecture

The entire application was built using C++, using an object-oriented and event-based approach. The programming infrastructure is provided by Ogre3D game engine. Two C++ libraries were created (fig.3). One called BaseProject acts has a framework that takes care of all Ogre3D initializations, IO management, scene and level context and sound management. The other library is UtilityProject, an utility toolkit used to create simple custom objects, custom lines, custom materials and additional useful resources.

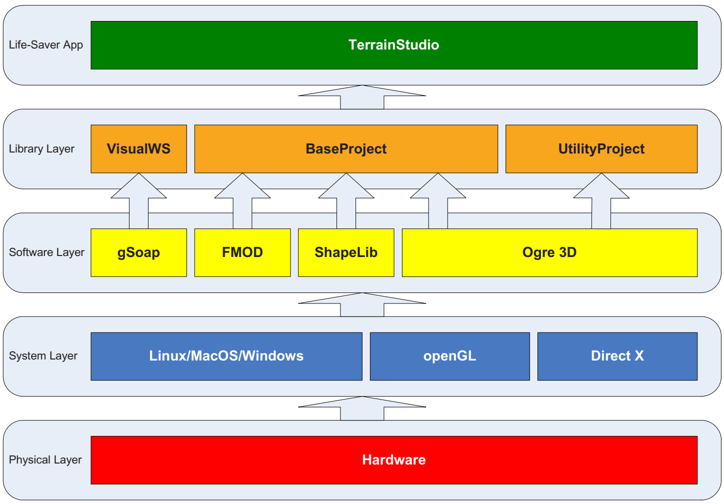


Fig. 3. Diagram of the Visualization Component. The LIFE-SAVER project is responsible for the two top layers.

The code was organized with a set of Manager Classes. These follow the "Singleton" code pattern, which means that for each class there is only one instance of the class running. This is important when, for example, it is necessary to have a global InputManager class that must be accessible from many parts of the code and has to be unique in the running program.

One of the manager classes is the StateManager. This class runs the main render loop of the program. The main advantage of this class is that it has a stack of Scenes. Each Scene is a state of the program and it can be a new level or a configuration scene. Each scene can be "pushed", "popped" or redirected to other scenes. This allows the fast creation of Scene flows.

Another important method was the use of an event based approach. When an object interacts with the mouse it is registered as a listener in the Input Manager and then the desired handlers are implemented in the object's class. This allows a decentralized approach, where each object has his input code, as opposed to having all the input treated in one place.

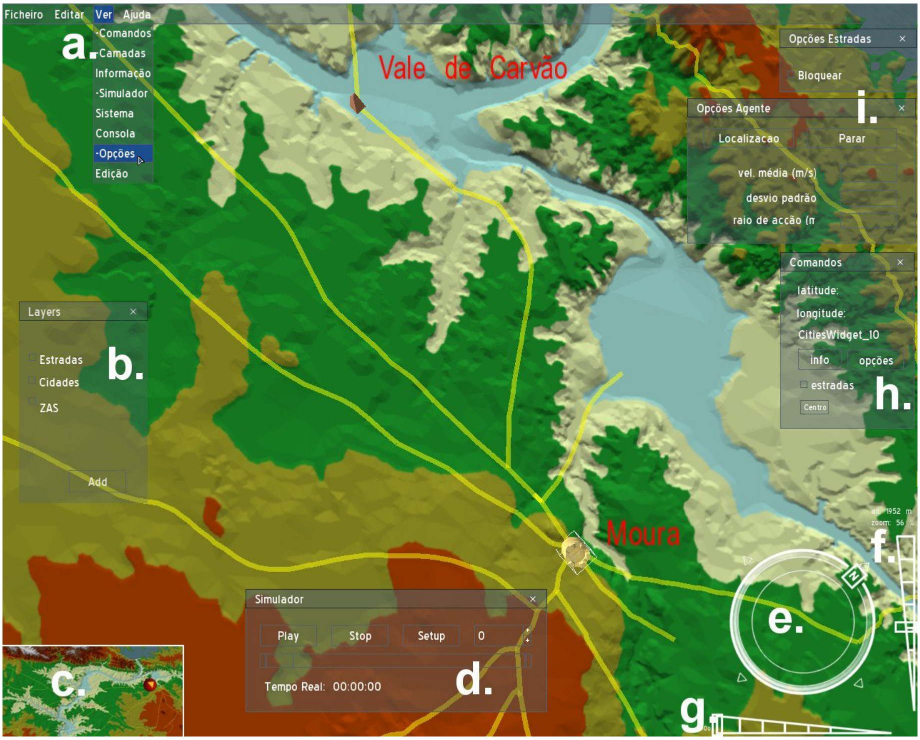


Fig. 4. Visual Interface: a) main menu, b) Layers, c) mini-map, d) simulator controller, e) map rotation, f) zoom, g) pitch, h) command console, i) agent options. To move the map the user has to drag the map.

6.2 Graphic Implementation and Challenges

Through the partnership with LNEC, a large quantity of relevant data was acquired. This data includes maps and geographic information including digital information about roads, villages, buildings in risk, flooding zones and flood wave parameters.

To generate the 3D terrain was necessary to create a gray scale height map. This map was generated taking into account the contour lines of the terrain and the topographical height points of the river valley.

Using the PLSM plug-in of Ogre3D the map was divided in several small squares that are dynamically loaded as needed.

The representation of elements in the scenario required some developments in the computer graphics area. For example, for the roads representation it was necessary to create a line with thickness that followed the terrain's relief.

Besides Ogre3D other external C++ libraries were studied and integrated. To open ArcGis data the Shapelib [14] was used. To use sound the FMOD [15] library was tested. To create visual GUI interfaces with common window components such as buttons, editboxes or checkboxes the Crazy Eddies GUI library (CEGUI) was used.

A web service was established between the Visual component and the Simulation component. To use the web services it was necessary to integrate an open source library called GSoap [16].

6.3 Interface

The current application interface is essentially centered on the world navigation and on the simulation agent interaction options. Being a project related with emergencies planning and simulation the prototype application is designed to be used in a control center where many people interact collaboratively. Because of that we elected as our primarily input device the Interactive White Board (IWB) and the interface reflects that. All menus can be accessed using only the left click (equivalent to the screen touch).

The navigation is made using the bottom right interface. The knob rotates the world indicating the North, the vertical slider changes the zoom and the horizontal slider changes the pitch of the viewer enabling a 3D perspective. To move the world the user only has to drag the terrain. This interface has similarities with the Google Earth's configuration.

The left mini-map allows a fast recognition of where the user is. If the user clicks on it the viewing area is redirected to that area. This is a similar behavior to the RTS games.

The graphic user interface (GUI) features several menus as seen in figure 4. When the user selects an agent or a terrain feature, additional information is displayed on the Information menu. The Options displays the available options for that specific agent. The Simulation menu controls the simulation flow. An additional menu is the Layers menu that allows the user to switch on and off layers of information.

7 Interaction

One of the main goals of the LIFE-SAVER project is to research new forms of interaction that depart from the traditional mouse and keyboard setup.

There are several devices that are being tested to be used in the LIFE-SAVER project. What is important about each device is its availability, degree of supported interaction and amount of people that can interact at each time with the application. Currently, we are supporting the mouse, digitizer pen and the touch-sensitive board. Additionally several studies are being made to use collaborative multipoint interfaces.

7.1 Interactive White Board and Tablets

The Interactive Whiteboard (IWB) is a large white interactive board where the computer image is projected and enables navigation on the screen, by touching it. It only supports one touch at any given time. Whenever the board is touched, it assumes that the touch position is where the mouse cursor is. The idea is to use the IWB as the central piece of the civil protection command room, placing it in a wall or on a table (fig.1). The IWB is a very practical hands-on device but has some limitations. It only handles absolute position, when the user touches the board, and only relative position when dragging on the board. Mainly, it only supports left clicking. A right click button is available, but it is not intuitive and easy to reach. Another handicap is that users usually interact with their fingers and fingers tend to be less precise than a pen or a mouse. There are also some synchronization problems between the position touched by the user and the cursor's position.

In this project, the IWB is used with large buttons and draggable areas, where the user can do all the world moving operations with his hands.

The same interface that works with the IWB is naturally usable in a Tablet PC. For demonstration purposes the Tablet is usually used due to its portability.

7.2 Multipoint Research

The main disadvantage from the IWB is that it only allows a single touch point. This makes the interaction rather limited. Using a multi-touch interface will lead to more invisible interfaces. All the operations will be dealt with finger gestures. We will be able to rotate and zoom with two fingers, dragging them on the board [6].

To create a multipoint interface several solutions were studied [5], including one using computer vision.

Our multipoint device is a table with a glass top (fig.5). Beneath the glass there is a common USB camera aimed right up. The glass is covered with a white polyethylene plastic that makes objects far away diffuse and is see-through with near objects. On the top of the glass we place multiple objects that have special marks on the bottom. The goal is to capture the marks with the camera and use detection algorithms to determine their positions on the table.

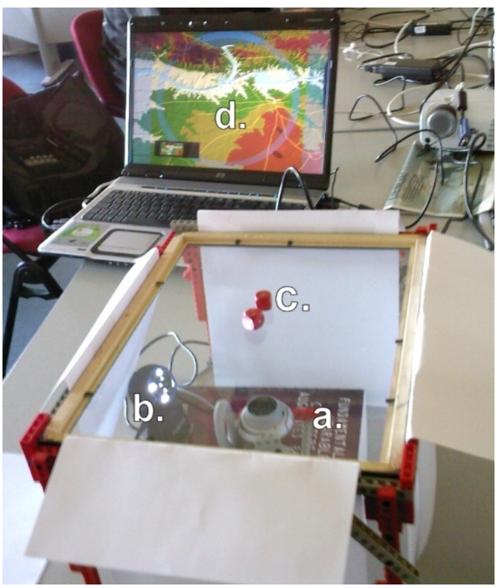


Fig. 5. Multipoint research device: a) USB web camera, b) illumination, c) two markers that sit on top of the glass, d) computer where the image is processed and the position of the marks is detected

Fig. 6. Camera view showing the AR toolkit markers. The algorithm detects the position of the black squares.

The users interact with the computer by moving the several marked objects (fig.6). Each object is represented on the screen with a different colored cursor. The marks and detection algorithm used are based on the Augmented Reality Toolkit (AR toolkit) [19]. Besides the marker position it is also possible to detect if the mark is touching the glass and rotation of the mark. This allows creating "click" events and rotation events that work like knobs.

The main problems of this device results essentially from lightning conditions and the low frequency of the camera (30 fps).

This research is still in an initial stage and is not ready for real users' deployment but shows promising results.

8 Conclusions and Future Work

This paper presents a system for emergency simulation, focusing on the user interface and design decisions for this component. Using a game engine to support the interface, enables a rich set of interaction possibilities, which can be augmented as more actions are needed. Supporting multiple devices enables different usage scenarios, including decision makers, actors that are involved in the emergency, and the general public. In the system, we are also experimenting with several metaphors for navigating in the emergency area and further evaluations will be carried out. In the next months, we will be testing the different devices while adding functionality and evaluating the new features with the intended users of the system.

Acknowledgements

Project LIFE-SAVER is funded by the Portuguese Foundation for Science and Technology through grant PDCT/AMB/57247/2004. We would like to thank LNEC (Laboratório Nacional de Engenharia Civil - National Civil Engineering Lab) for all the geographic and scientific information and support. We thank all the people in IMG (Interactive Multimedia Group) at FCT-UNL.

References

1. Betâmio de Almeida, A., Matias Ramos, C., Santos, M. A., Viseu, T.: Dam Break Flood Risk Management in Portugal. LNEC, Lisbon (2003)
2. Santos, M. A., Gamboa, M.: Plano de Emergência Externo da Barragem de Alqueva (Alqueva Dam External Emergency Plan). Report nº 326/03_NTI, p.169. LNEC, Lisbon (2003)
3. José, H., Canós, J.H., Alonso, G., Jaén, J.: A Multimedia Approach to the Efficient Implementation and Use of Emergency Plans. IEEE Computer Society Press, Los Alamitos (2004)
4. Tangible Media Group, `http://tangible.media.mit.edu/`
5. Patten, J., Ishii, H., Hines, J., Pangaro, G.: Sensetable: a wireless object tracking platform for tangible user interfaces. In: SIGCHI conference on Human factors in computing systems (CHI 2001), pp. 253–260. ACM Press, New York (2001)

6. Han, J.Y.: Low-Cost Multi-Touch Sensing through Frustrated Total Internal Reflection. In: Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology. ACM Press, New York (2005)
7. ArcGis, http://www.esri.com/
8. HazMat, http://www.etc.cmu.edu/projects/hazmat_2005/
9. Ogre3D, http://www.ogre3d.org/
10. Virtools, http://www.virtools.com/
11. Kot, B., Wuensche, B., Grundy, J., Hosking, J.: Information visualisation utilising 3D computer game engines case study: a source code comprehension tool. In: Proceedings of the 6th ACM SIGCHI New Zealand chapter's international conference on Computer-human interaction: making CHI natural. ACM Press, New York (2005)
12. PLSM2 Ogre3D Plug-in, http://tuan.kuranes.free.fr/Ogre.html
13. Sabino, A., Nóbrega, R., Correia, N., Rodrigues, A.: Life-Saver: Flood Emergency Simulator. In: ISCRAM, Washington D.C. (2008)
14. ShapeLib, http://shapelib.maptools.org/
15. FMOD, http://www.fmod.org/
16. Gsoap, http://www.cs.fsu.edu/engelen/soap.html
17. Dix, A., Finlay, J., Abowd, G., Beale, R.: Human-Computer Interaction. Prentice-Hall, Englewood Cliffs (1998)
18. Buxton's site, http://www.billbuxton.com/multitouchOverview.html
19. AR toolkit, http://www.hitl.washington.edu/artoolkit/

3D Mesh Exploration for Smart Visual Interfaces

Georg Fuchs, Mathias Holst, René Rosenbaum, and Heidrun Schumann

Institute for Computer Science, University of Rostock

Abstract. Today's fast growth of both the number and complexity of digital 3D models results in a number of research challenges. Amongst others, the efficient presentation of, and interaction with, such complex models is essential. It therefore has become more and more important to provide the user with a smart visual interface that presents all the information required in the context of the task at hand in a comprehensive way. In this paper, we present a two-stage concept for the task-oriented exploration of 3D polygonal meshes. An authoring tool uses a combination of automatic mesh segmentation and manual enrichment with semantic information for association with specified exploration goals. This information is then used at runtime to adapt the model's presentation to the task at hand. The exploration of the enriched model can further be supported by interactive tools. 3D lenses are discussed as an example.

1 Motivation

Today, more and more processes related to product development, production and servicing are implemented almost exclusively by means of the CAx software family. This results in a fast growth of both the number, as well as the complexity and fidelity, of digital 3D models that are used for a variety of purposes, such as system simulation, product presentations, or technician training.

This trend results in a number of research challenges [1,2]. Besides questions of how to create and (re)use models or how to preserve semantics across the entire modeling process, the efficient presentation of, and interaction with, such complex models is essential. In other words, it has become more and more important to provide the user with a *Smart Visual Interface* that presents all the information required in the context of the task at hand in a comprehensive way. This includes support of *Mesh Exploration* through such interfaces, i.e. to support the user in the process of understanding or verifying the structure and functionality of an object (modeled as a polygonal mesh) within the context of the task at hand. Enriching a model with additional information helps to communicate complex spatial relationships and to identify individual object components, e.g. parts of machinery or organs in medical data. Especially, such information can be used to adapt the presentation according to a user-specified exploration goal [3], e.g. by accentuating important components and hiding unimportant ones. Offering interactive exploration further discards the need for predefined animations.

Our definition of a smart visual interface is therefore based on two main aspects: (i) a description of the task performed, the *task model*, and (ii) the use of this model for *adaptive representation* of the 3D model for the task at hand.

M. Sebillo, G. Vitiello, and G. Schaefer (Eds.): VISUAL 2008, LNCS 5188, pp. 80–91, 2008.

2 Background and General Approach

The background for our investigations is a mobile maintenance support scenario in the scope of an industry research project [4]. Tasks associated with repair and maintenance work typically follow laid down procedures, which lend themselves to explicit notation in the form of a task model. Generally, a task model forms a hierarchy of compound tasks with subtasks. Subtasks have conditional and/or temporal relationships between them, representing the step-by-step nature of individual working steps. Thus, we proposed a concept of using task models in visual interfaces, as the core of an 'e-Manual' application that presents the user with documentation required for the completion of her current maintenance task.

[4] mainly considered the presentation of visual content that is most common to technical manuals: two-dimensional illustrations and schematics, especially digitized raster images from printed documentation. The challenge with this type of content is the absence of any structure or semantic information about the depicted object: the image is but a 'pixel soup'. In order to be able to highlight task-specific aspects of such an illustration, appropriate Regions of Interest (RoI) have to be defined and attributed with meta data beforehand. RoIs are then assigned a Degree of Interest (DoI) with respect to a given task. These information are finally used to adapt the presentation accordingly. This can for example be the use of Focus & Context distortion, or the accentuation of important regions through hue and saturation adjustments. Fig. 1 gives an illustrative example for this kind of task-based adaptation.

Extending the maintenance e-Manual to also include 3D representations would increase the learning efficiency. 3D models generally support the interactive exploration of complex spatial configurations much better than 2D illustrations. For example, the 3D representation of a piece of machinery can help to identify and locate an attachment part. 3D modeling software usually employ hierarchical structures (scene graphs) that represent a segmentation into independent components. However, such hierarchies and/or its associated meta data may not be available due to limitations of the data format, intellectual property protection considerations, or simply because no meta data was generated for the

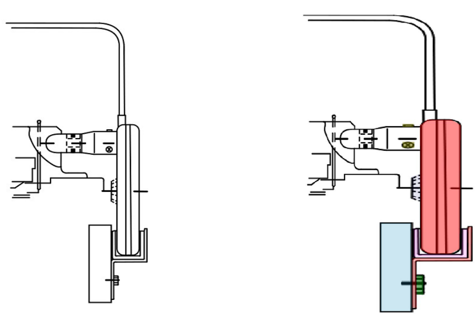


Fig. 1. Example for the task-based adaptation of a 2D raster image (left). Important image regions have been enlarged and accentuated by color adjustment (right).

model components in the first place. Analogous to raster graphics, such 3D models mostly are but a set of triangles without semantics, a ‘triangle soup’. In this case defining RoIs becomes a task of first segmenting the triangle mesh into meaningful components. Task-based adaptation of the representation could then include the automatic choice of a suitable initial viewpoint, the component-wise application of Level-of-Detail (LoD) techniques, highlighting important parts, and the omission or transparent representation of occluding geometry.

In this paper, we therefore concentrate on integrating task models with 3D representations to obtain a task-oriented visual output of complex triangle meshes. This imposes two additional requirements to the modeling process: (i) enriching the 3D model with semantics, and (ii) associating aspects of the model with the task at hand. These crucial stages of the process can not be performed solely by automatic means. While segmentation algorithms form the basis for a semantic decomposition of the triangle mesh, an interactive inspection step is needed nonetheless to ensure that components of the model can be appropriately associated with the (sub-)task at hand. We will show how this inspection can be supported by an appropriate exploration viewer and interactive 3D lenses.

The following Sect. 3 gives an overview of related work as well as references to our previous work on task-oriented 2D representations. The general approach is discussed in Sect. 4. Section 5 gives a summary and some concluding remarks.

3 Related Work

Enhancing raw geometry with additional high-level shape information is an important topic in both industry and academia. Thus results the requirement to decompose a model into meaningful components. To this end, several segmentation methods have been devised [2,5,6]. For a good recent overview, see [7].

The segmentation of a surface mesh can be carried out either according to purely geometric aspects [8,9] or semantic-oriented [10–16]. In the former case the mesh is decomposed into patches that are equal with respect to a certain property (curvature, distance to reference plane) whereas latter methods try to identify object parts that correspond to relevant features of the object’s shape. Geometry-based approaches can also be used as a pre-process to the detection of relevant features. Semantic-oriented approaches have enjoyed increasing attention in recent research because they support morphing, 3D surface reconstruction, skeleton extraction well as the modeling by composition paradigm which is based on the natural decomposition of surfaces. Yet the majority of approaches in computer graphics were not designed to detect features in a specific context such as product modeling in the industrial production process [17]. Generally, different algorithm perform better for certain types of model features. For this reason, Attene et al. [18] propose what they call *multi-segmentation*, i.e. to use several segmentation algorithms in parallel and to mix-and-match their individual segmentation results.

Our approach is based on the *feature point & core extraction* approach presented in [12] that aims at creating pose- and proportion-invariant segmentations. It works in three steps: transformation of the mesh vertices into a pose-invariant MDS (multi-dimensional scaling) representation, robust extraction of prominent feature points from the MDS domain, and finally the extraction of the object's core component. The latter is done by mirroring the transformed mesh vertices on the object's bounding sphere and subsequent vertex classification. The algorithm achieves consistent segmentations of similar models that differ only in pose or component proportions. It also adheres to the minima rule [19] by detecting limbs and protrusions that can be visually separated along local concavities from the object's core. This makes the method suitable for a wide range of both natural and technical models.

Besides segmentation methods for the semantic decomposition of triangle meshes, the efficient presentation of these complex models for various application purposes is another important research topic. Interactive tutoring systems need to communicate specific aspects of a complex model (e.g. [3]). A primary question therefore is how to accentuate relevant features while minimizing the visual impact of irrelevant or occluding objects. Non-photorealistic rendering styles [20] are a good way to highlight important shape features. Transparency [21,22] and cutaway views [22,23] are very efficient static methods to remove or subdue irrelevant occluding geometry. Some interactive tutoring systems automatically determine optimal views which minimize occlusions [24] or plan camera movements to guide the interactive exploration [25]. Dynamic labeling is a method to establish co-referential relations between textual and visual elements [26].

In addition to static accentuation/de-accentuation methods, so-called *lens techniques* can be used to dynamically adapt the presentation in a user-defined Region of Interest. Magic Lenses were first introduced by Bier et al. [27,28] as a focus & context technique for 2D interfaces. A Magic Lens is a movable, semi-transparent user interface element that changes the representation of data viewed through it. Generally, lenses can be used not only for magnification but also numerous other effects, such as filtering (remove objects under lens to reduce clutter, e.g. [29]), adaptation of the presentation (e.g. a change of rendering style) and LoD (data through the lens is rendered at higher resolution, e.g. [30]). Several lenses can be combined to produce composite effects where they intersect.

The Magic Lens metaphor was extended to three dimensions by Viega et al. [31]. They implemented two types of 3D lenses: a 'flat' lens that projected a volume of influence into the scene, and a volumetric lens that affected content falling within the space of a cube. Both approaches exploited hardware support for clipping planes which made it possible to divide the scene into 'lensed' and 'un-lensed' spaces in real-time. The use of 3D lenses as interactive tools for exploration and interaction with virtual 3D scenes also attracted more recent research in the domain of augmented reality applications [32].

To summarize, there exist a number of segmentation algorithms, semantic-oriented approaches in particular, that can form the basis for the annotation of a 3D model, as well as rendering methods for their efficient presentation in the context of the task at hand. There are approaches [18,33] similar to ours. These however concentrate on ontology browsing for shape annotation and retrieval, especially [18] having a strong emphasis on the exploitation of multi-segmentation. For our purposes, the choice of a semantic-oriented algorithm with good overall applicability for decomposing arbitrary models into meaningful components, which can be associated with (sub-)tasks from the task model, is sufficient as a starting point. We therefore chose the feature point & core extraction algorithm [12] to illustrate our approach.

4 Concept

The task-oriented approach can be broken down into two stages: the model authoring phase (including inspection), and rendering the model in the *Exploration Viewer* as part of the smart visual interface, using the task model for adaptation of the presentation (Fig. 2).

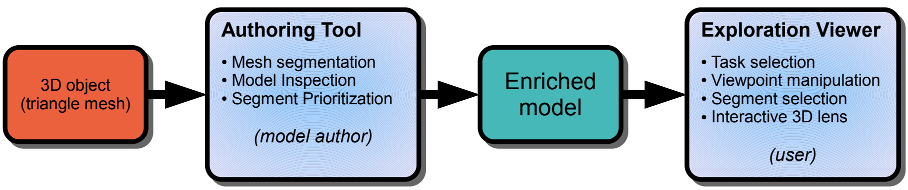


Fig. 2. Overview of the task-based model exploration process

The following Sect. 4.1 discusses the individual steps of the authoring process. The adaptation of the enriched model representation for a given task and the use of interactive 3D lenses is described in Sect. 4.2.

4.1 Model Authoring

The authoring phase can be broken down into three subsequent operations that are performed on the model to obtain a suitable model enrichment:

1. Mesh Segmentation
2. Model Inspection
3. Segment Prioritization

Mesh Segmentation is the first step in the authoring process. As a preprocessing step to the actual segmentation, a mesh simplification is performed to build a hierarchy with decreasing LoD. This not only improves performance of the segmentation algorithm, it also reduces the susceptibility of automatic algorithms for small perturbances of the surface, such as scanning artifacts. Moreover, it later allows to select different LoDs per segment to adapt the visual representation (cf. Sect. 4.2).

The model author starts by selecting the LoD in the hierarchy at which the segmentation should be performed using a slightly modified feature point & core extraction algorithm. In the original approach [12], the object core was extracted by mirroring the mesh vertices on the bounding sphere. Any vertex that maps to the convex hull of the mirrored vertex set is classified as belonging to the core. In order to improve the segmentation results for arbitrary objects that may exhibit bumpy surfaces or scanning artifacts, we introduced an additional parameter to classify vertices with a maximal hull distance $d \geq 0$ as core vertices as well. Also, unlike proposed in [12], our prototype does not yet apply a MDS transformation of the mesh vertices before feature point detection. However, for rigid technical models that are not posed the results are quite acceptable.

As discussed in Sect. 3, the performance of automatic algorithms generally depends strongly on the model geometry and its prevalent features. Using multi-segmentation [18] is one way to ensure meaningful results. However, our approach is to iteratively refine the initial results with the help of interactive inspection of the partially segmented model instead. This ensures that besides geometric features, also application-specific features can be taken into account. An technical domain example are weld seams in the concavities between model segments that would hardly be detected as independent segments by automatic methods.

Model Inspection. During model inspection, the author can verify whether the detected segments match the desired result. To this end, the model is rendered with the segment patches assigned random colors. The author can manipulate the 3D view (zoom, rotate) to inspect each segment patch. Individual segments can be selected, either by picking them in the 3D view or from a list control (cf. Fig. 4), and sub-segmented further by iteratively applying the segmentation algorithm to the associated triangle patch.

Another option during the inspection phase is to manually segment parts of the model. Manual segmentation offers great flexibility to overcome algorithm limitations as it does not rely on its suitability for a given geometry. By interactively placing simple shapes, such as spheres or cubes, in the 3D scene the author can quickly define RoIs (Fig. 3a). Parts of the mesh intersecting these volumes can be assigned to a new segment, or added to an existing one.

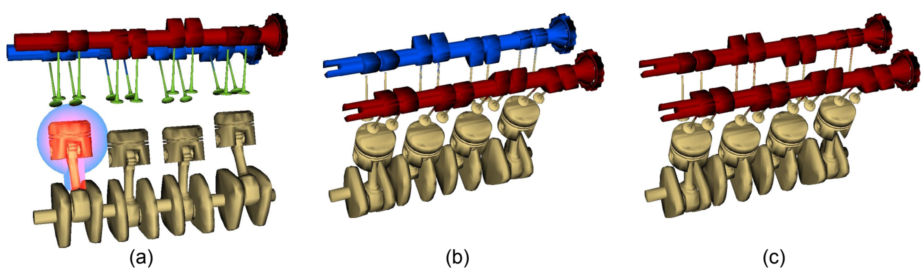


Fig. 3. Definition of 3D RoIs of an object to obtain a first coarse segmentation (a). Refinement is possible by selecting and further subdividing segments, or by merging multiple segments (b, c).

The resulting segmentation of the current iteration can be refined even further by three additional operations:

– Expanding a segment by adding a triangle strip at its boundary,
– Shrinking a segment by removing a triangle strip from its boundary, and
– Merging two segments into one (Fig. 3b-c).

Expanding and shrinking segments can also be applied to obtain a disjunctive, exhaustive segmentation of the entire model, if so desired. This is done by expanding patches into mesh areas yet unassigned to any segment, and shrinking overlapping segments in the respective areas. Note that the final segmentation does neither need to be disjunctive – triangles may be associated to more than one segment with fuzzy borders – nor exhaustive if some regions of the surface do not belong to any *meaningful model component*, i.e. relevant to at least one task.

Finally, after the segmentation for the selected LoD has been completed, the result is mapped onto all other levels of the model hierarchy. The surface mesh can therefore be presented with variable LoD in the exploration viewer.

Segment Prioritization. After the desired segmentation has been achieved that reflects the meaningful model components, the author can annotate them with semantic information. Basis of the annotation process is the task model for a given maintenance task (cf. [4]).

Therefore, our authoring tool allows to associate, for every task, each segment a priority or *importance* normalized to the [0..1] value range. Therefore, importance values express the Degree of Interest with respect to the task at hand and are therefore stored in the nodes of the task model. The default importance value is 0, i.e. the component is irrelevant, while segments with assigned values > 0 constitute meaningful model components in the context of the task at hand.

Optionally, segments can also be associated with basic rendering attributes (color, material, transparency) and style (e.g. contour sketches, wire frame, translucent, solid). This can be used to enforce visual grouping of related components, for example by color-coding functional groups of components. Also, the author can define a good initial viewpoint on the model for each task.

Fig. 4 shows a screenshot of the authoring tool with the inspection 3D view and the task pane for importance value adjustment.

4.2 Exploration Viewer

The exploration viewer as part of the smart visual interface renders the model and allows to interactively explore it. The viewer automatically creates an *initial view* of the 3D model according to the information from the task model. The importance value thereby is the primary mechanism to adapt the presentation of the model to the task at hand. It is used to control the following aspects:

Initial viewpoint selection: For an optimal view on the 3D representation for a given task, the virtual camera should be positioned so that at least the most important model components are initially visible. This viewpoint is either stored in the associated task model node, or the importance values from the task model can be used to control viewfinding algorithms like [34].

Level of Detail selection: Components with the maximum importance use the highest-resolution base mesh, whereas less relevant components are rendered at correspondingly lower LoD. Irrelevant segments (zero importance) are rendered at the lowest LoD, or are even not rendered at all.

Accentuation: To efficiently convey the important aspects of the task at hand, the attention of the user should be guided to the relevant components. This is achieved by accentuating important model components and visually subduing (de-accentuating) unimportant ones according to the values from the task model. There is a vast number of methods for accentuation (cf. Sect. 3). Examples include simple adjustment of material properties such as hue, saturation or alpha value (Fig. 5a), selection of rendering styles (Fig. 5b) including NPR techniques [20], and creation of cutaway views [23].

Moreover, the task model can provide additional information that may be used in conjunction to the enriched model description. Two examples are

- To provide names or short designations that can be used to dynamically label segments (our labeling approach is described in [35]), and
- To provide a textual description of work activities on, or additional information about the component(s). These can be displayed as text juxtaposed to the 3D view, or communicated via Text-to-Speech output in multi-modal interfaces, as discussed in [4].

As these information reside in the task model they can be collected and maintained independently from the model authoring process.

However, while the above measures generate appropriate initial views, a single illustration often does not suffice to depict all salient visual objects due to (partial) occlusions. Therefore, the ability to interactively *explore* the model is still important.

The exploration viewer supports this in a number of ways. Besides the obvious view manipulation, it is always possible to select and thus focus on an arbitrary model component. This overrides the importance value for the associated segment from the task model and the visual representation is updated accordingly (Fig. 5c). Like in the authoring tool, component selection is realized either through picking in the 3D view or through selecting the associated list entry (cf. Fig. 4).

Another tool for exploring the model in a flexible way is the use of interactive 3D lenses that can be positioned freely in 3D space (Fig. 6). In doing so, the user defines a dynamic RoI in which rendering parameters are changed independently of segment boundaries. In our prototype, we so far employ only a single lens that changes the LoD of the mesh. This allows a user to view those parts of the model in full fidelity even if their predefined relevance for the task at hand is minimal. However, other lens functions such as switching the rendering style or an 'X-ray' effect to interactively reduce occlusion could be realized as well.

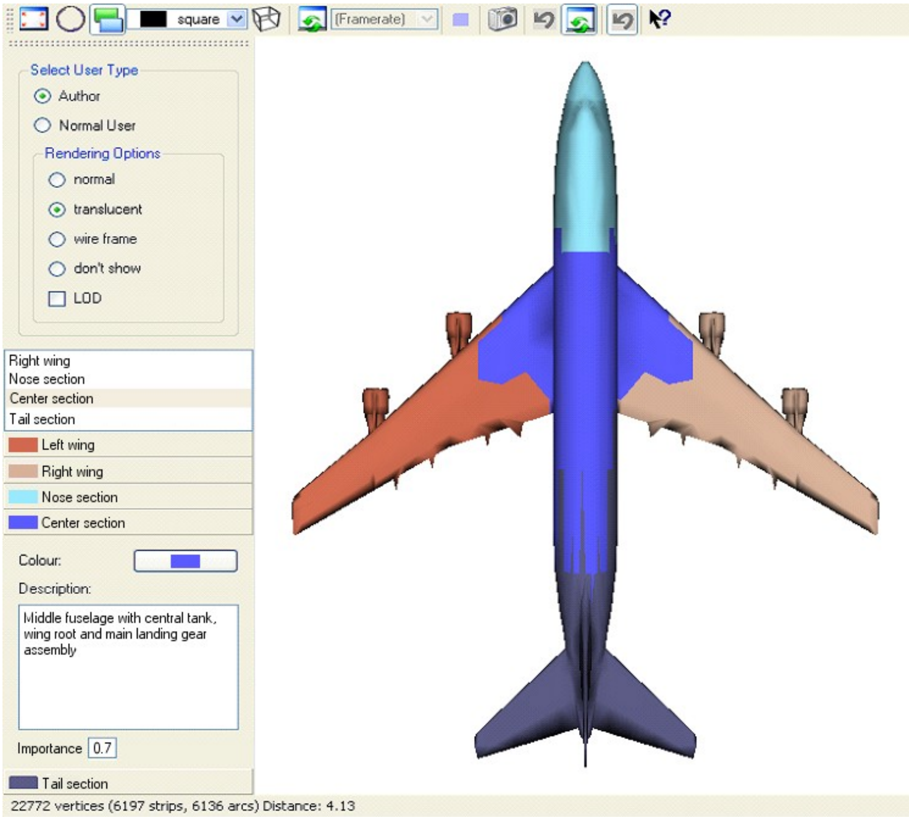


Fig. 4. Screenshot of the authoring tool prototype. The 3D inspection view shows the current segmentation of the model. The control panel on the left is used to set the segment importance value and other properties for the currently selected task.

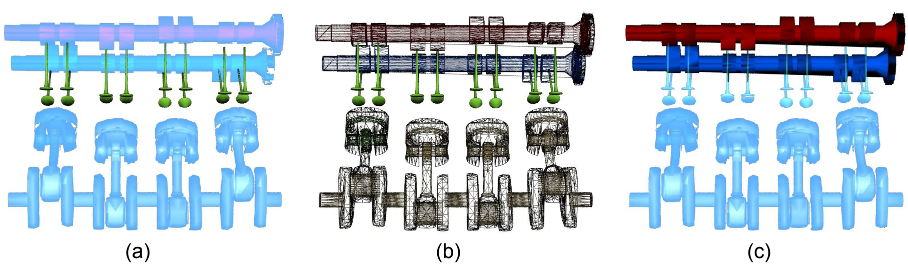


Fig. 5. Example for the application of different rendering styles for segment accentuation (a, b). The valves are defined as the most important component for the current task, other components are visually subdued. (c) shows the result of manually selecting two other segments (both camshafts) as the focus, overriding the importance values from the task model.

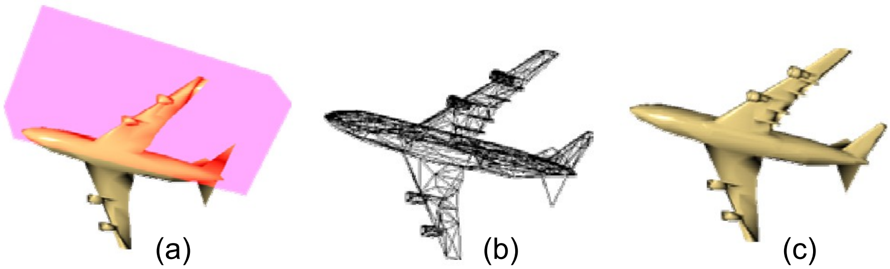


Fig. 6. Applying a 3D lens to manipulate the level of detail in selected object regions. (a) Lens volume, (b) resulting hybrid LoD mesh, (c) visual result.

5 Summary

In this paper we presented a concept for the task-oriented exploration of 3D polygonal meshes. It uses a task model as well as a combination of automatic mesh segmentation and manual enrichment of segments. The association of model components with importance values in the task model is a versatile means for the adaptation of the presentation to the task at hand. The exploration of the enriched model can further be supported by interactive tools such as 3D lenses.

The concept has been prototypically implemented using C++, OpenGL and Qt. The authoring tool supports hierarchical mesh simplification, automatic and manual segmentation and segmentation prioritization. Automatic segmentation algorithms are integrated as plug-ins. Currently, only a slightly modified feature point & core extraction algorithm [12] is available. The exploration viewer so far supports the importance-driven selection of the segment's LoD and its rendering style (culled, wire frame, translucent, darkened solid, solid).

Future work will go along several lines. The automatic segmentation capabilities could be improved by implementing additional algorithm plug-ins. Good candidates for technical models are Plumber [15] and Hierarchical fitting primitives [10]. Further improvements to the exploration viewer could include support for importance-driven NPR (e.g. from line sketches to stippling to fully shaded), improved viewpoint selection, and additional lens effects for interactive exploration. We further strive to evaluate the proposed concept and planned extensions with real-world data in cooperation with our industry partner within the research project [4]. The ECS International Ltd. is a supplier of software solutions for Product Data Management (PDM) in engineering, where interaction with and inspection of complex technical 3D models features prominently.

References

1. Mayer, J.: The New Main Research Areas of Fraunhofer IGD. Computer Graphics topics 19(2), 18–19 (2007)
2. Attene, M., Biasotti, S., Falcidieno, B., Mortara, M., Patanè, G., Spagnuolo, M.: Topological, geometric and structural approaches to enhance shape information. In: Proceedings of Eurographics Italian Chapter, Catania (I), pp. 7–13 (2006)

3. Preim, B., Ritter, F.: Techniken zur Hervorhebung von Objekten in medizinischen 3d-Visualisierungen. In: Simulation und Visualisierung, pp. 187–200 (2002)
4. Fuchs, G., Reichart, D., Schumann, H., Forbrig, P.: Maintenance Support – Case Study for a Multimodal Mobile User Interface. In: IS&T/SPIE's 16th Ann.Symp.Electronic Imaging: Multimedia on Mobile Devices II (2006)
5. Chazelle, B., Dobkin, D.P., Shouraboura, N., Tal, A.: Strategies for polyhedral surface decomposition: an experimental study. In: SCG 1995: Proceedings of the eleventh annual symposium on Computational geometry, pp. 297–305. ACM Press, New York (1995)
6. Li, X., Toon, T.W., Huang, Z.: Decomposing polygon meshes for interactive applications. In: I3D 2001: Proceedings of the 2001 symposium on Interactive 3D graphics, pp. 35–42. ACM Press, New York (2001)
7. Attene, M., Katz, S., Mortara, M., Patanè, G., Spagnuolo, M., Tal, A.: Mesh segmentation - a comparative study. In: IEEE International Conference on Shape Modeling and Applications (SMI 2006) p. 7 (2006)
8. Garland, M., Willmott, A., Heckbert, P.S.: Hierarchical face clustering on polygonal surfaces. In: I3D 2001: Proceedings of the 2001 symposium on Interactive 3D graphics, pp. 49–58. ACM, New York (2001)
9. Mangan, A.P., Whitaker, R.T.: Partitioning 3d surface meshes using watershed segmentation. IEEE Transactions on Visualization and Computer Graphics 5(4), 308–321 (1999)
10. Attene, M., Falcidieno, B., Spagnuolo, M.: Hierarchical mesh segmentation based on fitting primitives. Vis. Comput. 22(3), 181–193 (2006)
11. Katz, S., Tal, A.: Hierarchical mesh decomposition using fuzzy clustering and cuts. ACM Trans. Graph. 22(3), 954–961 (2003)
12. Katz, S., Leifman, G., Tal, A.: Mesh segmentation using feature point and core extraction. The Visual Computer 21(8–10), 649–658 (2005)
13. Liu, R., Zhang, H.: Segmentation of 3d meshes through spectral clustering. In: PG 2004: Proceedings of the Computer Graphics and Applications, 12th Pacific Conference, Washington, DC, USA, pp. 298–305. IEEE Computer Society, Los Alamitos (2004)
14. Mortara, M., Patanè, G., Spagnuolo, M., Falcidieno, B., Rossignac, J.: Blowing bubbles for multi-scale analysis and decomposition of triangle meshes. Algorithmica 38(1), 227–248 (2003)
15. Mortara, M., Patanè, G., Spagnuolo, M., Falcidieno, B., Rossignac, J.: Plumber: a method for a multi-scale decomposition of 3d shapes into tubular primitives and bodies. In: SM 2004: Proceedings of the 9th ACM symposium on Solid modeling and applications, Genoa, Italy, pp. 339–344. Eurographics Association (2004)
16. Zhou, Y., Huang, Z.: Decomposing polygon meshes by means of critical points. In: MMM 2004: Proceedings of the 10th International Multimedia Modelling Conference, Washington, DC, USA, p. 187. IEEE Computer Society Press, Los Alamitos (2004)
17. Bespalov, D., Shokoufandeh, A., Regli, W.C., Sun, W.: Local feature extraction using scale-space decomposition. In: Computers und Information in Engineering Conference (DETC 2004-57702). ASME Design Engineering Technical Conferences. ASME Press (2004)
18. Attene, M., Robbiano, F., Spagnuolo, M., Falcidieno, B.: Semantic annotation of 3d surface meshes based on feature characterization. In: Falcidieno, B., Spagnuolo, M., Avrithis, Y., Kompatsiaris, I., Buitelaar, P. (eds.) SAMT 2007. LNCS, vol. 4816, pp. 126–139. Springer, Heidelberg (2007)

19. Hoffman, D., Richards, W.: Parts of recognition. In: Pinker, S. (ed.). Visual Cognition, pp. 65–96. MIT Press, Cambridge (1985)
20. Strothotte, T., Schlechtweg, S.: Non-Photorealistic Computer Graphics: Modeling, Rendering, and Animation. Morgan Kaufman Publisher, Los Altos (2002)
21. Diepstraten, J., Weiskopf, D., Ertl, T.: Transparency in interactive technical illustrations. Computer Graphics Forum 21(3), 317–325 (2002)
22. Bruckner, S., Groller, M.E.: Volumeshop: An interactive system for direct volume illustration. p. 85 (2005)
23. Diepstraten, J., Weiskopf, D., Ertl, T.: Interactive cutaway illustrations. Computer Graphics Forum 22(3), 523–532 (2003)
24. Götzelmann, T., Hartmann, K., Strothotte, T.: Contextual grouping of labels. In: Schulze, T., Horton, G., Preim, B., Schlechtweg, S. (eds.) 17th Conference on Simulation and Visualization, pp. 245–258. SCS Publishing House (2006)
25. Beckhaus, S., Ritter, F., Strothotte, T.: Guided exploration with dynamic potential fields: the cubicalpath system. Computer Graphics Forum 20(4), 201–210 (2001)
26. Ali, K., Hartmann, K., Strothotte, T.: Label Layout for Interactive 3D Illustrations. Journal of WSCG 13(1–3), 1–8 (2005)
27. Bier, E.A., Stone, M.C., Pier, K., Buxton, W., DeRose, T.D.: Toolglass and Magic Lenses: The See-Through Interface. Computer Graphics 27 (Annual Conference Series), 73–80 (1993)
28. Bier, E.A., Stone, M.C., Fishkin, K., Buxton, W., Baudel, T.: A taxonomy of see-through tools. In: Proceedings of CHI 1994 conference, pp. 358–364. Addison-Wesley, Reading (1994)
29. Ellis, G., Bertini, E., Dix, A.: The sampling lens: Making sense of saturated visualizations. In: Proceedings of CHI 2005, Portland, OR, USA, pp. 1351–1354. ACM Press, New York (2005)
30. Trapp, M., Glander, T., Buchholz, H., Döllner, J.: 3d generalization lenses for interactive focus + context visualization of virtual city models. In: Proceedings 12th International Conference Information Visualisation (IV 2008), London, UK, IEEE Computer Society Press, Los Alamitos (2008)
31. Viega, J., Conway, M.J., Williams, G., Pausch, R.: 3d magic lenses. In: UIST 1996: Proceedings of the 9th annual ACM symposium on User interface software and technology, pp. 51–58. ACM Press, New York (1996)
32. Looser, J., Billinghurst, M., Cockburn, A.: Through the looking glass: the use of lenses as an interface tool for augmented reality interfaces. In: GRAPHITE 2004: Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Australasia and South East Asia, pp. 204–211. ACM, New York (2004)
33. Götzelmann, T., Götze, M., Ali, K., Hartmann, K., Strothotte, T.: Annotating Images through Adaptation: An Integrated Text Authoring and Illustration Framework. Journal of WSCG 15(1–3), 115–122 (2007)
34. Alcocer, P.V.: On the Selection of Good Views and its Application to Computer Graphics. Phd thesis, University of Catalana, Barcelona, Spain (2004)
35. Fuchs, G.A., Luboschik, M., Hartmann, K., Ali, K., Strothotte, T., Schumann, H.: Adaptive labeling for interactive mobile information systems. IV, 453–459 (2006)

Dynamic Buffer Areas Obtained by EFCM Method in GIS Environment

Ferdinando Di Martino[1,2,*] and Salvatore Sessa[1]

[1] Università degli Studi di Napoli Federico II
Dipartimento di Costruzioni e Metodi Matematici in Architettura
Via Monteoliveto 3, 80134 Napoli, Italy
fdimarti@unina.it
[2] Università degli Studi di Salerno
Dipartimento di Matematica e Informatica,
Via Ponte Don Melillo, 84084 Fisciano (Salerno), Italy
Tel.: ++390812538907, Fax: ++390812538909
sessa@unina.it

Abstract. Impact areas in spatial analysis are determined by using density clustering methods which have an elevated computational complexity. We propose an alternative method based on the Extended Fuzzy C-Means (EFCM) and we derive dynamic buffer areas as hypersphere volume prototypes in a GIS environment.

Keywords: Extended fuzzy C-means, GIS, hotspot.

1 Introduction

A buffer area in spatial analysis is a zone around to features of a theme at a specified distance. Buffer areas are calculated in many fields of spatial analysis for determining bounded or dangerous zones: for example, areas around to the epicenter of an earthquake or areas of industrial pollution, etc. An hotspot can be defined as an area containing clusters of localized events and these areas are determined via clustering techniques (see [1] for an exhaustive list). The Extended Fuzzy C-Means (EFCM) is presented in [2]: the shape of the clusters are hyperspheres and here it is implemented in a tool GIS. The advantages of the EFCM method with respect to the classical FCM method consist essentially in the determination of the number of the clusters and the shape of every cluster volume prototype on the geographic map is circular and considered as hotspot area. In Sections 2 we describe the EFCM algorithm. In Section 3 we give an application of the EFCM algorithm in the specific problem of fire prevention of a forest area, located in New Mexico, by constructing circular areas which represent dangerous buffer areas.

2 The EFCM Algorithm

It is well known that the classical algorithm FCM presents two shortcomings:

* Corresponding author.

M. Sebillo, G. Vitiello, and G. Schaefer (Eds.): VISUAL 2008, LNCS 5188, pp. 92–95, 2008.

- the number C of the prototypes must be defined a priori or one calculates C as minimum or maximum of a suitable function [2],
- the cluster centers tend to locate in areas with high concentrations of features and the zones with low density data points could be relevant.

In order to overcome the above disadvantages, the EFCM was firstly proposed in [2] and [3]. There the volume prototypes are hyperellissoids but these become hyperspheres if one uses the Euclidean metric. If d_{ij} is the distance between the feacture vector $\mathbf{x_j} = (x_{1j}, x_{2j}, \ldots, x_{nj}) \in R^n$, with j=1,…,N, and the center $\mathbf{v_i} = (v_{1i}, v_{2i}, \ldots, v_{nj})^T$ of the i-th volume prototype V_i and if r_i is the radius of V_i, with i = 1,…,C, we say that x_j belongs to V_i if $d_{ij} \le r_i$. The covariance matrix $\mathbf{P_i}$ associated to the i-th cluster V_i is calculated as [3]:

$$\mathbf{P_i} = \frac{\sum_{j=1}^{N} u_{ij}^m (\mathbf{x_j} - \mathbf{v_i})(\mathbf{x_j} - \mathbf{v_i})^T}{\sum_{j=1}^{N} u_{ij}^m} \tag{1}$$

where m ≥1 is the fuzzifier parameter, u_{ij} is the degree of belongness of x_j to the i-th cluster V_i calculated as

$$u_{ij} = \frac{1}{\left(\sum_{k=1}^{c} \frac{d_{ij}^2}{d_{kj}^2} \right)^{\frac{2}{m-1}}} \tag{2}$$

The determinant of $\mathbf{P_i}$ gives the volume of the i-th cluster V_i. The objective function to be minimized is the following [10]:

$$J = \sum_{i=1}^{C} \sum_{j=1}^{N} u_{ij}^m (d_{ij}^2 - r_i^2) \tag{3}$$

The formulas (3) are update at each step of the algorithm (cfr. [3] for further details) and it works till to get $\|U^{(s)} - U^{(s-1)}\| < \varepsilon$, being ε >0 a prefixed parameter and $U^{(s)} = \left(u_{ij}^{(s)}\right)$ the matrix **U** of the membership degrees $u_{ij}^{(s)}$ calculated at the s-th step.

3 Tests with Hotspot Fire Events

The authors of [1] have implemented the EFCM method in a GIS environment created with the tool ESRI/ARCGIS. In spatial analysis, a circle on the geographic map, representing a prototype volume obtained with the usage of the EFCM algorithm, can be considered as a good approximation of an hotspot. Indeed the exact shape of the clusters could be determined by using the density based clustering method as in crime analysis (cfr., e.g., [4]), but this method is more expensive than EFCM in terms of

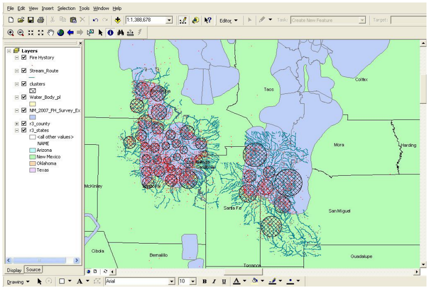


Fig. 1. Cluster volume prototypes for "Fire History" point data

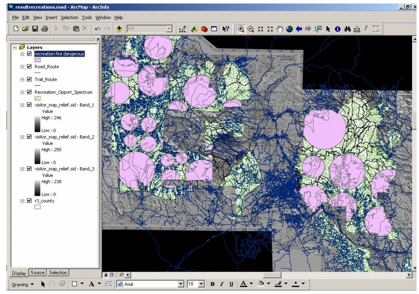


Fig. 2. Fire dangerous recreation opportunity areas

computational complexity. The patterns included in the i-th volume prototype have a membership value equal to 1 for the i-th cluster and 0 for the remaining clusters. Hence we can interpret a volume prototype as a geographic buffer area. Here we experiment the use of the EFCM algorithm on hotspot events considered as input data of the punctual features of the theme "Fire History" of USDA Santa Fe Service National Forest. The "Fire History" points represent the location at which fires began in the last three years. The latitude and longitude data was mapped to X,Y locations.

The data-test is made about of 5000 features distributed in the geographic area of New Mexico, mainly covered by the counties of Los Alamos, Sandoval, San Miguel, Mora, Rio Arriba and Santa Fè. The "Fire History" data are shown on the map of Figure 1, where are displayed also the stream routes, the water bodies, the base vegetation location and the surveyed forest areas in New Mexico. Our aim is to use the EFCM method to determine circular areas with high fire frequencies considered as fire dangerous buffer areas. In Figure 2 are shown the Fire History points, the Santa Fè National Forest visitor map revised in 2005, the roads, the trial routes and the recreation opportunity areas. We are interested to analyze fire dangerous buffer areas: indeed we study, for instance, their impact with the recreation opportunity areas. We have tested the use of the EFCM to determine these buffer areas varying the initial number of clusters, as in Table 1.

Table 1 proves that the EFCM method is quite stable since the final number of clusters is always 38. The circular volume prototypes are shown in Figure 1.

Table 1. Number of clusters and corresponding number of features

Initial number of clusters	Final number of clusters	Difference $\|U^{(s)} - U^{(s-1)}\|$
130	38	0.69×10^{-4}
120	38	0.71×10^{-4}
110	38	0.67×10^{-4}
100	38	0.85×10^{-4}
90	38	0.53×10^{-4}
80	38	0.49×10^{-4}
70	38	0.66×10^{-4}
60	38	0.88×10^{-4}
50	38	0.72×10^{-4}

In Figure 2 we show the intersection of the recreation opportunity areas with the cluster prototypes. This figure shows that we can use the cluster volume prototypes derived using the EFCM method to determine in a GIS risk and impact areas based on point event data. The above experiments have pointed out that we can use the EFCM clustering method in spatial analysis for determining dynamically circular buffer areas. Over the geographic map these areas can be considered as a good approximation of a classical hotspot.

4 Conclusions

In spatial analysis usually impact areas are determined by using density clustering methods which have an elevated computational complexity. Here we have suggested the EFCM method because of the following advantages: robustness to noise and outliers, linear computational complexity and automatic determination of the optimal number of clusters. In this method we derive dynamic buffer areas as hypersphere volume prototypes which become circles in the case of bi-dimensional pattern data, like in the case of event point data in the presented GIS application, devoted to the analysis of hotspot fire events in the Santa Fè district (NM), downloaded from web pages at the URL http://www.fs.fed.us/r3/gis/sfe_gis.shtml.

References

1. Di Martino, F., Loia, V., Sessa, S.: Extended Fuzzy C-Means Clustering Algorithm for Hot Spot Events in Spatial Analysis. Internat. Jour. of Hybrid Intelligent System 4, 1–14 (2007)
2. Kaymak, U., Babuska, R., Setnes, M., Verbruggen, H.B., van Nauta Lemke, H.M.: Methods for Simplification of Fuzzy Models. In: Ruan, D. (ed.) Intelligent Hybrid Systems, pp. 91–108. Kluwer Academic Publishers, Dordrecht (1997)
3. Kaymak, U., Setnes, M.: Fuzzy Clustering with Volume Prototype and Adaptive Cluster Merging. IEEE Transactions on Fuzzy Systems 10(6), 705–712 (2002)
4. Murray, A.T., McGuffog, I., Western, J.S., Mullins, P.: Exploratory Spatial Data Analysis Techniques for Examining Urban Crime. British Journal of Criminology 41, 309–329 (2001)

Automatic G^1 Surface Reconstruction from Serial Cross-Sectional Images

Fatimah Yahya[1], Jamaludin Md Ali[2], Ahmad Abdul Majid[2], and Arsmah Ibrahim[1]

[1] FTMSK, Universiti Teknologi Mara, Shah Alam, Selangor, Malaysia
[2] Pengajian Sains Matematik, Universiti Sains Malaysia, P. Pinang, Malaysia
fatimahy2003@yahoo.co.uk, jamaluma@cs.usm.my, majid@cs.usm.my, arsmah@tmsk.uitm.edu.my

Abstract. Biomedical imaging facilities today like MRI, PET, CT scan and confocal microscopy produce sequentially parallel cross sectional images. Reconstructing trustworthy 3D explicit models which enable better understanding of the topology and shape of structure is crucial in facilitating diagnosis, improves surgical planning and aid in biological research.

Our technique produce a surface from G^1 cross sectional contour curves of images. Surface accuracy is controlled by a tolerance measure. Features can be isolated and identified for corresspondence.

First the boundaries of the region of interest are extracted and corner points detected. G^1 rational Bezier cubics, iteratively determined, are fitted piecewise between these corners and approximating the boundary. as close as need be. Adjacent contour curves are blended together to form the surface. Technique is fully automatic.

Keywords: Surface reconstruction, corner detection, serial cross-sectional contours, arc length, curve fitting, reparameterization.

1 Introduction

There are two ways in creating 3D models from 2D images; volume visualizations and polygonal modelling. In the polygonal modelling approach, the outlines of structure of interest are obtained at each slices. These outlines or contours are stacked and a polygonal mesh is created between sucessive contours by connecting vertices (points) on one contour to vertices on the other. Meshes are either triangular or rectangular. This gives rise to the problem of corresspondence- how to connect vertices between contours, and tiling – how to create meshes from edges.Fuchs [1], Bajaj et al [2], Stewart and McCracken [3], Fujimura and Kuo [4] and Chen et al [5] are examples that address the problem of reconstructing surfaces from contours.

Most existing approaches to shape reconstruction yield C^0 or C^1 triangular surfaces which may give rise to a visual artifact in the reconstructed shapes due to discontinuity in surface normal. Such discontinuity is less noticeable for rectangular tiling or free

M. Sebillo, G. Vitiello, and G. Schaefer (Eds.): VISUAL 2008, LNCS 5188, pp. 96–99, 2008.

form surfaces (Fujimura and Kuo [4]). Triangular tilings which is piecewise linear also makes it difficult to incorporate feature corresspondence.

Our technique produce a surface from G^1 cross sectional contour curves of images. Surface accuracy is controlled by a tolerance measure. A low tolerance produces finer, more accurate images; albeit with more computations. Features can be identified from the significant points data and isolated. Thus parts of the whole surface can be produced separately; for example the nose or ears from a face. Technique is fully automatic except for determining the first boundary point of the first cross-section.

The surface reconstruction involves several steps. First, CT images are digitized and boundary of the region concerned obtained. Corner points of the boundary is then detected and curves fitted to these points. Here the problem of corresspondence and tiling is overcome by blending contour curves through their relative length producing rectangular meshes of the surface. All programming and rendering done in Mathematica while Slicer-3d is used to access the scanned images and for region thresholding.

2 Corner Detection

Here we use a method that utilize region of support with ideas of eigenvalues of covariance matrix as a way of measuring corners. The ratio; $\frac{\lambda_s}{\lambda_s + \lambda_l}$ is used as a measure of curvature as it is better in dealing with differences in magnitudes of data and sizes of supports. λ_1 is the larger eigenvalue of the covariance matrix. Corner points are the data points whose eigenvalue ratios are local maximums. See [8].

3 Curve Fitting

G^1 rational cubic Bezier curves interpolate corner points and approximate the other boundary data points. The curves are made close enough to the boundary data point as desired by adjusting the weights of the rational cubic. Automatic adjustments and determination of these weights is obtained by optimizing at each iterating reparameterization.

The iterative process may not end with the desired curve i.e. within tolerance. The limits of the iterative process is reached when further iteration does not lower the maximum distance of curve to digitized data points. When this happens and tolerance has not been satisfied, the curve is then split into two, and the fitting process repeated for the two new segments obtained. This process is repeated until tolerance is achieved. The construction of these curves have been discussed in Yahya et al. [8].

4 Arc Length Reparameterization

In blending curves to form surfaces, corresponding points on adjacent contour curves need to be found. This is done by matching length of the contours. Pairs of contours are matched up by its relative length. Thus the contours need to be arc length parameterized. A fast technique using cubic Bezier; based on Walter & Fournier's [6] method; is used to reparameterize the contour curves as closed functions of arc length. The technique here, is the inverse to that presented by Walter & Fournier [6].

5 Contour Blending

Contours interpolated using cubic Hermite polynomials with tangents to the surface as introduced by Goodman,Ong and Unsworth [7].

6 Implementation, Results and Discussion

A set of 120 CT scanned dicom images of the head at 1.5 mm apart was obtained. Fig. 1 shows some CT scanned cross-sectional images of a head. Digital images are obtained

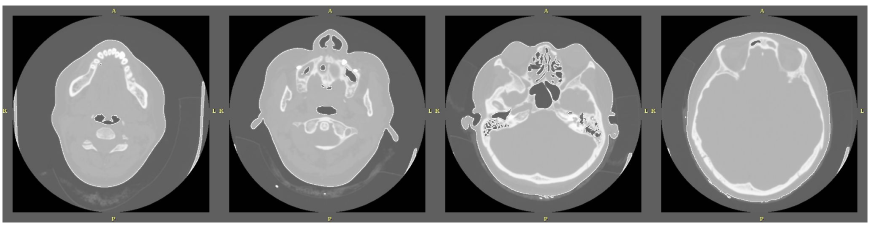


Fig. 1. Some of the CT scan images of the whole head

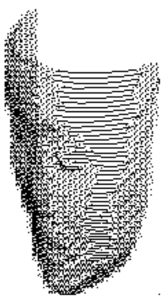


Fig. 2. Stacked contours

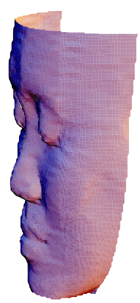


Fig. 3. Reconstructed face

Fig. 4. Nose- side view

Fig. 5. Nose- front-right view

from the scanned slices. Preprocessing is done and the resulting digital image is then converted into binary image by thresholding. A closed contour of one pixel thick is extracted from the binary image.

Boundary extraction, corner detection, curve fitting, contour interpolating and rendering done in Mathematica. Fig. 2 shows the stacked contours while Fig. 3 shows the reconstructed face. Fig.4 and 5 shows an isolated nose reconstructed on its own (not taken from the face).

The solid reconstruction of the face is good with clear representation of the face features. However it can be seen that it is not so smooth between pieces. This area will be the focus of our future work.

Although this paper only present examples of open surfaces made up of open contours, the technique is of course applicable to closed contours with little adjustments to the programming.

Diverse shape and length between adjacent contours can cause distortion to the mesh which is carried on to the surface. This is due to the corresspondence being no longer right. It can be corrected by realligning the contours by their corner points and the subsequent readjustment of arc length reparameterization. However, we do not face such problems in producing the face here (Fig. 3) or the nose (Fig. 4 &5).

References

1. Fuchs, H., Kedem, Z.M., Uselton, S.P.: Optimal surface reconstructionfrom planar contours. Commun. ACM 20, 693–702 (1977)
2. Bajaj, C.L., Coyle, E.J., Lin., K.-N.: Arbitrary topology shape reconstruction from planar cross sections. Graph Model Image Process 58, 524–543 (1996)
3. Stewart, C.-H., McCracken, T.: A system for 3D surface models from 2D sectional Imagery. In: International Conference on Mathematics and Engineering Technique. Medicine and Biological Sciences (2002)
4. Fujimura, K., Kuo, E.: Shape reconstruction from contours using isotopic deformation. Graph Model. Image Process 61, 127–147 (1999)
5. Chen, Y.-C., Chiang, A.-S., Hsieh, K.-S.: A reliable surface reconstruction system in biomedicine. Computer Methods and Programs in Biomedicine 86, 141–152 (2007)
6. Walter, M., Fournier, A.: Approximate Arc Length Parameterization. In: Proceedings of the 9th Brazilian Symposium on Computer Graphics and Image Processing, pp. 143–150 (1996)
7. Goodman, T.N.T., Ong., B.H., Unsworth, K.: A new algorithm for the reconstruction of 3D objects from cross-sections. In: Proc. Of the ASEE 5th Inter. Conf. on Engineering Computer Graphics and Descriptive Geometry, Australia, pp. 546–551 (1992)
8. Yahya, F., Ali, J., Majid, A., Ibrahim, A.: An automatic generation of G1 curve fitting of Arabic characters. In: Proc. Inter. Conf. on Comp. Graph., Imag. & Visual. CGIV 2006, Australia, pp. 542–547. IEEE Computer Society, Los Alamitos (2006)

Operational Risk Visualization

Concetto Bonafede[1] and Roberto Marmo[2]

[1] University of Pavia, Dep. of Statistics "L. Lenti", 27100 Pavia, Italy
concetto.bonafede@unipv.it
http://www.unipv.it/dipstea
[2] University of Pavia, Faculty of Engineering, Laboratory of Computer Vision, 27100 Pavia, Italy
roberto.marmo@unipv.it
http://vision.unipv.it

Abstract. Operational risk management is a new paradigm for decision making, that involves the two-way exchange of information between interested parties in order to manage risks according to strategic background. The database used to develop our graph comes from a telecommunication service company which provides custom communication equipments to client. We provide graphical representations useful to decide pro-actively if a new client is a potential source of some kind of problems that lead to growth of some risks. We study the usability of graphical formats using eye tracker.

1 Introduction

Risk analysis has been considered as part of the process of planning a technological system and addressed the risk inherent in its day-to-day operations [1]. The risk is measured in terms of a combination of two variables concerning two different aspects of an harmful event: Frequency and Impact. The context is either the Operational Risk Management (ORM) or the Business Continuity Management (BCM) [2] [3] [4]. The ORM framework handle with harmful events due to People, System, Process, External Event (such as flood).

The use of graphs helps visualisation and cognition by grouping information that is used together, so reducing search times [5]. Risk visualization involves the two-way exchange of information between interested parties in order to make decisions about how best to manage risks [6]. We provide graphical representations useful to decide pro-actively if a new client is a potential source of some kind of problems that lead to growth of some risks. These risks could be due to either System or Process or others (as in ORM) from which service interruptions are generated (as in BCM).

2 Data Description

We have used a database coming from a telecommunication company, related to PBX (Private Branch Exchange). Source of service interruption risk are due

M. Sebillo, G. Vitiello, and G. Schaefer (Eds.): VISUAL 2008, LNCS 5188, pp. 100–103, 2008.

to Hardware, Software, Network, Interface and Security happen to PBX. Users share outside lines for making telephone calls external to the PBX. The information available are about:

- The type of customer as Health for hospital and similar, Banking for bank, Finance for credit institute and so on;
- The type of new client within the categories of customer type;
- The type of problems and date, that lead to an interruption;
- The severity (categorized in three levels) that describe the magnitude of the interruption in function of the problem and customer;
- Some other information related to the clients as number of smart phone, number of lines, etc.

3 Description of Visualizations

We plot information about Banking as Severity Customer Type, but this approach can be generalized to the other clusters of information: Banking, Computers, Construction, Cooperatives, Defense, Education, Elderly citizens home, Electronics, Finance, General, Government, Government owned company, Health, Hotels, Industry, Lawyer accountant offices, Leisure Consumerism, Municipalities, Operating company, Transportation.

We propose 4 sub-plots and a colormap. In first group we draw two markers on the same row instead of two bars on different row, a sort of dot plot [5]. The circle as marker denotes information about the New_client and the x-mark denotes information about the Old_client, in each case the value is between 0 and 1. Each sub-plot is divided on 5 rows, from the top: Hardware (black color of signs), Interface (magenta), Network (ciano), Security (brown), Software (blue). The red bar between circle and x-mark highlights a difference major than 0.2, so the Old_client is more suitable compared to the New_client. The yellow bar between circle and x-mark highlights a difference major than 0.2, so the New_client is more suitable compared to the Old_client. Note that in case of x-mark inside a circle we visually pop-out that the choice is the same because the values are similars. The last sub-plot is the parallel coordinates plot that deals with high dimensional number of feature [5] [7].

Each observation is represented by sequence of coordinate values plotted against their coordinate indices, so on x-axis there are vertical lines representing each feature about High, Mean and Low and on y-axis there are the value for related feature between -1 and 1. The color of each line is related to color corresponding to one information on the row of other sub-plots. It is possible to describe the trend of values corresponding to difference between New_client and Old_client. In case of value under 0 level the related information on the New_client is more suitable compared to the Old_client, in other case the related information on the Old_client is more suitable compared to the New_client. On right side a colormap help the visual comparison between Old_client and New_client: first case above mentioned corresponding to green color that is a sort of good choice, second case corresponding to yellow color that is a sort of

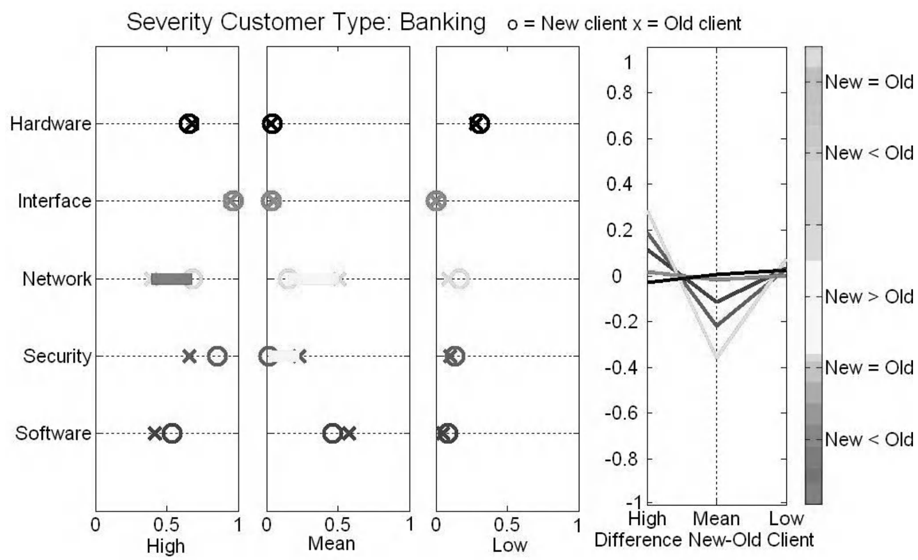


Fig. 1. Proposed visualization composed of 3 sub-plot and on right hand a visual digest

doubt or not relevance choice, third case corresponding to red color that is a sort of bad choice. It is possible to read directly the fourth graph to obtain a visual digest, a summary of relevant information creating a focus useful for decision making, then other graphs provide details and comparison for each information. We arrange all the graphs for all the clusters of information composing an informational dashboard [8] to display a dense array of information in a small amount of space in a manner that communicates clearly and immediately.

4 Usability Design and Discussion

Eye tracking can help us to understand usability to reduce the number of operations performed [9] [10]. We use the Tobii 1750 which integrates camera and infrared lighting into a monitor [11] to obtain the path followed by eyes when looking at something composed by saccade-fixation-saccade sequence of eye movement, the scanpath represented by circles with radii proportional to fixation times, and to get an indication of those areas of the screen which were most watched, that is the heatmap [12] [13]. We tracked the eye movements of five participants whilst they carried out tasks.

As usability measures the following performance data was collected:

1. Time taken and error rate to answer the question;
2. Number of fixations and average total fixation duration;
3. Scan path length for the duration;
4. Participants' rating of graph usability.

The question regards: yes or no about acquisition of the new client. There are no wrong answers, because there is a redundant visualization of relevant information, on three sub-plots and summary on right side. Rating of graph usability on range [1,4] where 1 denotes low rating and 4 denotes high rating, average: 3. When volunteer look as first time, it is necessary to take confidence with each visual object. The fixation path of the user on the graph illustrates that the user glanced at the legend and the title upon first seeing the graph and then read the text and each subplot area, requiring more time to learning. But on further visualization, volunteer requires low time to read the graph, just to analyze only each subplot area.

Acknowledgments

The authors acknowledge financial support from national Italian grant MIUR-FIRB 2006-2009 and European grant EU-IP MUSING (contract number 027097). We also remark the whole MUSING team in Pavia, for useful comments.

References

1. Beroggi, G.E., Wallace, W.A.: Operational Risk Management: A New Paradigm for Decision Making. IEEE Transactions on Systems, Man and Cybernetics 24(10), 1450–1457 (1994)
2. Bonafede, C., Cerchiello, P., Giudici, P.: Statistical models for business continuity management. Journal of Operational Risk 2(4), 79–96 (2007)
3. COSO: Enterprise risk management integrated framework, Executive summary. Committee of Sponsoring Organization of the Treadway Commission (2004), `http://www.coso.org`
4. PAS56: Guide to business continuity management (2003)
5. Spence, R.: Information Visualization. Addison-Wesley, Reading (2001)
6. Lipkus, I.M., Hollands, J.G.: The Visual Communication of Risk. Journal of the National Cancer Institute Monographs 25, 149–163 (1999)
7. Fayyad, U., Wierse, A., Grinstein, G.: Information visualization in data mining and knowledge. Morgan Kaufmann, San Francisco (2005)
8. Few, S.: Information dashboard design: the effective visual communication of data. O'Reilly, Sebastopol (2006)
9. Renshaw, J.A., Finlay, J.E., Tyfa, D., Ward, R.D.: Designing for Visual Influence: an Eye Tracking Study of the Usability of Graphical Management Information. In: Rauterberg, et al. (eds.) Human-Computer Interaction IFIP, pp. 144–151 (2003)
10. Milekic, S.: The more you look the more you get: intention-based interface using gaze tracking. In: Proceedings of 7 annual museum and the web conference, North Carolina, March 19-22 (2003)
11. Tobii Technology, A.B.: Tobii 1750 Eye tracker Release B (2003)
12. Duchowski, A.T.: Eye Tracking Methodology: Theory and Practice. Springer, Berlin (2003)
13. Tai, R.H., Loeher, J.F., Brigham, F.J.: Introduction to the use of eye-gaze tracking to study participant behavior during problem-solving. In: Narst (ed.) AnnualMeeting of the National Association for Research in Science Teaching, NARST, Vancouver, BC (2004)

Classification-Based Color Constancy

Simone Bianco[1], Gianluigi Ciocca[1], Claudio Cusano[1], and Raimondo Schettini[1]

DISCo (Dipartimento di Informatica, Sistemistica e Comunicazione),
Università degli Studi di Milano-Bicocca, Viale Sarca 336, 20126 Milano, Italy
{bianco,ciocca,cusano,schettini}@disco.unimib.it

Abstract. In this work, we investigate how illuminant estimation techniques can be improved using indoor/outdoor classification. The illuminant estimation algorithms considered are derived from the framework recently proposed by Van de Weijer and Gevers. We have designed a strategy for the selection of the most appropriate algorithm on the basis of the classification results. We have tested the proposed strategy on a subset of the widely used Funt and Ciurea dataset. Experimental results clearly demonstrate that our strategy outperforms general purpose algorithms.

1 Introduction

Computational color constancy aims to estimate the actual color in an acquired scene disregarding its illuminant. The different approaches can be broadly classified into color invariant and illuminant estimation [1]. The former approaches derive from the image data invariant color descriptors without explicitly estimating the scene illuminant. The latter is actually a two stage procedure: the scene illuminant is estimated from the image data, and the image colors are then corrected on the basis of this estimate to generate a new image of the scene as if it were taken under a known, canonical illuminant. Many illuminant estimation solutions have been proposed in the last few years although it is known that the problem addressed is actually ill-posed as its solution lacks uniqueness or stability. To cope with this problem different solutions usually exploit some assumptions about the statistical properties of the expected illuminants and/or of the object reflectances in the scene.

An excellent review of illuminant estimation algorithms has been given by Hordley [1]. To further improve illuminant estimation several promising studies investigated the idea of using high-level information: Gasparini and Schettini [2] applied an adaptive mixture of the white balance and gray world procedures. Van de Weijer et al. [3] selected the best illuminant out of a set of possible ones by segmenting the images into regions corresponding to a set of semantic classes. Gijsenij and Gevers [4] used natural image statistics to identify the most important characteristics of color images and to achieve the selection and/or combination of color constancy algorithms.

The idea investigated here is that the effectiveness of illuminant estimation techniques may be improved if the images are automatically assigned to semantic

M. Sebillo, G. Vitiello, and G. Schaefer (Eds.): VISUAL 2008, LNCS 5188, pp. 104–113, 2008.

cathegories by an image classifier. We considered the indoor and outdoor classes, which correspond to categories of images usually taken under different illumination conditions. Therefore, these two classes of images may require different color constancy techniques.

Let us suppose to have several distinct color constancy algorithms; based on their performances on a training set that is manually labeled into indoor and outdoor classes, it is possible to identify the best algorithm for the two classes considered and the best algorithm for the whole training set. Given an input image a decision forest trained to discriminate between indoor and outdoor images is used to select the most appropriate illuminant estimation algorithm.

In our experiments we used the dataset presented by Ciurea and Funt [5]. Since the images in this dataset were frames extracted from video clips, and they show high correlation between each other, we applied a video-based analysis to automatically select the images to be included in our data set. Experimental results show that the accuracy of color constancy techniques can be significantly improved if specific algorithms are chosen for each image class.

2 Computational Color Constancy

An image acquired by a digital camera can be seen as a function $\boldsymbol{\rho}$ which is mainly dependent on three physical factors: the illuminant spectral power distribution $I(\lambda)$, the surface spectral reflectance $S(\lambda)$ and the sensor spectral sensitivities $\mathbf{C}(\lambda)$. Using this notation, the sensor responses at the pixel with coordinates (x, y) can be thus described as:

$$\boldsymbol{\rho}(x, y) = \int_{\omega} I(\lambda) S(x, y, \lambda) \mathbf{C}(\lambda) \mathrm{d}\lambda \ , \tag{1}$$

where ω is the wavelength range of the visible light spectrum, $\boldsymbol{\rho}$ and $\mathbf{C}(\lambda)$ are three-component vectors. Since the three sensor spectral sensitivities are usually respectively more sensitive to the low, medium and high wavelengths, the three-component vector of sensor responses $\boldsymbol{\rho} = (\rho_1, \rho_2, \rho_3)$ is also referred to as the sensor or camera $\mathbf{RGB} = (R, G, B)$ triplet. Assuming that the color $\mathbf{I}$ of the scene illuminant as seen by the camera only depends on the illuminant spectral power distribution $I(\lambda)$ and on the sensor spectral sensitivities $\mathbf{C}(\lambda)$, then computational color constancy is equivalent to the estimation of $\mathbf{I}$ by:

$$\mathbf{I} = \int_{\omega} I(\lambda) \mathbf{C}(\lambda) \mathrm{d}\lambda \ , \tag{2}$$

given only the sensor responses $\boldsymbol{\rho}(x, y)$ across the image. It has been demonstrated that this is an under-determined problem [6], and thus further assumptions and/or knowledge is needed to solve it. Typically some information about the camera being used are exploited, and/or assumptions about the statistical properties of the expected illuminants and surface reflectances.

Several computational color constancy algorithms exist in the literature, each based on different assumptions. In this work, we chose six algorithms, but a different set of algorithms could be used.

Recently Van de Weijer et al. [7] proposed a framework which unifies a variety of algorithms. These algorithms correspond to instantiations of the following equation:

$$\left(\iint \left| \nabla^n \boldsymbol{\rho}_\sigma(x,y) \right|^p \mathrm{d}x\mathrm{d}y \right)^{\frac{1}{p}} = k\mathbf{I} \ , \tag{3}$$

where n is the order of the derivative, p is the Minkowski norm, $\boldsymbol{\rho}_\sigma(x,y) = \boldsymbol{\rho}(x,y) \otimes G_\sigma(x,y)$ is the convolution of the image with a Gaussian filter $G_\sigma(x,y)$ with scale parameter σ, and k is a constant to be chosen so that the illuminant color I has unit length. In this work, by varying the three variables (n,p,σ) we have generated six algorithm instantiations that correspond to well known and widely used color constancy algorithms: Gray World (GW) [8], White Point (WP) [9] (also known as Maximum RGB), Shades of Gray (SG) [10], General Gray World (gGW) [11,7], Gray Edge (GE1) [7] and Second Order Gray Edge (GE2) [7]. The algorithms are reported in Table 1 together with the parameters settings that have to be made into (3) to generate them and the assumptions on which they are based. As can be noticed, the instantiations of GW and WP have all three parameters (n,p,σ) fixed; SG instead, has the parameter p that can be opportunely tuned for a particular image; while gGW, GE1 and GE2 have two parameters (p and σ) which must be tuned.

For the tuning of the parameters and for performance evaluation of the algorithms we used the dataset of images presented by Ciurea and Funt [5]. The dataset is commonly used in the evaluation of color constancy algorithms as it is labeled with the ground truth illuminants. In this dataset, 15 digital video clips were recorded (at 15 frames per second) in different settings such as indoor, outdoor, desert, markets, cityscape, etc... for a total of two hours of videos.

Table 1. Overview of the different computational color constancy algorithms together with the assumptions on which they are based

Algorithm name	Parameters (n,p,σ)	Assumptions
Gray World (GW)	$(0,1,0)$	the average reflectance in a scene is achromatic
White Point (WP)	$(0,\infty,0)$	the maximum reflectance in a scene is achromatic
Shades of Gray (SG)	$(0,p,0)$	the $p-$th Minkowski norm of a scene is achromatic
General Gray World (gGW)	$(0,p,\sigma)$	the $p-$th Minkowski norm of a scene after local smoothing is achromatic
Gray Edge (GE1)	$(1,p,\sigma)$	the $p-$th Minkowski norm of the first order derivative in a scene is achromatic
Second Order Gray Edge (GE2)	$(2,p,\sigma)$	the $p-$th Minkowski norm of the second order derivative in a scene is achromatic

From each clip, a set of images was extracted resulting in a dataset of more than 11,000 images. A gray sphere appears in the bottom right corner of the images and was used to estimate the true color of the scene illuminant. Since the dataset sources were video clips, the images extracted show high correlation. To remove this correlation only a subset of images should be used from each set. Taking into account that the image sets came from video clips, we applied a two stage video-based analysis to select the image to be included in the final illuminant dataset.

In the first stage, a video clip is reconstructed from each set of images removing the right part of the images containing the gray sphere. The video clip is fed to a key frame extraction algorithm [12] which dynamically selects a set of candidate images by analyzing the visual contents of consecutive frames. Clips showing high variability in their pictorial contents will have a high number of images extracted while clips showing little or no variability will have only a single image extracted.

As a trade-off between the number of images to be included in the dataset and the correlation problem, we set the parameters of the key frame extraction algorithm so that the images extracted correspond to at least 10% of the clip size. The final subset consists of 1,135 images.

In the second stage, we further processed the extracted images with a visual summary post-processing algorithm [13]. For this work, we exploited only the key frame grouping processing step that eliminates pictorially similar images, using a hierarchical clustering algorithm.

The clustering algorithm further removes redundancies within the set of images. At each step one image is removed from the set and the clustering process stops when the number of remaining images is exactly 10% of the clip size.

To evaluate the performance of the algorithms on the dataset, we have used an intensity independent error measure. As suggested by Hordley and Finlayson [14], we use the angle between the RGB triplets of the illuminant color ($\boldsymbol{\rho}_w$) and the algorithm's estimate of it ($\hat{\boldsymbol{\rho}_w}$) as error measure:

$$e_{ANG} = \arccos\left(\frac{\boldsymbol{\rho}_w^T \hat{\boldsymbol{\rho}_w}}{\|\boldsymbol{\rho}_w\| \|\hat{\boldsymbol{\rho}_w}\|}\right) \quad . \tag{4}$$

It has been shown also [14] that the median error is a good descriptor of the angular error distribution.

Four of the color constancy algorithms considered (SG, gGW, GE1, GE2), needed a training phase to opportunely tune the parameters (n, p, σ). We use the median error as a cost function for the choice of the best parameters. Being the median error a non-linear statistic, we needed a multidimensional non-linear optimization algorithm: our choice was to use a Pattern Search Method (PSM). PSMs are a class of direct search methods for nonlinear optimization [15,16].

We need a training set on which to perform the tuning of the parameters. For this, 300 images (150 indoor and 150 outdoor) were randomly extracted from the 1,135 images of the illuminant dataset and used as training set. Since we expect different behavior of the algorithms on the different classes considered, the parameters were tuned independently for the indoor class, for the outdoor class and for the whole training set.

Table 2. The parameters found by the pattern search algorithm. Only the values reported in bold have been computed, the others have been set according to the definition of the algorithms.

	Indoor			Outdoor			General purpose		
	n	p	σ	n	p	σ	n	p	σ
GW	0	1	0	0	1	0	0	1	0
WP	0	∞	0	0	∞	0	0	∞	0
SG	0	**1.27**	0	0	∞^*	0	0	**1.06**	0
gGW	0	**1.32**	**1.00**	0	∞^*	**0.00**	0	**1.08**	**0.83**
GE1	1	**0.60**	**1.72**	1	**1.10**	**0.83**	1	**1.10**	**1.08**
GE2	2	**1.06**	**2.96**	2	**1.91**	**0.04**	2	**1.55**	**1.83**

* Values which diverge towards infinity.

The parameters found by the pattern search algorithm are reported in Table 2. It can be seen that the optimal values found for the parameter p for the indoor class are lower than the ones for the outdoor class. The optimal values found for the parameter σ instead, are higher for the indoor class than the ones for the outdoor class. While for the indoor class the pattern search optimization found different values for the parameters of the different color constancy algorithms, we can see that in the choice of the parameters for the outdoor class of the gGW and SG algorithms, they tended to be asymptotically convergent to the WP's ones. Regarding the general-purpose class, we can see instead that the optimal choice for the parameters of the SG algorithm is very similar to the GW's ones.

3 Image Classification

The key idea investigated here is that illuminant estimation may be improved if additional information about the acquisition conditions is known. For this we designed a classifier which is able to automatically distinguish between indoor and outdoor images. We considered these two classes of images because they usually correspond to different illumination conditions.

The problem of indoor/outdoor classification has been already addressed in previous works (see, for example, Szummer and Picard [17], Vailaya et al. [18], and Schettini et al. [19]). As a classifier we used a decision forest [20] composed of several classification trees built accordind to the CART methodology [21]. The trees are generated using a supervised learning process which requires a training set of images manually labeled as indoor or outdoor. First, the images are described by a set of low-level features, that is, features which can be automatically computed without any a-priori knowledge about the content. The features we considered are:

- The first and second statistical moments of the components in the YCbCr color space, computed on the pixels of the images divided into seven horizontal sub-bands;

– A 27 bin histogram of the color of the pixels in the RGB color space;
– A 18 bin histogram of the directions of the edges of the images;
– The average absolute value and the standard deviation of the coefficients of a 3-level wavelets transform which yields to a subdivision in ten sub-bands.

Each image is thus described by a feature vector of 107 components.

Each classification tree of the forest is produced by a recursive procedure where each node corresponds to a subset of the training set (the root of the tree corresponds to the whole training set). At each step a node of the tree is divided in two subnodes by a split, that is, a binary partition of the subset associated to that node consisting in the comparison of the value of a low-level feature with a threshold. Among all possible splits (i.e. all possible combination of features and thresholds) the one which minimizes the average impurity of the two sub-nodes is selected. Impurity is meausured by a diversity function; here we used the Gini diversity index:

$$i(t) = 2p(indoor|t)p(outdoor|t) \ , \tag{5}$$

where $p(\omega|t)$ is the resubstitution estimate of the conditional probability of class ω in node t, that is, the probability that a case found in node t is a case of class ω. When the impurity cannot be further improved by a split, the node is considered as terminal.

When the tree growing process is finished, terminal nodes are labeled as indoor or outdoor:

$$label(t) = \begin{cases} indoor & \text{if} \quad p(indoor|t) \geq p(outdoor|t) \\ outdoor & \text{otherwise} \end{cases} \ . \tag{6}$$

The growing process usually produces big trees which overfit the training data. A pruning process increases the generalization capabilities of the trees by reducing their dimension. Pruning is based on a cost-complexity criterion: given a tree T, its performance is measured by:

$$R_\alpha(T) = R(T) + \alpha|T| \ , \tag{7}$$

where $R(T)$ is the probability of misclassification estimated on the training set, $|T|$ is the size of the tree (the number of terminal nodes), and α is a parameter which weights prediction errors and complexity of the tree. Starting from the original tree T_0 (corresponding to $\alpha = 0$) and increasing the value of α the sequence of best subtrees $T_0, T_1, \ldots, T_M$ is considered. Each subtree corresponds to the optimal subtree of T_0 for a range of values of the parameter α. The best subtree of the sequence is selected on the basis of the errors made on a validation set.

The decision forest is built by producing several bootstrap replicates of the training set. Each variation of the training set is used to grow an inital tree, and the cases not selected by the bootstrap procedure are used in the pruning process as a validation set.

For the training of the decision forest we collected 6,785 images, downloaded from the web, or acquired by a scanner or digital cameras. The images have been

Table 3. Confusion matrix obtained on the test set by the decision forest. The number of misclassifications was 169 (86 indoor and 83 outdoor images).

	Predicted indoor	Predicted outdoor
True indoor	82.1%	17.9%
True outdoor	12.7%	87.3%

manually annotated yielding to 2,105 indoor images and 4,680 outdoor images. No enhancement procedure (such as white balancing) has been applied to the images. We partitioned the images into a training set of 2,000 images (1,000 indoor and 1,000 outdoor), and a validation set containing the remaining 4,785 images. After a model selection phase we selected the four features described above and we set the size of the decision forest to 50 trees.

Within this framework we obtained a classification accuracy of about 85.1% on our subset of the Ciurea-Funt image dataset. The results are summarized in Table 3 as a confusion matrix.

Better results could probably be obtained if homogeneous datasets are used for both training and test. Most of the misclassification errors correspond to images with little information about the context in which they were taken. For instance, there are several close-ups of various objects, which cannot be classified as indoor or outdoor without exploiting high level knowledge and reasoning.

4 Experimental Results

To validate our strategy we compared its performance with the results obtained by general purpose algorithms. In order to compare the whole error distribution between different algorithms, together with the median angular error, we used the Wilcoxon Sign Test (WST) [22]. A score is generated by counting the number of times that the performance of a given algorithm has been considered to be better than the others.

The best algorithms for each class was chosen by assessing their performance on the training set introduced in Sect. 2. The results are reported in Table 4. More in detail, the three groups of columns refer to the results obtained on indoor images, outdoor images, and on the whole training set.

On the basis of the results obtained we can select the best algorithm for each class. For indoor images both SG and gGW may be chosen. In fact, the WST score shows that they exhibit an indistinguishable behavior. We selected the SG because its angular error is slightly better. For outdoor images the GE2 algorithm is clearly superior to the others. According to the WST scores, on the whole training set (without performing any classification) there is no algorithm which significantly outperforms the others.

We can see that the errors on the two classes are very different. In particular, with the exception of the GW algorithm, all the errors obtained on the outdoor

Table 4. Median angular error obtained by the color constancy algorithms on the training set. The best results for each column are reported in bold.

	Indoor Images†		Outdoor Images†		Whole Training Set	
	Median	WSTs	Median	WSTs	Median	WSTs
GW*	4.91	3	7.86	0	5.62	**1**
WP*	11.83	0	2.81	2	7.76	0
SG	**4.31**	**4**	2.81	2	5.56	**1**
gGW	4.32	**4**	2.81	2	5.57	**1**
GE1	5.40	1	3.72	1	**5.45**	**1**
GE2	5.57	1	**2.48**	**5**	5.47	**1**

* Algorithms with fixed, class independent, parameters.
† Test with algorithms tuned specifically for the class.

Table 5. Median angular error obtained by the color constancy algorithms on the test set. The best results for each column are reported in bold.

Method	Median	WSTs
GW	5.95	0
WP	5.48	3
SG	5.80	0
gGW	5.80	0
GE1	4.47	4
GE2	4.65	4
Classification-based	**3.78**	**6**

class are significantly lower than those obtained on the indoor class. This can be explained by analyzing the dataset. The majority of the outdoor images were shot under a near ideal illumination condition (clear sky without any color cast). Moreover, the outdoor illuminants do not present the same variability of the indoor illuminants. Looking at the parameters and results of the WP, SG, and gGW algorithms on the indoor images, it can be seen that they behave identically as a WP algorithm. This can be explained considering that real outdoor images tend to exhibit color channel clipping in the high intensity range. These very bright pixels are taken as reference white by the three WP-like algorithms and this reference is often very close to the real scene illuminant.

Table 5 summarizes the results on the 835 images (336 indoor and 499 outdoor) in the test set. Here we compare the median angular error obtained by the single illuminant estimation algorithms and our classification-based approach, which applies SG to images classified as indoor and GE2 to those classified as outdoor.

Our results demonstrate that, at least on the dataset we considered, a classification based strategy outperforms general purpose algorithms.

5 Conclusions

We presented a classification-based strategy for automatic illuminant estimation. To improve illuminant estimation accuracy, an image classifier is trained to classify the images as indoor and outdoor, and the best performing algorithm on each class is applied. The proposed strategy can be used with any set of color constancy algorithms.

Experimental results, performed on a suitable subset of the widely used Funt and Ciurea dataset, demonstrate that using different algorithms for indoor and outdoor images improves the results with respect to general purpose algorithms. From our experiments, the best combination of algorithms consisted in the selection of the Shades of Gray algorithm for the indoor images and Second Order Gray Edge for outdoor images. This combination reduced the median angular error by 15.4% with respect to the best general purpose algorithm.

As a future work we are considering the introduction of an uncertainty class. Images for which the classifier is not confident enough will be processed with a more conservative algorithm. We also plan to include other single and combined color constancy algorithms in our strategy.

References

1. Hordely, S.D.: Scene illuminant estimation: Past, present, and future. Color Research & Application 31(4), 303–314 (2006)
2. Gasparini, F., Schettini, R.: Color balancing of digital photos using simple image statistics. Pattern Recognition 37(6), 1201–1217 (2004)
3. van de Weijer, J., Schmid, C., Verbeek, J.: Using High-Level Visual Information for Color Constancy. In: Proc. IEEE 14th International Conference on Computer Vision, pp. 1–8 (2007)
4. Gijsenij, A., Gevers, T.: Color Constancy using Natural Image Statistics. In: Proc. International Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2007)
5. Ciurea, F., Funt, B.: A Large Image Database for Color Constancy Research. In: Proc. IS&T/SID 11th Color Imaging Conference, pp. 160–164 (2003)
6. Funt, B., Barnard, K., Martin, L.: Is machine colour constancy good enough? In: Burkhardt, H.-J., Neumann, B. (eds.) ECCV 1998. LNCS, vol. 1406, pp. 445–459. Springer, Heidelberg (1998)
7. van de Weijer, J., Gevers, T., Gijsenij, A.: Edge-based Color Constancy. IEEE Transactions on Image Processing 16(9), 2207–2214 (2007)
8. Buchsbaum, G.: A spacial processor model for object color perception. Journal of Franklin Institute 310, 1–26 (1980)
9. Cardei, V., Funt, B., Barndard, K.: White point estimation for uncalibrated images. In: Proc. IS&T/SID 7th Color Imaging Conference, pp. 97–100 (1999)
10. Finlayson, G., Trezzi, E.: Shades of gray and colour constancy. In: Proc. IS&T/SID 12th Color Imaging Conference, pp. 37–41 (2004)
11. Barnard, K., Cardei, V., Funt, B.: A comparison of computational color constancy algorithms; part two: Experiments with image data. IEEE Tansactions on Image Processing 11(9), 985–996 (2002)

12. Ciocca, G., Schettini, R.: An Innovative Algorithm for Key Frame Extraction in Video Summarization. Journal of Real-Time Image Processing 1(1), 69–88 (2006)
13. Ciocca, G., Schettini, R.: Supervised And Unsupervised Classification Post-Processing for Visual Video Summaries. IEEE Transactions on Consumer Electronics 2(52), 630–638 (2006)
14. Hordley, S.D., Finlayson, G.D.: Re-evaluating Color Constancy Algorithms. In: Proc. 17th International Conference on Pattern Recognition, pp. 76–79 (2004)
15. Lewis, R.M., Torczon, V.: Pattern search algorithms for bound constrained minimization. SIAM Journal on Optimization 9, 1082–1099 (1999)
16. Lewis, R.M., Torczon, V.: Pattern search methods for linearly constrained minimization. SIAM Journal on Optimization 10, 917–941 (2000)
17. Szummer, M., Picard, R.: Indoor-outdoor image classification. In: Proc. Int. Workshop on Content-Based Access of Image and Video databases, pp. 42–51 (1998)
18. Vailaya, A., Figueiredo, M., Jain, A.K., Zhang, H.-J.: Image classification for content-based indexing. IEEE Transactions on Image Processing 10(1), 117–130 (2001)
19. Schettini, R., Brambilla, C., Cusano, C., Ciocca, G.: Automatic classification of digital photographs based on decision forests. International Journal of Pattern Recognition and Artificial Intelligence 18(5), 819–845 (2004)
20. Breiman, L.: Bagging predictors. Machine Learning 24, 123–140 (1996)
21. Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J.: Classification and Regression Trees, Wadsworth and Brooks/Cole (1984)
22. Wilcoxon, F.: Individual comparisons by ranking methods. Biometrics 1, 80–83 (1945)

An Optimized Fuzzy Based Short Term Object Motion Prediction for Real-Life Robot Navigation Environment

Vijay S. Rajpurohit[1] and M.M. Manohara Pai[2]

[1] Department of Computer Science and Engg.,
vijaysr2k@yahoo.com
[2] Department of Information and Communication Technology,
Manipal Institute of Technology, Manipal 576 104, India
mmm.pai@manipal.edu

Abstract. Short term motion prediction for moving objects in a real life Robotic navigation environment involves uncertainty and temporal validity of the results.Prediction of accurate position of a moving object and responding with quick action are the main objectives of Robot motion planning. This paper proposes a novel algorithm for short term motion prediction involving a fuzzy based predictor. Because of the multi valued nature of the fuzzy logic, this approach enjoys high robustness in dealing with noisy and uncertain data. The knowledge captured by the Rulebase has been optimized by defining directional space. In the proposed work the predictor has been evaluated with three well known defuzzification techniques. Based on the analysis of results, it has been found that Mean Of Maximum defuzzification technique has lower response time and better accuracy. The predictor is tested for various motion patterns in real life environment.

Keywords: Short Term Motion Prediction, Fuzzy Rule base, Rule base Optimization, Directional Space, Fuzzy Predictor Algorithm, Defuzzification.

1 Introduction

Short Term object motion prediction in a real-life Robot navigation environment refers to the prediction of the next instance position of a moving object based on the previous history of its motion. The living beings and vehicles charecterize the dynamic environment and exhibit motion in various directions with different velocities. In an unmanned Robot navigation system, the Robot has to acquire the information of moving objects and predict their future positions for the next instance based on their previous history of motion in order to make efficient path planning[3][7]. The sensors available to read the position of the moving object should send accurate data in quick time succession and the Robot should process and generate the predicted position within limited time as the validity of the result is very short. Real life data often suffer from inaccurate readings due to

M. Sebillo, G. Vitiello, and G. Schaefer (Eds.): VISUAL 2008, LNCS 5188, pp. 114–125, 2008.

environmental constraints, sensors, size of the objects and possible change in motion pattern of the moving objects. This needs the system to be Robust to handle these uncertainties and predict next instance object position as accurate as possible. Research literature has addressed solutions to the short term object motion predictions with different methods such as Curve fitting or Regression methods [9] [22]Neural network based approaches [1][2][5] , Hidden Markov stochastic models [23], Bayesian Occupancy Filters [6], Extended Kalman Filter[12][18] and Stochastic prediction model [21].

Based on the literature survey it is observed that i) The existing models lack flexibility in handling the uncertainties of the real life situations. ii) Probabilistic models sometimes fail to model the real life uncertainties. iii) The existing prediction techniques show poor response time due to their complex algorithmic structure. iv) Most of the approaches validate the results with simulated data or simple navigational environments.

The present work overcomes these difficulties with a simple solution for short term motion prediction using fuzzy inference method. It is assumed that the Robot is instantaneously stationary and observes the moving object through vision based sensors. Position of the moving object is sampled by the Robot at two definite time intervals. With the current sampling positions, the proposed Fuzzy predictor algorithm predicts the future object's position in the next sampling duration.The model is flexible as the navigation environment considered is fuzzy in nature. In the initial step the Fuzzy rulebase is generated by the simulator with different object motion patterns within the navigational environment. Inconsistent and redundant rules are removed by defining directional space within navigational environment. Even though the fuzzy inference rules are relatively more in number , still the response time of the predictor is very small due to the short circuit evaluation of rules[13]. Min_Max , Centre Of Area and Mean Of Max defuzzification techniques are used to generate crisp output to identify the suitabile defuzzification technique in terms of accuracy and quick response time for the current application.The algorithm is tested on both simulated and real-life benchmark data provided by CAVIAR project . The complete process of short term motion prediction is represented in Fig1.

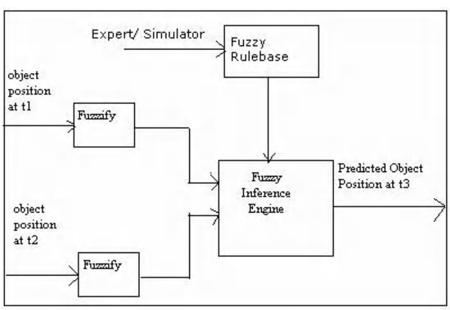


Fig. 1. Short term motion prediction

The paper is organized as follows. In section 2 the Fuzzy membership functions, Fuzzy rule base, optimization of the rulebase,inference process and defuzzification of the outputs for the predictor are discussed. Fuzzy predictor algorithm is presented in section 3. In Section 4 experimental results and the analysis of the results are presented. Finally, concluding remarks are given in section 5.

2 Fuzzy Based Object Motion Prediction

Robot navigation system has to handle large amount of uncertaint data in real life environment. Fuzzy logic addresses this problem as it takes uncertain data, processes it and obtains certain and finite data[4][11][14] [16] [17] [20]. The object motion prediction is done by incorporating human experience in the form of fuzzy inference rules. It is assumed that, the environment is observed through stereo vision technique. The observed environment covers semi circular area in front of the Robot and is also referred to as the planning window. The planning window divides the navigation space into 7 Fuzzy categories in range wise and 7 Fuzzy categories in direction wise. Each rule considers the object positions at time t1 and t2 as antecedents and predicted position at t3 as consequent. The approach considers multiple conditional constructs of the form

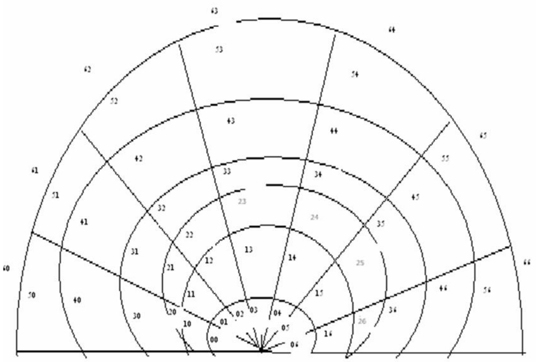
$$IF(R1, \theta 1) \text{ and } (R2, \theta 2) \text{ THEN } (R3, \theta 3)$$

(where R and θ are the Range and Angle of object from Robot at two different time instances).

2.1 Fuzzy Membership Functions

The Robot is capable of visualizing the navigation environment in front (about 180^0 in semi circular range). Fuzzy regions in front of the Robot are defined according to the visualization capability of the sensors [15]. Each object detected has a distance variable from the Robot. This range data has a different membership in each of the 7 range subsets defined as VVFAR (VVF), VFAR (VF), FAR (F), MODERATE(M), NEAR (N), VNEAR(VN), VVNEAR(VVN). The direction of universe is divided into 7 subsets. The linguistic variables that describe the Angle heading are VVLEFT(VVL), VLEFT(VL), LEFT(L), FRONT(F), RIGHT(R), VRIGHT(VR), VVRIGHT(VVR).

The Range and Angle information has to be represented by a suitable membership function. Many authors have addressed critical issues relating to the selection and performance of fuzzy membership functions for various real-time robot control applications[8][10][19]. In most of the cases Triangular membership function has proved superior over other membership functions like trapezoidal, Gaussian, bell shaped, polynomial-PI,and Sigmoidal. For our application as the prediction needs to be more accurate and the strength of the rule/ rules fired can make remarkable difference in the prediction, selection of triangular membership function for representing Angle and Range values is inevitable. The selection of 07 Fuzzy subsets for Range and Angle is moderate as selecting 05 categories will



a. Division of space in Fuzzy subsets of Range and Direction

Fuzzy Range	Numeric Notation	Fuzzy Angle	Numeric Notation
Vvnear	0	vvleft	0
Vnear	1	vleft	1
Near	2	Left	2
Mod	3	Front	3
Far	4	Right	4
Vfar	5	Vright	5
Vvfar	6	Vvright	6

b. Numeric notation for Fuzzy range and Angle Values

Fig. 2. Division of Navigation Space space into Fuzzy subsets of Range and Direction

have less number of fuzzy rules but quality of prediction may reduce if navigation space is large, selecting 09 or more number of categories will increase the number of fuzzy rules as well as the complexity of the system which could reduce the response time of the predictor. Both Range and Angle subsets are normalized between 0-1.

The Fuzzy representation of the environment is shown in Figure 2 with numerical notation for each region. The Fuzzy representation divides the whole navigation environment in to different regions like VVFAR VLEFT (61), FAR RIGHT (44) and NEAR FRONT(23) etc.

2.2 Fuzzy Rule Base

Initial fuzzy rule base is generated by the simulator by generating different motion patterns within the navigation environment with various source and destinations for the objects moving in the environment.

For the current application, the Fuzzy rulebase is of the form

$$\begin{aligned}&If\ (Range\ is\ r1\ and\ Angle\ is\ \theta 1\ \ at\ \ t1)\ AND\ (Range\ is\ r2\ and\ Angle\ is\ \theta 2\ \ at\ t2)\\ &THEN\ (Range\ is\ r3\ and\ Angle\ is\ \theta 3\ \ at\ t3)\end{aligned} \tag{1}$$

Where t1, t2 represent the two equal time instances at which the object is observed and t3 represent the time instance at which the object is predicted.

2.3 Optimization of the Fuzzy Rulebase

The number of fuzzy rules generated by the input-output pairs is usually large. Inconsistent and redundant rules are inevitable. Rulebase optimization is an important task in the desgn of Fuzzy inference system as it ensures the correctness and robustness by detection of anomalies.

Many of the rules defined in the system look inconsistent such as

1. IF object at t1 is VFAR,VVLEFT and object at t2 is FAR,VVLEFT THEN object predicted at t3 is MOD,VVLEFT.
2. IF object at t1 is VFAR,VVLEFT and object at t2 is FAR,VVLEFT THEN object predicted at t3 is FAR,VVLEFT.

Where the antecedents are same but the consequents are different. The reason for the inconsistency is due to the direction of traversal of the object. Future motion of the object is dependent on the history of the direction of the traversal of the object. To overcome this type of inconsistency while defining the rule base, partitioning of the navigational space is done. Considering the navigational space that is tessellated in eight geographical directions, the sensor readings of the object positions taken at previous two time intervals forms a trajectory in one of these directions.

{SW(d1),S(d2),SE(d3),E(d4),NE(d5),N(d6),NW(d7),W(d8)}

A separate directional space is created for each direction (Figure 3) and rules are clustered based on the direction of traversal object. Depending on the direction of traversal of the object, only those rules which belong to that directional space will be selected for processing.

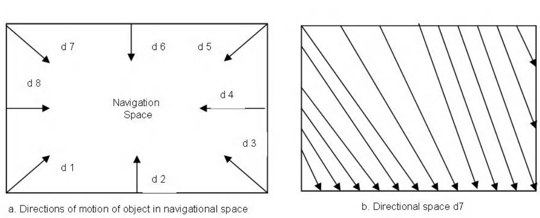


Fig. 3. Division of navigational space into Directional Space

2.4 Fuzzy Inference Process and Defuzzification

Inference process is done by Mamdani model.One of the most widely used fuzzy models for inference is the Mamdani model which consists of the following linguistic rules that describe a mapping from U1 x U2 x U3. . . .Ur to W

Given the inputs of the form

X1 is A'1 ,x2 is A'2 , ...Xr is A'r

Where A'1 , A'2,. . . .A'r are fuzzy subsets of U1,U2,. . . Ur. The contribution of rule Ri to Mamdani model's output is a fuzzy set whose fuzzy membership function is computed by

$$\mu c'i(y) = (\alpha i1 \wedge \alpha i2 \wedge .... \wedge \alpha in) \wedge \mu ci(y) \quad (2)$$

where αi is the matching degree of rule Ri, and where αij is the matching degree between xj and Ri 's condition about xj. $\mu ci(y)$ is the membership function for set Ci(y)

$$\alpha\, ij = \sup_{x\, j}(\mu A`j(xi) \wedge \mu Aij(xi)) \tag{3}$$

Where ˆ denotes the min operator. The final output of the model is the aggregation of outputs from all rules using max operator.

$$\mu c(y) = \max\{\mu c'1(y), \mu c'2(y), ...\mu c'L(y)\} \tag{4}$$

Defuzzification of the final output is done to get the crisp value. Three most commonly used defuzzification techniques are considered i.e. Fuzzy OR method/ Min-Max, Center of Area and Mean Of Maxima methods.

Min-Max/Fuzzy Or method. The output of all the rules fired are combined using disjunction or maximum operator. This final output is multiplied by the range of output variable.

$$S = ((\, s1 \wedge s2) \vee (s3 \wedge s4) \vee (s5 \wedge s6) \vee (s7 \wedge s8)........(sn1 \wedge sn2)) * s \max_{out} \tag{5}$$

Where S is the final crisp output produced, s1,s2...sn are the output of the fuzzy membership functions, $smax_{out}$ is the range of the output variable.

Mean of Maximum. The Mean of maximum (MOM) defuzzification method calculates the average of those output values that have the highest possible degrees. Suppose "y is A" is a fuzzy conclusion to be defuzzified, then we can express the MOM defuzzification method using the following formula.

$$MOM(A) = \sum(y*)/\, |P| \quad where\ y{*}\varepsilon\, P \tag{6}$$

Where P is the set of output values y with highest possibility degree in A i.e.
$P = \{y*|\mu_A(y*) = sup_y \mu_A(y)\}$

Center of Area/Gravity. This method is also known as center of gravity or center of area defuzzification. This defuzzification technique can be expressed as

$$R = \frac{\sum di\ \mu A(di)}{\sum\ \mu A(di)} \tag{7}$$

Where R represents the crisp output variable i.e. the predicted position of the object , di is the ith domain value and μA (di) is the truth membership value for that domain point.

3 Fuzzy Predictor Algorithm

The short term prediction for a moving object is represented in the form of a prediction algorithm. Positions of moving objects are sampled by the Robot's sensor system at two definite time intervals.

$C(r1, \theta 1)^{t1}$ represents the object position at time t1, where r1 represents range and $\theta 1$ represents angle. $C(r2, \theta 2)^{t2}$ represents the object position at time t2, where r2 represents range and $\theta 2$ represents angle. With the current sampling positions, the Fuzzy predictor model predicts the future object's positions in the next sampling duration. $C(r3, \theta 3)^{t3}$ represents the predicted range and angle of the object at time instance t3.

The Fuzzy Predictor algorithm
start
Read the object position $c(r1, \theta 1)^{t1}$ *at time t1*
Read the object position $c(r2, \theta 2)^{t2}$ *at time t2*
Fuzzify the object positions at t1 and t2
$c(r1, \theta 1)^{t1} \rightarrow fr1^{t1}$ *the fuzzy Range at t1*
$f\theta 1^{t1}$ *the fuzzy Angle at t1*
$c(r2, \theta 2)^{t2} \rightarrow fr2^{t2}$ *the fuzzy Range at t2*
$f\theta 2^{t2}$ *the fuzzy Angle at t2*
Generate Fuzzy rulebase
Apply rule optimization to remove redundant and inconsistent rules
Apply Fuzzy inference method to calculate predicted position at t3
$pred((fr1^{t1}, f\theta 1^{t1}) \hat{} (fr2^{t2}, f\theta 2^{t2})) \rightarrow (fr3^{t3}, f\theta 3^{t3}))$
Defuzzify the output using Min_Max Defuzzification
$Defuzzify\_Min\_Max((fr3^{t3}, f\theta 3^{t3})) \rightarrow C\_Min\_Max(r3, \theta 3)^{t3}$
Defuzzify the output using MOM Defuzzification
$Defuzzify\_MOM((fr3^{t3}, f\theta 3^{t3})) \rightarrow C\_MOM(r3, \theta 3)^{t3}$
Defuzzify the output using COA Defuzzification
$Defuzzify\_COA((fr3^{t3}, f\theta 3^{t3})) \rightarrow C\_COA(r3, \theta 3)^{t3}$
end

4 Experimental Results and Analysis

Object motions with different motion patterns are generated by a simulator in different directions to generate the initial rulebase. The rules generated are clustered based on the direction of the motion pattern into the directional space clusters. Table1 shows the number of rules remained in each directional space after removing inconsistencies and redundancies.

Our predictor algorithm is tested for real life bench mark dataset available at: http://homepages.inf.ed.ac.uk/rbf/CAVIAR/. (EC Funded CAVIAR project/IST 2001 37540) to check for relative error. The data set consists of different human motion patterns observed at INRIA Lab at Grenoble, France

Table 1. Optimized Fuzzy rules in each directional space

d1	d2	d3	d4	d5	d6	d7	d8
143	178	146	152	141	172	144	183

Fig. 4. A typical Scenario observed at INRIA Lab and Shopping Centre (courtesy : CAVIAR Project)

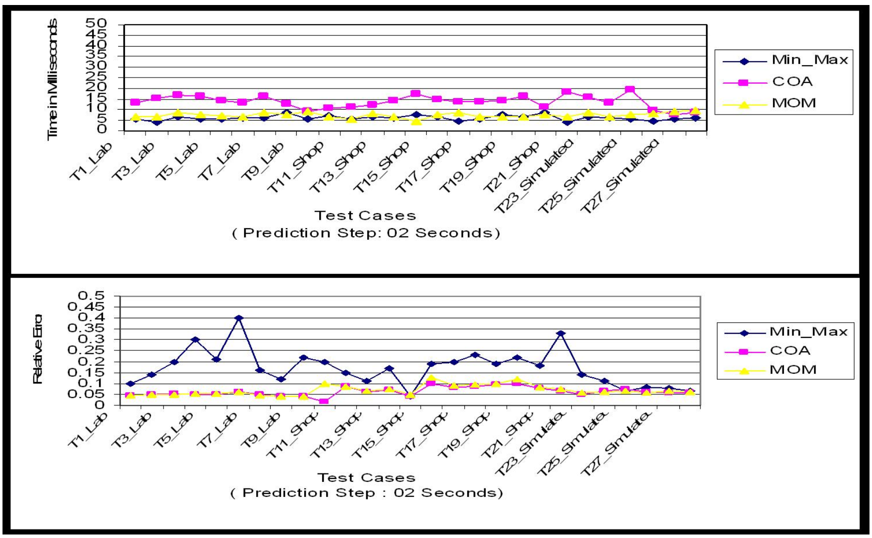


Fig. 5. Average Response time and Relative Error of the predictor at prediction step:02 seconds

and Shop Centre . These motion patterns consists of frames captured at 25 frames/second. A typical scenario of the INRIA Lab and the Shop Centre are shown in fig 4.

We define the relative error (RE) for M sample test data as

$$RE = (\sum_{M}(|da - dp|/da))/M \tag{8}$$

Where da is the actual distance, dp is the predicted distance.

The Fuzzy predictor was run on 1.66GHz machine in VC++ environment. Fig.5 represent the average relative error observed for the prediction algorithm for various test cases using Min_Max, MOM and COA defuzzification techniques. For each test case average response time is also calculated to find its suitability to real life environment. The prediction algorithm is tested by processing the

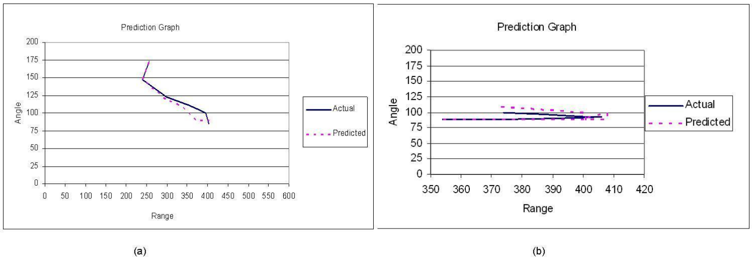


Fig. 6. Prediction graphs showing the some of path prediction solutions with MOM Defuzzification

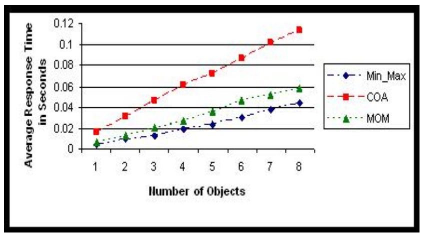


Fig. 7. Response Time of the Predictor for Multiple Number of Objects

frame data of moving human patterns stored in the data base at intervals of 50 frames (02 Seconds). It is observed that the predictor with MOM defuzzification performs better in terms of response time and less relative error.

The navigation environment is presented in the form of Prediction graph where x axis represents the Range parameter and the y axis represents the Angle parameter. The predicted Angle and Range values are compared with actual values obtained from real-life environment In most of the cases the predicted values are in the region of the actual measured range and Angle values. Figure 6. illustrates some of the results obtained for path prediction using MOM defuzzification.

Table 2 compares some of the well known prediction techniques with our fuzzy predictor with MOM defuzzification for response time, relative error and Environmental constarints.

The performance of the predictor is tested when more than one objects are sensed by the sensor. The response time of the predictor for all the objects should be acceptable for real-life applications. The tests are carried out assuming at most 6-8 objects can be visible and can affect the decisions to be made regarding robot traversal(Figure 7).

4.1 Analysis of Results

Based on the results obtained it can be concluded that the Fuzzy predictor with MOM defuzzification has less relative error and quick response time as compared

Table 2. Comparison of Short term predictors

Short Term Predictor	Environmental Constraints if any	Relative Error	Response time in seconds
ANN predictor [5]	Simulated Environment With Rectilinear Paths,	6-17%	560x 10^{-3} sec
Bayesian Occupancy Filters[6]	Only for small scale environments due to the intrinsic complexity of the related inferences and numerical computations. Not applicable to large scale Environments	Not Specified	100 x 10^{-3} sec
Polynomial NN[1]	Simulated Environment, Quality of prediction is poor for un trained regions	1-10%	Not Specified
Auto Regressive model[23]	Simulated Environment, Quality of prediction depends on accurate identification of model coefficients, implemented for linear paths	Not Specified	Computationally intensive
Extended Kalman Filter[19]	Simulated Environment, Quality of prediction decreases with increase in planning space	1-10%	Not Specified
Fuzzy Predictor with MOM	**Real life Environment, Any Real life Environment can be mapped to Fuzzy Navigation Environment**	**1-10%**	**07x 10^{-3} sec to 09 x 10^{-3} sec**

to other defuzzification techniques. In real life applications the Robot has to deal with multiple number of objects and the total response time for all the objects motion prediction should be less than the time gap between two sensor readings. Our predictor is able to generate results suitable for real life situations. Based on the comparison with existing prediction techniques(Table2), our predictor is more flexible, simple to implement and deals with noisy and uncertain data of real life situations.The relative error of 5-10% is acceptable for our system as the predicted fuzzy region and the fuzzy region of actual position remains the same.

5 Conclusion

In a dynamic navigation system the Robot has to avoid stationary and moving objects to reach the final destination. Short Term motion prediction for moving objects in such an environment is a challenging problem. This paper proposes a simplified approach for predicting the future position of a moving object using fuzzy inference rules derived from expert knowledge. Fuzzy based prediction is more flexible, can have more real-life parameters, comparable to the existing approaches and suited for real-life situations. The results of the study indicate that the that the Fuzzy predictor algorithm with Mean of Maximum(MOM) defuzzification gives better performance when compared with other prediction techniques. The authors are in the process of further optimizing the rulebase and improve the response time of the predictor.

Acknowledgements. The authors are thankful to the benchmark dataset provided by EC Funded CAVIAR project/IST 2001 37540.We are indebted to AICTE, Government of India, for funding our project.

References

1. Foka, A., Trahanias, P.E.: Predictive Autonomous Robot navigation. In: Proceedings of the 2002 IEEE/RSJ International Conference on Intelligent Robots and Systems, EFPL, Lausanne, Switzerland, october, pp. 490–494 (October 2002)
2. Foka, A., Trahanias, P.E.: Predictive control of Robot velocity to avoid objects in dynamic environments. In: Proceedings of the 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems, October 27-31, pp. 370–375. Las Vegas, Nevada (2003)
3. Bruce, A., Gordon, G.: Better Motion Prediction for People-tracking. In: Proceedings of. IEEE International Conference on Robot. Automation (ICRA 2004) (2004)
4. Fayad, C., Web, D. P.: Optimized Fuzzy logic based algorithm for a mobile robot collision avoidance in an unknown environment. In: 7th European Congress on Intelligent Techniques & Soft computing, Aachen, Germany, September 13-16 (1999)
5. Charles, C., Song, K.-T.: Dynamic motion planning based on Real-Time object Prediction. In: Proceedings of the 1996 IEEE Intl. Conference on Robotics and Automation Minnesota,
6. Laugier, C., Petti, S., Vasquez, A.D., Yguel, M., Fraichard, T., Aycard, O.: Steps Towards Safe Navigation in Open and Dynamic Environments. In: Laugier, C., Chatila, R. (eds.) Autonomous Navigation in Dynamic Environments: Models and Algorithms. Springer Tracts in Advanced Robotics Series (STAR). Springer, Heidelberg (2006)
7. Vasquez, D., Large, F., Fraichard, T., Laugier, C.: High speed autonomous navigation with motion prediction for unknown moving objects. In: Proceedings of IEEE Intelligent Robots and systems,2004. Sendai (JP), pp. 82–87 (October 2004)
8. Huang, H., Chang, L.: Simulation study on the membership functions of the basic logic controller for the electric vehicle propulsion systems. In: The proceedings of the Canadian Conference on Electrical and Computer Engineering, Calgary, pp. 1004–1007 (May 1996)
9. Yu, H., Su, T.: A Destination Driven Navigator with Dynamic Object Motion Prediction. In: The proceedings of International Conference on Robotics 2L Automation, Seoul, Korea, May 21-26, pp. 2692–2697 (2001)
10. Zhao, J., Bose, B.K.: Evaluation of membership functions for fuzzy logic controlled induction motor drive. In: IECON Proceedings (Industrial Electronics Conference), vol. 1, pp. 229–234 (2002)
11. Yen, J., Langari, R.: Fuzzy logic: intelligence, control, and information. Prentice-Hall, Inc, Englewood Cliffs (1998)
12. Kalman, R.: A New Approach to Linear Filtering and Prediction Problem. Transactions of the ASME-Journal of Basic Engineering 82, 35–45 (1960)
13. Louden, K.C.: Principles of Programming Languages: Principles and Practice. Brooks/Cole, Thomson (2004)
14. Webb, P., Fayad, C., Brettenbatch.: The integration of an optimized Fuzzy logic navigation algorithm into a semi-autonomous Robot Control System. In: Paper presented at the International Workshop on Recent Advances in Mobile Robots, Leicester, UK, June 29 (2000)

15. Memon, Q., Khan, S.: Camera Calibration and Three-Dimensional World Reconstruction of Stereo-Vision Using Neural Networks. International Journal of Systems Science 32, 1155–1159 (2001)
16. Leyden, M., Toal, D., Flanagan, C.: A Fuzzy logic based navigation system for a mobile Robot. In: The 2nd Wismar Symposium on Automatic Control, Wismar (1999)
17. Cao, M., Hall, E.: Fuzzy logic control for an automated Guided Vehicle. In: The SPIE international conference, Boston (November 1998)
18. Madhavan, R., Schlenoff, C.: Moving Object Prediction for Off-road Autonomous Navigation. In: Proceedings of the SPIE Aerosense conference, Orlando, April 21-25 (2003)
19. Berkan, R.C., Trubatch, S.L.: Fuzzy Systems Design Principles: Building Fuzzy If-Then Rule Bases. John Wiley & Sons Inc, Chichester (1997)
20. Yen, J., Pfluger, N.: A Fuzzy logic based extension to Payton and Rosenblatt's command fusion method for mobile robot navigation. IEEE Transactions on Systems, Man and Cybernetics 25, 971–978 (1995)
21. Nam, Y.S., Lee, B.H., Kim, M.S.: View-Time Based Moving Object Avoidance using Stochastic Prediction of Object Motion. In: Proceedings of the 1996 IEEE International Conference on Robotics and Automation Minneapolis, Minnesota (April 1996)
22. Hui-Zhong, Z., Shu-Xin, D., Tie-jun, W.: On-line real-time path planning of mobile Robots in dynamic uncertain environment. Journal of Zheing University Science A, 516–524 (2006)
23. Zhu, Q.: Hidden Markov Model for Dynamic Object Avoidance of Mobile Robot Navigation. IEEE Transactions on Robotics and Automation, 390–396 (1991)

Experiments on Selection of Codebooks for Local Image Feature Histograms*

Ville Viitaniemi and Jorma Laaksonen

Department of Information and Computer Science, Helsinki University of Technology, P.O. Box 5400, FI-02015 TKK, Finland
{ville.viitaniemi,jorma.laaksonen}@tkk.fi

Abstract. Histograms of local features have proven to be powerful representations in image classification and object detection. In this paper we experimentally compare techniques for selecting histogram codebooks for the purpose of classifying 5000 images of PASCAL NoE VOC Challenge 2007 collection. We study some well-known unsupervised clustering algorithms in the task of histogram codebook generation when the classification is performed in post-supervised fashion on basis of histograms of interest point SIFT features. We also consider several methods for supervised codebook generation that exploit the knowledge of the image classes to be detected already when selecting the histogram bins.

1 Introduction

Histograms of local features have proven to be powerful representations in image classification and object detection. Consequently their use has lately become very commonplace in image content analysis tasks (e.g. [8,11]). For instance, the most successful entries to the PASCAL NoE Visual Object Classes (VOC) Challenge 2006 image analysis performance evaluation competition were based on histogram techniques. In the 2007 Challenge [1] most of the participants utilised local feature histograms solely or in combination with other approaches.

Use of local image feature histograms for supervised image classification and characterisation requires performing several steps. Image locations have to be selected and suitable local features extracted from them. Subsequently, one needs to generate histograms of the extracted features. The histograms are then fed into some supervised classification algorithm. All the parts of the analysis chain are subject of continuous study. In this paper we concentrate on the histogram-forming part and consider the feature extraction and classification techniques to be given.

An essential question in histogram generation is the selection of histogram bins, i.e. the histogram codebook. Usually the selection is done by applying some unsupervised clustering algorithm. In this case the supervised classification procedure is called *post-supervised* as the supervision is applied only after the

* Supported by the Academy of Finland in the *Finnish Centre of Excellence in Adaptive Informatics Research* project.

M. Sebillo, G. Vitiello, and G. Schaefer (Eds.): VISUAL 2008, LNCS 5188, pp. 126–137, 2008.

unsupervised codebook selection. Alternatively, information of the class labels of the data can be used already when forming the codebook for the histograms. In this case the classification process is called *pre-supervised*.

The conventional unsupervised clustering algorithms are devised for other purposes and their suitability to codebook generation is not self-evident. In this paper we experimentally compare the histogram generation capabilities of several clustering algorithms in the task of classifying images of one specific collection—the images of the VOC Challenge 2007—according to the objects they contain. Additionally, we consider a few alternative techniques for supervised codebook generation and test them in the same image classification task.

The rest of this paper is organised as follows. Section 2 briefly outlines prototype selection algorithms that will be used in the experiments. In Sect. 3 we detail the image analysis task and the experimental procedures. In Sections 4 and 5 we describe experiments with unsupervised and supervised generation of codebooks, and report the results of the experiments. Finally, in Sect. 6 we present our conclusions and discussion.

2 Prototype Selection Algorithms

In this section we briefly introduce prototype selection algorithms that will be applied later in our experiments. These algorithms try to find a pre-specified number of prototype vectors $\{\mathbf{p}_k\}$ that describe a vectorial data set $\{\mathbf{v}_i\}$ as well as possible. The algorithms described here divide into two groups: unsupervised and supervised.

The unsupervised clustering methods use only the knowledge of the distribution of the vectors $\{\mathbf{v}_i\}$. One way to quantify the quality of the clustering is to calculate the squared quantisation error when the vectors of the data set are approximated with their nearest prototype vectors:

$$E = \sum_i \|\mathbf{v}_i - \mathbf{p}_{\mathrm{q}(\mathbf{v}_i)}\|^2, \tag{1}$$

where the index i enumerates all the data set vectors and $\mathrm{q}(\mathbf{v})$ is the index of the prototype vector nearest to the vector $\mathbf{v}$.

The k-means algorithm [7] is an iterative algorithm that approximately optimises the above mentioned quantisation error E. The algorithm starts with some initial positions for the prototype vectors $\mathbf{p}_k$. In each iteration of the algorithm's ISODATA variant, each prototype vector is moved to the centroid of the data vectors for which the prototype is currently the closest one. The algorithm terminates when the prototype vectors no longer change their position. In our experiments we started with randomly partitioning the data vectors to prototypes and taking the initial prototype positions to be centroids of the partitions. We stopped the algorithm before full convergence when the number of data vectors that had different nearest prototype vectors on consecutive rounds of the iteration became sufficiently small.

The Linde-Buzo-Gray (LBG) algorithm [4] is another iterative clustering algorithm. It starts with only one prototype vector and continually doubles the number of prototypes. In the algorithm the position of the prototypes is first optimised in an iteration that resembles the k-means algorithm. The quantisation error is monitored and the iteration is stopped when the relative improvement in one iteration round no longer exceeds a threshold. If the codebook already contains the pre-specified number of prototype vectors the algorithm stops. Otherwise the codebook size is doubled and the iteration is started over.

The Self-Organising Map (SOM) [2] is an iterative algorithm that can be interpreted to approximately optimise

$$E_{\mathrm{SOM}} = \sum_i \sum_k \mathrm{h}_{k,\mathrm{q}(\mathbf{v}_i)} \|\mathbf{v}_i - \mathbf{p}_k\|^2, \tag{2}$$

where the index i enumerates the data set and k the set of prototype vectors. $\mathrm{h}_{k,l} \in [0, 1]$ is a pre-specified neighbourhood function that reflects topological closeness of map units k and l. This cost function otherwise corresponds to the squared quantisation error, but enforces also ordering of the prototype vectors according to a pre-specified topology expressed in terms of the neighbourhood function. In our experiments we use neighbourhood functions imposed by a two-dimensional rectangular topology. The Tree-Structured SOM (TS-SOM) [3] algorithm is a computational shortcut to the SOM algorithm that speeds-up the search for nearest prototype vector by defining a hierarchy of SOMs of increasing size. After forming a codebook with the TS-SOM algorithm, we use only its bottom level for the histogram generation.

The Learning Vector Quantisation (LVQ) algorithms [2] are supervised prototype selection algorithms that iteratively update the position of the prototype vectors so that the 1-nearest-neighbour classification boundaries approximate the true Bayesian classification boundaries. These algorithms work satisfactorily in the case of binary classes, and also in the case of multiple classes when the overlap structure of the classes is simple. In our experiments we used the LVQ1 algorithmic variant with optimised learning rate adaptation (OLVQ1).

3 Image Collection and Experimental Procedures

In this paper we consider the image ranking problem in a supervised setting. This problem can be approached by extracting visual features from both training and test images and applying supervised learning techniques to representations of these features. In the experiments reported here we keep other parts of the ranking pipeline fixed, but concentrate on methods of representing the features with histograms.

A powerful and nowadays popular approach to visual feature extraction is to extract a large number of local features and describe each image with the histogram of such features. For our experiments we standardise the local feature extraction and determine the SIFT features [5] of interest points detected with the Harris-Laplace detector [6]. We then compare different methods of selecting the codebooks for the histograms. The histograms are then fed into a SVM

Table 1. The 20 object classes of VOC Challenge 2007

Category	Class	Images	Images in partitions
Person	person	2008	983–1025
Animal	bird	330	150–180
	cat	337	145–192
	cow	141	69–73
	dog	421	188–233
	horse	287	128–159
	sheep	96	36–60
Vehicle	aeroplane	238	112–126
	bicycle	243	111–132
	boat	181	81–100
	bus	186	82–104
	car	713	337–376
	motorbike	245	104–141
	train	261	116–145
Indoor	bottle	244	105–139
	chair	445	204–241
	dining table	200	86–114
	potted plant	245	111–134
	sofa	229	107–122
	tv/monitor	256	114–142

classifier that utilises a χ^2-kernel. Further details of the classification procedure can be found in [10].

The experiments reported in this paper have been performed on the images and object annotations of the PASCAL NoE Visual Object Classes Challenge 2007 collection [1]. In the collection there are altogether 9963 photographic images of natural scenes. In the experiments we use one half of the images, selected for training and validation and denoted "trainval" by the challenge organisers. Each of the images contains at least one occurrence of the 20 object classes detailed in Table 1. The presences of these objects in the images were manually annotated by the organisers. In many images there are objects of several classes present. Altogether the images of the whole collection contain 24640 objects. The most common object class ("person") is present in 40% of the images, the rarest (sheep) in 1.9%. Often there are several objects of the same kind in an image. The "trainval" half of the collection consists of 5011 images. These images are represented by the SIFT features of approximately 1200 interest points per image on average, approximately 6 million interest points altogether.

The experiments are performed in supervised learning setting. To this end, the used 5011 images are partitioned approximately equally into training and test sets. Every experiment was performed separately for each of the 20 object classes. Given the labelled training images, the goal of the system was to produce a ranking of the test images according to descending estimated likelihood of the images to contain the target object. The goodness of the produced rankings was assessed by comparing the ranking with manually determined ground truth annotations. In this paper we report the results using the non-interpolated average precision (AP) performance metric [9]. We have also evaluated the experiments using the receiver operating characteristic (ROC) area under curve (AUC) metric. Those results pointed to the same direction as the AP results, but showed somewhat more statistical variation.

The statistical fluctuations of the results were easily observed to be rather pronounced. To estimate the statistical reliability of the results, most of the experiments were repeated with six different partitionings of the images into training and test sets. In all the partitionings the training and test sets contain 2501 and 2510 images, respectively. One of the partitionings coincides with "train" and "val" sets defined by the challenge organisers. This partitioning was used for those experiments that we were not able to repeat for all the partitionings. Other partitionings were selected randomly within the constraint on the total number of images in the training and test halves. The number of occurrences of the 20 object classes in the partitions was not constrained. Table 1 shows the ranges of the variation in the number of images of individual classes in the partitions, confirming that the class distribution is reasonably similar in all of the partitions.

In the reported results we show the 95% confidence interval of the mean of the results over random partitionings. As usual, the interval is determined based on assumption on normal distribution and is not to be taken literally. The statistical fluctuation in most of our results for the individual 20 object classes was found to be so strong that our experiments could not reliably resolve the order of the codebook generation methods in classification within individual object classes. We thus report only the mean average precision (MAP) over all the object classes that fluctuates less due to its averaging nature.

4 Experiments on Unsupervised Codebook Selection

In this section we describe experiments where the codebooks for histogram generation were obtained by applying unsupervised clustering methods to the set of the extracted SIFT feature vectors. We report here the ranking performance when these histograms were used as bases for supervised classification. To speed up the clustering algorithms, the original set of six million SIFT feature vectors was sampled to 20 vectors per image, i.e. to approximately 100 000 vectors in total. As the number of retained vectors still highly exceeds the considered codebook sizes, we estimate that the sampling did not to essentially affect the codebook quality.

We formed codebooks using all the unsupervised clustering algorithms of Sect. 2: the k-means, LBG, SOM and TS-SOM algorithms. For comparison, we also included an alternative where randomly selected vectors of the data set were chosen as prototype vectors. As the LBG algorithm produces codebooks with sizes that are powers of two, we selected similar sizes for the codebooks produced by the other algorithms to ensure comparability of the results. We thus used codebooks with 128, 256, 512, 1024, 2048, 4096 and 8192 units. The

Table 2. Codebook statistics

Algorithm		MSE	H/bits	unused
random	256	0.362	6.8	1
	512	0.331	8.0	2
	1024	0.306	9.0	5
	2048	0.283	10.0	20
	4096	0.259	11.0	78
	8192	0.233	12.0	339
LBG	128	0.260	6.8	0
	256	0.242	7.7	0
	512	0.226	8.7	5
	1024	0.213	9.4	17
	2048	0.201	10.1	61
	4096	0.190	10.7	226
	8192	0.178	11.2	1444

Algorithm		MSE	H /bits	unused
k-means	256	0.238	7.8	0
	512	0.222	8.8	0
	1024	0.207	9.8	1
	2048	0.193	10.7	1
	4096	0.178	11.5	3
SOM	16x16	0.268	7.9	0
	25x20	0.256	8.9	0
	40x25	0.249	9.88	0
	50x40	0.232	10.9	0
	80x50	0.223	11.9	0
TS-SOM	16x16	0.288	7.7	0
	64x64	0.223	11.54	0

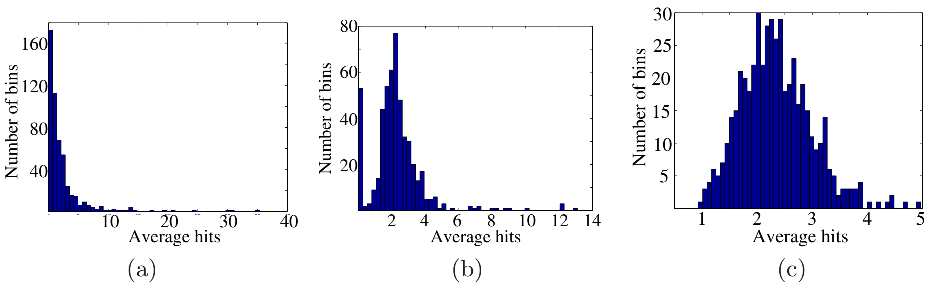


Fig. 1. Histograms of average hit counts of the histogram bins for three codebooks: (a) random codebook with 512 units, (b) LBG codebook with 512 units, and (c) SOM codebook with 25x20 units

SOM codebooks originated from earlier experiments and therefore the number of units in them were not exactly powers of two. However, the differences are insignificantly small.

Table 2 shows some statistics of the codebooks. The numbers appearing next to the algorithm identifiers are the sizes of the generated codebooks. In the "MSE" column we have evaluated the mean quantisation error in the training data set. In this respect, the k-means and LBG codebooks are clearly superior to the random codebooks, k-means being slightly more accurate. This is balanced by the significantly higher computational cost of the k-means algorithm. SOM and TS-SOM codebooks are between these two and the random selection with respect to their quantisation accuracy, TS-SOM codebooks being computationally essentially lighter generate.

The next column ("H/bits") of the table characterises the evenness of the distribution of hits to various histogram bins by means of entropy. We see that the values are rather close to the theoretical maximum of even distribution (e.g. 8 bits for the histograms of size 256), although some deviating structure can clearly

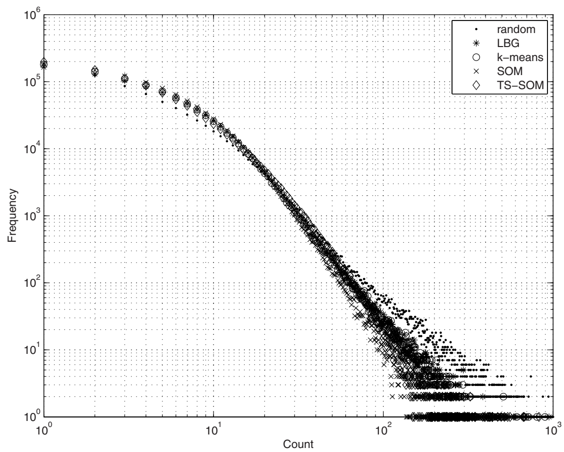


Fig. 2. Distributions of histogram bin hit counts within individual images for different codebooks with 256 units

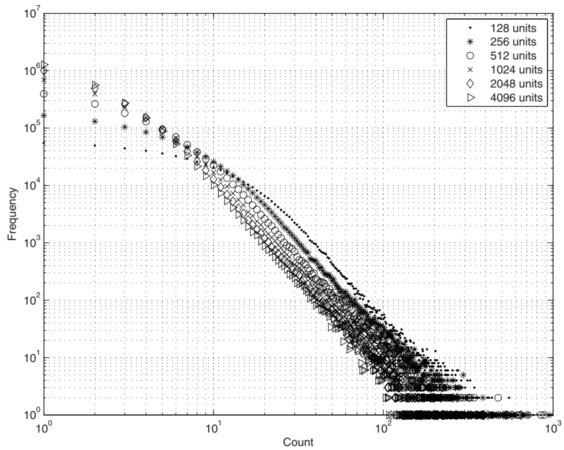


Fig. 3. Distributions of histogram bin hit counts within individual images for LBG codebooks of different sizes

be observed in the distributions of Fig. 1. The various clustering algorithms are slightly different in this sense. The rightmost column of the table shows that the codebooks do not have very many dead units that do not encode any portion of the data, with the LBG codebooks being an exception. Figures 2 and 3 illustrate the distribution of hit counts the histogram bins receive within individual images. In the former image different symbols correspond to equally sized codebooks resulting from different clustering algorithms. In the latter figure the symbols correspond to different sizes of LBG codebooks. The overlap of the symbols indicates that the distributions are essentially quite similar. Even in the larger

Table 3. MAP performance resulting from unsupervised codebook selection

Algorithm		MAP norm	MAP nonorm	Algorithm		MAP norm	MAP nonorm
random	256	0.339±0.004	0.355±0.007	k-means	256	0.366±0.010	0.381±0.012
	512	0.363±0.010	0.376±0.010		512	0.380±0.007	0.391±0.007
	1024	0.376±0.008	0.385±0.011		1024	0.390±0.006	0.400±0.007
	2048	0.384±0.002	0.388±0.008		2048	0.397±0.008	0.399±0.008
	4096	0.385±0.003	0.384±0.006		4096	-	0.394±0.008
	8192	-	0.376±0.009	SOM	16x16	0.330±0.007	0.347±0.007
LBG	128	-	0.357±0.009		25x20	0.345±0.007	0.360±0.009
	256	0.360±0.008	0.376±0.008		40x25	0.353±0.003	0.365±0.005
	512	0.376±0.007	0.389±0.009		50x40	0.361±0.010	0.365±0.008
	1024	0.385±0.008	0.397±0.008		80x50	0.349±0.007	0.354±0.006
	2048	0.397±0.007	0.400±0.007	TS-SOM	16x16	0.328±0.008	0.343±0.007
	4096	0.398±0.010	0.404±0.008		64x64	0.367±0.005	0.368±0.007
	8192	-	0.398±0.009				

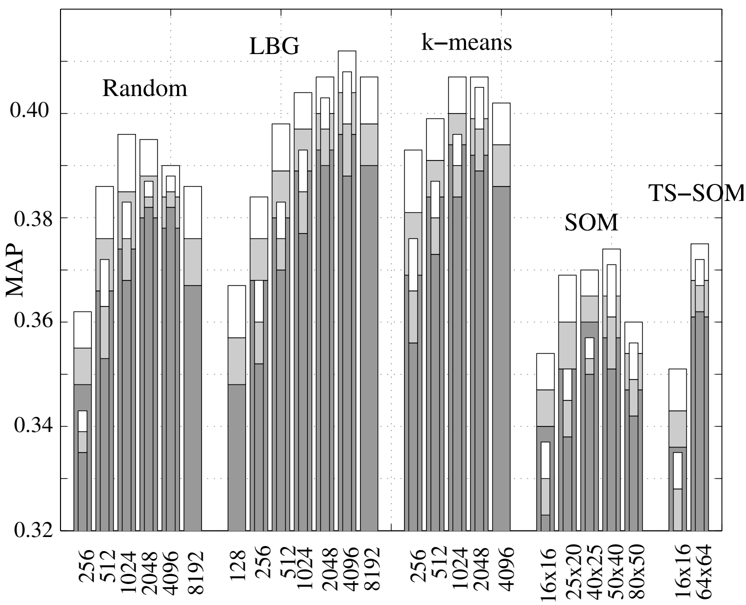


Fig. 4. MAP performances resulting from unsupervised codebook selection, grouped by selection algorithms

codebooks the histogram bin counts are clearly not binary, as has sometimes been suspected.

Table 3 shows the MAP ranking accuracies resulting from the use of the different codebooks. The difference between column “norm” and “nonorm” is that for the former the histograms have been normalised within each image to sum to one, whereas for the latter column such normalisation has not been

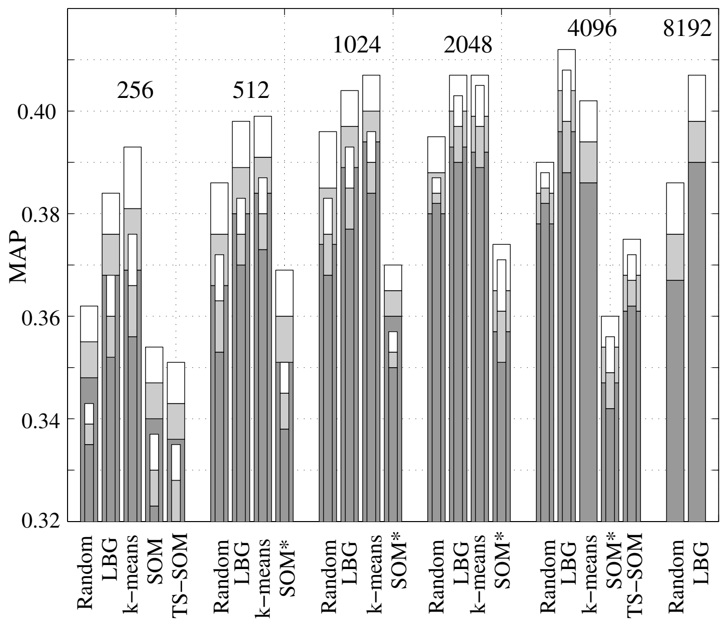


Fig. 5. MAP performances resulting from unsupervised codebook selection, grouped by codebook size. The SOM codebooks marked with an asterisk (*) are only approximately of the indicated size.

performed. In light of these results such normalisation does not appear useful. Figures 4 and 5 illustrate the same results pictorially. In the former the methods are grouped by the codebook selection algorithm, in the latter by equal codebook size. The lighter shades on top of the bars indicate 95% confidence intervals. The narrow bars correspond to MAP of the normalised codebooks, the broader bars to MAP of the unnormalised ones.

5 Experiments on Supervised Codebook Selection

In this section we report our results from the supervised generation of codebooks. Some of the results are, however, still more suggestive than definitive as we have not yet been able to perform full statistical testing of some of the presented alternatives due to the requirement to generate codebooks separately for each of the 20 object classes and each of the six partitionings to training and test data.

We implemented three alternative methods for supervised selection of codebooks. In the first two, a separate codebook was generated for each of the 20 object classes. In the first method we synthesised the codebooks by merging class-specific codebooks with background codebooks of equal size. Both codebooks were formed with the LBG algorithm. The background codebook originated from the unsupervised experiments of Sect. 4. For forming the class-specific codebooks

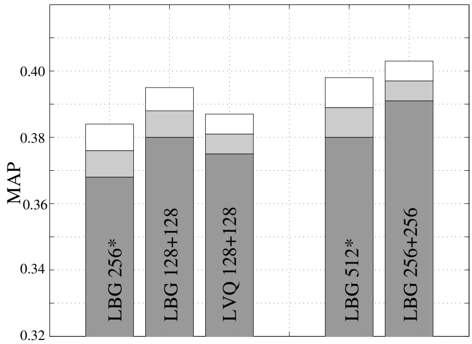


Fig. 6. MAP performances resulting from supervised codebook selection. The performance of unsupervisedly selected LBG codebooks of the same sizes are included for comparison and are marked with an asterisk (*).

Table 4. MAP performance resulting from supervised codebook selection. The results from the unsupervised LBG clustering are included in the table and marked with an asterisk (*).

Six times randomised			Single experiment		
Algorithm		MAP	Algorithm		MAP
LBG	128+128	0.388±0.007	LBG	128+128	0.379
	256+256	0.397±0.006		256+256	0.392
LVQ	128+128	0.381±0.006		512+512	0.391
LBG (*)	256	0.376±0.008	LVQ	128+128	0.376
	512	0.389±0.009		256+256	0.396
	1024	0.397±0.008	LBG	20X128	0.395
				20X256	0.393
				20X512	0.395

we used the interest points from the class images only, sampled down to 150 interest points per image. In the second method the codebooks were obtained by using the LVQ algorithm on the training set in the one-class-against-all-others fashion. For both the class and its complement the initial set of prototype vectors was obtained with the LBG algorithm. The LVQ algorithm was then run to refine the prototypes. The last one of the three methods comprised of first forming LBG codebooks for each of the 20 classes and then merging the codebooks.

Table 4 lists the MAP performances resulting from the use of supervised codebook selection techniques. All the results are for unnormalised histograms as these were found to outperform normalised ones in the experiments of Sect. 4. On the left of the table are more reliable results that could be statistically tested, on the right results of only suggestive value. The results are illustrated also in Fig. 6.

6 Conclusions and Discussion

Every experiment was not performed with exactly every algorithmic alternative, therefore the matrices in the result tables contain empty slots. However, we think the results are comprehensive enough in order to be able to make justified conclusions. The performance improvement from the worst to the best performing codebook—23% from 0.328 to 0.404—is very clear and far beyond the observed statistical fluctuation. Investing some thought for codebook selection is therefore justified. For instance, in the recent PASCAL VOC Challenge 2007 image classification evaluation, performance changes that large would have affected the rankings of the participating methods significantly.

Using our classification methodology, the use of rather large codebooks seems to be beneficial. The largest of the tested codebooks could have reached the size where the performance starts to saturate and finally degrade. This cannot be concluded very definitely as the codebook sizes growing in powers of two make the larger codebook sizes only sparsely covered. Practical reasons of computational time complexity, and to a degree also memory space concerns, make the use of this large codebooks somewhat cumbersome—both during the codebook generation itself, but also during histogram generation when the closest prototype vectors in the codebooks have to be searched. Also this makes comprehensive experiments with large codebooks problematic. One of the next steps could be finding approximate solutions without essential reductions in the classification accuracy.

Of the considered unsupervised clustering techniques the k-means and LBG algorithms produced the best performing codebooks, with the k-means codebooks probably performing slightly better. The difference is, however, within the possibilities of statistical fluctuation. These two clustering algorithms also perform best in terms of the quantisation error. The quantisation error and the histogram quality are, however, not necessarily in direct correlation. This was one of the motivations for including algorithms such as SOM and TS-SOM in this comparison, despite the fact that they are known to compromise quantisation error for topology preservation. These algorithms generate codebooks with the classification performance worse than that of randomly selected codebooks, in spite of clearly smaller quantisation error. This demonstrates the non-existence of direct link between quantisation error and histogram quality. One direction for future experiments would be the investigation of additional clustering algorithms in order to better understand the qualities of the clusterings that affect the usefulness of the generated histograms.

The tested supervised codebook generation methods were quite elementary. The class-background LBG method and the refinement of the LBG codebooks with LVQ provide, however, some performance improvement over unsupervised methods. The improvement is not very drastic and may not always be worth the added complication of generating separate codebooks for all the object classes to be recognised. Our experiments failed to reveal significant differences between the performances of the LBG and LVQ methods although the results might be interpreted to hint that the LVQ-based approach could perform slightly worse. In this light refining the LBG codebooks with the LVQ algorithm does not appear appealing especially as it is computationally rather costly. One possible reason for the failure of LVQ could be that it considers interest points just one by one. It may be, however, that individual interest points are not characteristic enough to function as a basis for object classification. Constellations of several interest points might be necessary for this. In our experiments, the concatenations of all the 20 class-specific codebooks did not seem to bring improvements over well-selected unsupervised codebooks of comparable sizes.

One possible direction for future research could be studying the possibility of performing variable selection on the bins of the histograms once the histograms have been extracted. However, this might be hindered by the same reasons of individual interest point types not being characteristic enough for an object class.

References

1. Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A.: The PASCAL Visual Object Classes Challenge (VOC2007) Results (2007), `http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html`
2. Kohonen, T.: Self-Organizing Maps, 3rd edn. Springer Series in Information Sciences, vol. 30. Springer, Berlin (2001)
3. Koikkalainen, P., Oja, E.: Self-organizing hierarchical feature maps. In: Proc. International Joint Conference on Neural Networks, San Diego, CA, USA, vol. II, pp. 279–284 (1990)
4. Linde, Y., Buzo, A., Gray, R.M.: An algorithm for vector quantizer design. IEEE Transactions on Communications 28(1), 84–95 (1980)
5. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
6. Mikolajcyk, K., Schmid, C.: Scale and affine point invariant interest point detectors. International Journal of Computer Vision 60(1), 68–86 (2004)
7. Schalkoff, R.J.: Pattern Recognition: Statistical, Structural and Neural Approaches. John Wiley & Sons, Ltd (1992)
8. Sivic, J., Zisserman, A.: Video Google: A text retrieval approach to object matching in videos. In: Proc. of ICCV 2003, vol. 2, pp. 1470–1477 (2003)
9. TREC 2006, common evaluation measures, Online proceedings of 15th Text Retrieval Conference (TREC 2006), (2006), `http://trec.nist.gov/`
10. Viitaniemi, V., Laaksonen, J.: Improving the accuracy of global feature fusion based image categorisation. In: Falcidieno, B., Spagnuolo, M., Avrithis, Y., Kompatsiaris, I., Buitelaar, P. (eds.) SAMT 2007. LNCS, vol. 4816, pp. 1–14. Springer, Heidelberg (2007)
11. Zhang, J., Marszałek, M., Lazebnik, S., Schmid, C.: Local features and kernels for classification of texture and object categories: a comprehensive study. Int. J. of Computer Vision 73(2), 213–238 (2007)

Extending KDDML with a Visual Metaphor for the KDD Process

Valerio Grossi and Andrea Romei

Department of Computer Science, University of Pisa
Largo Bruno Pontecorvo 3, 56127 Pisa, Italy

Abstract. The spreading application of data mining techniques is clearly represented by the large number of suites supporting the knowledge discovery process. The latter can be viewed as real visual programming environments. Based on this assumption, we define some requirements which a typical data mining high-level graphical user interface should satisfy, in order to guarantee a good level of interactivity and expressiveness. The aim of this study is to use these requirements during the engineering and development of visual knowledge flow abstraction for the existing KDDML (Knowledge Discovery in Databases Markup Language) system. We introduce some features not only directly related to the visual metaphor, but also to the whole system, here intended as a real visual programming environment for the knowledge discovery process.

1 Introduction

The growing interest for Data Mining (DM) is clearly represented by the large number of tools supporting different approaches to Knowledge Discovery in Databases (KDD). Several systems are easily available on the web, many of which offer a visual metaphor to guide the development of complex data analysis processes. These softwares, which can be either free or commercial, are complex suites including statistical and data preprocessing components, DM techniques, visualization and evaluation tools, and can be viewed as real visual programming environments.

Our aim is to implement a visual language for the KDDML system [1,2] in compliance with some of the requirements outlined and examined in detail in this contribution. We shall not only concentrate on graphical, representational characteristics, but we shall also propose a detailed survey of the KDDML system architecture, viewing some aspects not only directly related to its visual appearance, but also to the whole system, intended as a programming environment for KDD.

In the remaining of this section, the KDDML system is outlined, while in section 2, we introduce some requirements related to several aspects of a KDD GUI environment, and we evaluate it from different perspectives. In section 3, we present how these requirements are implemented in KDDML. Finally, in section 4, we conclude examining some well-known existing systems and proposing our final considerations.

M. Sebillo, G. Vitiello, and G. Schaefer (Eds.): VISUAL 2008, LNCS 5188, pp. 138–149, 2008.

1.1 Background: KDDML as a Middleware Language

KDDML (Knowledge Discovery in Databases Markup Language) is a middleware XML-based language (and system) whose aim is to support the development of final KDD applications, in which a mixture of database access, data preprocessing, mining extraction and deployment is required. The term *middleware* means that higher abstraction levels must be built on top of it, such as vertical applications, high-level GUIs, or more declarative languages. The term *XML-based* refers to the fact that data, patterns (i.e. mining models) and KDD queries are represented through XML documents.

XML tags in KDDML correspond to operators with a functional semantics that allows nesting of KDD tasks, thus permitting to view the KDD as a process. The XML syntax of a generic operator is shown below:

```
<OPERATOR_NAME att1="v1" ... attM="vM">
      <ARG1_NAME> .... </ARG1_NAME>
      ...
      <ARGn_NAME> .... </ARGn_NAME>
</OPERATOR_NAME>
```

The attributes correspond to parameters of the operator (e.g., the target attribute of a tree miner algorithm). Sub-elements `ARG1_NAME`, ..., `ARGn_NAME` define arguments passed to the operator. As it can be expected, arguments must be of an appropriate type and sequence, i.e. an operator signature must be specified. Intuitively, we can have different types for data sources, for each mining model and for operators returning a scalar.

KDDML is implemented in Java and has a three-layer architecture. Each layer implements a specific functionality and supplies an interface to the layer above. The bottom repository layer manages the read/write access to repositories of data and models and provides programmatic read/write access to the model contents. The operators layer above is composed of the implementations of language operators. The interpreter layer accepts a validated KDDML query and recursively traverses it, returning the result to the top layers. For a complete description of KDDML and further details, see [1] on the system architecture, and [2] on the language specification.

Running example. Let us consider the problem of building a voting classifier among three classification trees. The KDDML query is reported in figure 1. The construction of the first tree (first tag <TREE_MINER>) takes place on a training set given as ARFF file (tag <ARFF_LOADER>), by applying the ID3 decision tree induction algorithm. Before applying the algorithm on the *weather* dataset, the `temperature` attribute is discretized into three intervals of equal width, by using a natural binning discretization method (tag <PP_NUMERIC_DISCRETIZATION>)[1]. The second tree is obtained again by means of the ID3 algorithm (second tag <TREE_MINER>) on new data loaded from the repository (tag <TABLE_LOADER>). We assume that the `temperature` attribute has been previously discretized.

[1] Recall that the ID3 algorithm does not work with numeric attributes.

```
<KDD_QUERY name="sample">
 <TREE_VIEWER style_sheet="tree.xsl">
   <TREE_META_CLASSIFIER xml_dest="weather_meta.xml"
                         combination_type = "committee">
      <TREE_MINER target_attribute="play">
         <PP_NUMERIC_DISCRETIZATION attribute_name="temperature">
            <ARFF_LOADER arff_file_name="weather.arff"/>
            <ALGORITHM algorithm_name="natural_binning">
               <PARAM name="number_of_intervals" value="3"/>
            </ALGORITHM>
         </PP_NUMERIC_DISCRETIZATION>
         <ALGORITHM algorithm_name="ID3"/>
       </TREE_MINER>
       <TREE_MINER target_attribute="play">
         <TABLE_LOADER xml_source="weather.xml"/>
         <ALGORITHM algorithm_name="ID3"/>
       </TREE_MINER>
       <TREE_LOADER xml_source="weather_tree.xml"/>
   </TREE_META_CLASSIFIER>
 </TREE_VIEWER>
</KDD_QUERY>
```

Fig. 1. A sample KDDML query

Finally, the third tree is loaded directly from the models repository (tag <TREE_LOADER>). The resulting XML voting classifier (tag <TREE_META_CLASSIFIER>) is transformed into a HTML browsable format via a XSL style sheet (tag <TREE-_VIEWER>).

2 Requirements for a Knowledge Discovery Graphical User Interface

In this section, some visual requirements that a KDD tool should satisfy, in order to guarantee a good level of interactivity and flexibility, are introduced. Furthermore, we present requirements dealing with the whole system and not strictly related to the graphical metaphor.

For the sake of brevity, in the present contribution, aspects such as *data source* availability, *scalability*, as well as specific features of each software component are ignored. We concentrate only on the visual metaphor and the services that a KDD GUI supplies to the user.

2.1 Visual Aspects

When developing a KDD process, the user should have a powerful graphical abstraction that simplifies the design, without limiting expressiveness. The most common abstractions are typically two, namely *Tree-like* and *Graph*.

In GUIs adopting the former abstraction, all the components involved in the analysis are arranged in a tree-like structure, in which the parent component passes data to its children. This kind of representation is quite "tidy" and systematic, it forces the user to follow the steps from data input, typically the root of the tree, to result visualization.

The way for representing the KDD process through this metaphor is clear and easy to understand, but its intrinsic simplicity complicates all the cases when a component needs more than one input. Using these components and observing the tree, only an input passed by the father component is clearly visible. Other implicit inputs are typically specified in the component parameter dialog box.

The graph abstraction represents the KDD process as a directed acyclic graph, where the nodes are operators, and the edges represent the knowledge flow from a node to another. Differently from tree-like abstractions, this one enables the explicit representation of all operators receiving data from more than one node. From a graphical point of view, this kind of metaphor is more powerful and expressive than the previous one. It leaves more freedom to the user, but it has the disadvantage of not guiding the user during the development. Compared to tree-like visualization, the user is not implicitly constrained to follow a set of predefined phases, such as data input, preprocessing, extraction, valuation and so on.

A KDD visual metaphor must supply several controls over component parameters, components linking, as well as the subdivision of the operators in phases. From a visual perspective, a GUI has to drive the user as much as possible on error tracking during the process construction, for example highlighting a "not configured" or "wrong configured" component.

***Req. Visual Metaphor*:** *A KDD system must supply a graphical metaphor to represent the KDD process. It has to be user-friendly and must supply a good level of interactivity.*

2.2 Visual Programming Support

The GUI of DM suites represents a sort of environment for KDD visual programming and as such, it has to guarantee some features typically related to a development environment.

Similarly to usual programming languages, we can have *static analysis*, which concerns all the controls to check syntactic and typing correctness. Natural limitations of static analysis can be partially solved, introducing more complex methods inherited from abstract interpretation techniques, which involve some information typically available at run-time.

From a logical point of view, data can be divided into two well distinct subclasses, namely *meta-data* and *physical-data*. The former describes "data morphology", such as the number and types of attributes really available on physical-data. Based on this broadly accepted feature, we can introduce the concept of *meta-execution*, where the system runs the process using only the available meta-data. The introduction of this control does not negatively affect the overall performance of the system, since meta-data are very small compared to physical ones.

Let us consider the example introduced in sect. 1.1. If we want to use a classifier committee, to obtain a correct result, all the classifiers involved must be extracted from training sets sharing the same meta-data. Before extracting the actual model, a preprocessing component is involved. Preprocessing is one of the most computationally expensive KDD tasks, and it represents a good example, where a component can modify not just physical-data. Anytime a preprocessing phase is involved, without meta-execution, we cannot guarantee that an attribute will continue to exist, or will have the same initial type. In the cited example, the `temperature` attribute changes its type from *continue* to *discrete*.

An error in this phase can invalidate the entire process, with significative loss of time due to data access and transformation. Given data inputs and using meta-data, we can virtually execute the user task without handling physical data.

Meta-execution can also be used to improve the checking and visualization of operators input parameters. Thus, it is related to visual abstraction. Accessing meta-data, we can provide more information on selected data, on the effective number, name and type of attributes available. It is important to point out that meta-execution is heavily influenced by the information available on meta-data.

***Req. Meta-execution*:** *Advanced techniques to guarantee the correctness of a KDD process during its construction must be provided, without accessing physical-data.*

2.3 Extensibility

Since its introduction, DM (and in general KDD) has been characterized by a continuous development of new algorithms and techniques. Until recently, DM systems provided only command-line interfaces and rarely offered some integration with other frameworks. Today, on the contrary, DM suites should be engineered to ensure an easy extensibility of the whole system.

This implies that the GUI level of the framework has not to be modified by the user every time a new operator is introduced. The architecture of the system should be conceived in such a way that adding a new operator does not imply any change to the GUI, but the newly inserted operator is automatically available at graphical level.

The GUI level must be independent from the actual algorithms and tools implemented in the system. It has to retrieve all the necessary information on the components (and their parameters) available, and as a consequence the system architecture must provide a common way to define all the necessary information, which must be independent from the actual implementation of the GUI level.

This characteristic is quite important, because we can assume that a data miner is able to write an algorithm, also maintaining a kind of programming interfaces, but we cannot assume that a user will be able to change the GUI level implemented by other programmers.

***Req. Extensibility*:** *The GUI level must be independent from the actual status of the system. The system must provide a common well-defined way to extend itself, guaranteeing all the necessary information at the GUI level.*

3 How Does KDDML Fit the Visual Requirements?

In this section we investigate the extent to which KDDML is capable of satisfying the features presented in the previous section. We recall that this paper is devoted to provide general guidelines that can be taken into account in the design of a visual metaphor for the KDD process. Figure 2 shows a snapshot of the GUI.

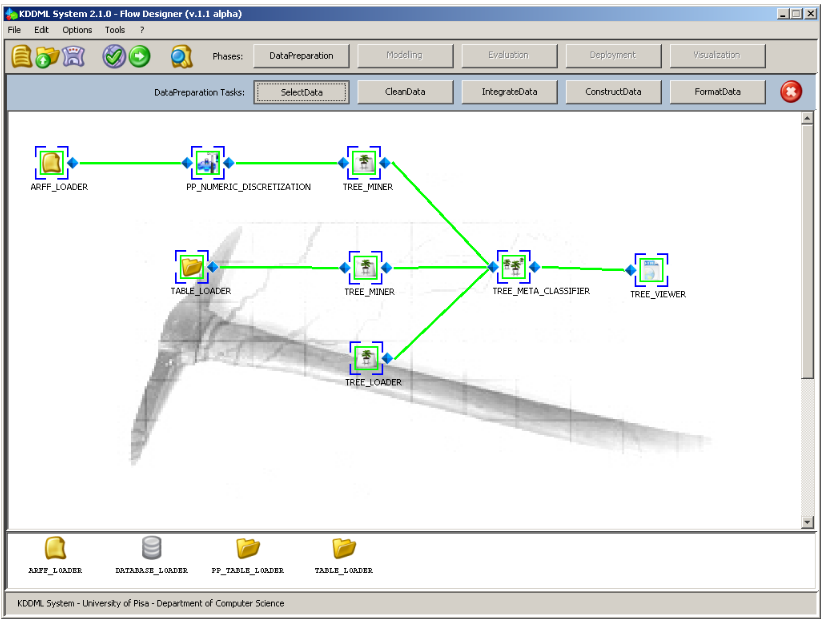
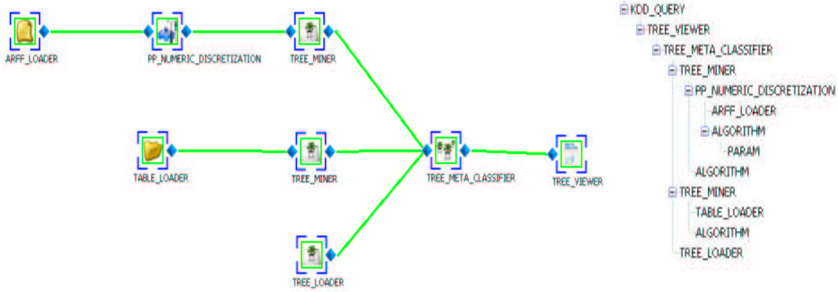


Fig. 2. The Flow Designer interface of KDDML

3.1 To Graph or Not to Graph

In implementing a high-level GUI for KDDML, we use a graph to represent the knowledge flow. The user selects components (i.e. KDD tasks) from a tool bar and connects them into a directed acyclic graph that processes and analyzes data. In our opinion, the graph represents a good trade-off between user-friendly usability and expressiveness. In fact, the visual metaphor of the KDD process as a graph of tasks (nodes) and flows (arcs) allows an intuitive definition of the KDD process and rapid prototyping.

In addition, a graph view is suitable for the KDDML language design for two main reasons. On the one hand, several KDDML tasks take inputs from different sources, due to the functional semantics of operators; e.g. the <TREE_META_CLASSIFIER> operator returns a voting classifier among a set of classification



(a) (b)

Fig. 3. The Knowledge Flow (a) and the DOM tree (b) of the query of fig 1

trees. On the other hand, the XML DOM[2] representing the query can be easily obtained by means of a right-to-left exploration of the graph. As an example, figure 3b shows the DOM representation of the query in example 1. Figure 3a depicts the KDDML knowledge flow as a graph. Notice that, the <ALGORITHM> XML tags are directly encapsulated as parameters of the <TREE_MINER> and <PP_NUMERIC_DISCRETIZATION> nodes, and they do not appear in the knowledge flow of figure 3a.

The definition of a KDD query partially follows the CRoss Industry Standard Process for Data Mining (CRISP-DM) [3]. It describes the DM methodology in terms of a hierarchical process model, consisting of sets of tasks at four levels of abstraction, namely phases, generic tasks, specialized tasks and process instances. This kind of solution also favors usability, when the number of available operators increases: the key idea is that operators belonging to a KDD phase/task are available whenever a user deals with that phase/task. As one could expect, a good trade-off between the number of phases and tasks and the total number of KDDML operators is required. In our design, the construction area has been divided into five main phases, i.e. data preparation, modeling, evaluation, deployment and visualization. Each phase breaks down the objective into generic tasks. Specifically, the data preparation phase includes data selection, data cleaning, data integration, data construction and data transformation. A third abstraction level consists in KDDML operators, e.g. data selection contains operators to remove attributes and instances or to sample data. Moreover, also the iterative nature of the KDD process is underway, for example, whenever a data preprocessing operator is applied to a result of a model application operator.

Finally, we can think our visual language as embedded into a user-friendly interface that compiles the KDD process into one or more KDDML queries. Basically, let us imagine the XML document of figure 1 enriched with XML elements and attributes that encapsulate visual information relative to the knowledge

[2] DOM (Document Object Model) provides a main-memory hierarchical object-oriented model of XML documents. The DOM tree of the KDDML query is directly processed by the KDDML interpreter in executing the query. See [1] for details.

flow; e.g. the coordinates of each node in the construction area. With this kind of solution, the KDDML interpreter can directly run the KDD query created with the visual interface: XML attributes and elements representing visual information are ignored during the parsing of the KDD query.

3.2 KDDML and Meta-execution

As mentioned earlier, KDD objects can be considered from two distinct points of view: the physical level and a higher abstraction logical level. Classical example of logical elements is the type of attributes (discrete, continuous, etc.) with weights that make sense in affecting the role of an attribute in the construction of a model. Generally, the logical level involves both source data and knowledge. For example, concerning knowledge, it can be useful to keep inside the model information about the logical schema of the relation used to build the model itself.

The PMML (Predictive Mark-up Modeling Language) standard [4] is used by KDDML to represent models as XML documents. PMML offers a typical example of meta-schema definition. Each PMML model is always composed by:

- A *data dictionary* containing the definitions of the fields used in the mining model, such as the type and value range;
- A *mining schema* that lists fields used by the model;
- Statistical information concerning the input dataset and the parameters used to build the model;
- The actual model.

The first three elements define the logical schema of the model, in addition to the physical one.

The same feature is achieved by KDDML for the meta-data of a data source. In KDDML, the data schema is stored as an XML file. It contains information about each attribute, which cannot be automatically derived without physical access from the attribute values. As an example, the following XML fragment contains basic statistics of the `weather` relation as used in the running example:

```
<SCHEMA log_name="weather" num_of_attributes="5" num_of_instances="20">
  <ATTRIBUTE name="outlook" num_of_missing_values="2" type="nominal">
     <NOMINAL_DESCR num_of_values="3">
        <VALUE value="rainy" cardinality="5"/>
        <VALUE value="sunny" cardinality="8"/>
        <VALUE value="overcast" cardinality="5"/>
     </NOMINAL_DESCR>
  </ATTRIBUTE>
  ...
  <ATTRIBUTE name="temperature" num_of_missing="0" type="numeric">
     <NUMERIC_DESCR mean="74.0" std_dev="35.3" min="64.0" max="85.0"/>
  </ATTRIBUTE>
</SCHEMA>.
```

Basically, in KDDML the physical component of each object is well-separated from the meta component. Meta-data and meta-knowledge are the so called meta-objects that do not include physical instances, nor the physical components of a model. Under this assumption, meta-execution, as described in sect. 2.2, acts by means of a method, namely `m-execute`, implemented in each language operator class, that takes a list of meta-objects as input and returns a meta-object as output. The method incorporates the control of correctness that, at this level, is available by scanning meta-objects. We have already provided several examples of such controls in sect. 2.2.

From the graphical point of view, this kind of meta-execution is transparent for the user. Whenever two nodes (i.e. operators) are connected by an arc (i.e. flow), meta-execution acts implicitly and the meta-information (i.e. the result of the `m-execute` method) is directly available for the target node. This information is used to validate the rest of the query. As an instance, let us consider a case in which one of the decision trees does not share the same meta-data, as in the running example of sect. 1.1. As a consequence, a warning is raised as soon as the user tries to create an arc from such `TREE_MINER` node to the `TREE_META_CLASSIFIER` father (see fig. 3).

In our opinion, this feature can increase the effectiveness of an environment in support to the KDD process and a high level visual knowledge flow must capture it. Obviously, meta-execution suffers from some limitations as well. For example, we are not able to check whether the number of leaves of an extracted tree satisfies a particular constraint, since the number of leaves is part of the physical model.

Currently, our KDD group is involved in a rigorous definition of such a study. Related works on meta-learning can be found in [5,6].

3.3 KDDML and Extensibility

In our view, a KDD architecture structured in layers is a design that favors extensibility. A visual language for the KDD process is thus a natural higher abstraction layer. Figure 4 depicts the main packages related to the four layers,

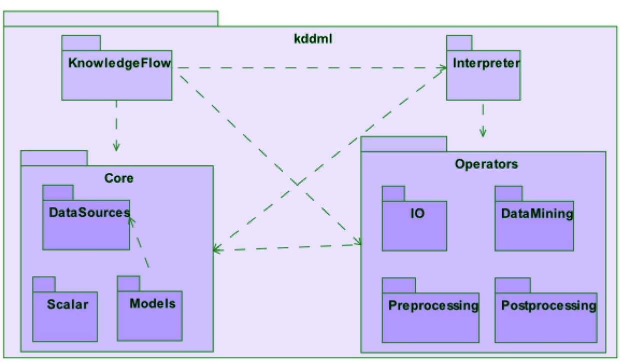


Fig. 4. The dependencies between the top-level KDDML packages

and above all, it shows the dependencies among them. It is important to note that the `KnowledgeFlow` package (i.e. the GUI layer) uses functionality of the `Core`, `Operators` and `Interpreter` packages, but not viceversa. As mentioned in sect. 1.1, the core layer encapsulates details on data sources and mining models and the operator layer encapsulates details on mining operators and algorithms. The GUI layer captures this information by means of Java interfaces.

For example, <OPERATOR_NAME> is implemented as a Java class satisfying the `KDDMLOperator` interface, which requires, among the others, the following methods used by the GUI layer:

- `KDDObjectType getArgumentTypeAt(int i)`, returns the expected type of the ith argument of the operator, where `KDDObjectType` is an object at the core layer over an enumeration of all types;
- `KDDObjectType getResultType()`, returns the output type of the operator;
- `KDDPhase getKDDPhase()`, returns the phase of the KDD process of the operator, where `KDDPhase` is an object at the core layer over an enumeration of CRISP-DM phases.
- `KDDTask getKDDTask(KDDPhase phase)`, returns the task given a particular phase. `KDDTask` is an object at the core layer over an enumeration of a set of possible tasks.
- `ScalarType getParameterTypeAt(int i)`, returns the expected type of the ith parameter of the operator, where `ScalarType` is an abstract class placed at the core layer, that incorporates the type related to a scalar value. Subclasses include for example string values, numeric ranges, and enumeration over a list of possible values.

These methods are invoked by the top GUI layer whenever details on operators and parameters are required. As one could expect, each KDDML object (i.e. models and data sources) provides details by means of similar interfaces placed at the core layer. The `KnowledgeFlow` package is not affected by the number (and type) of mining models, operators or algorithms, since they are loaded at run-time via Java reflection.

4 Conclusions

4.1 Related Work

Several modern DM and KDD tools[3] have been developed in the last few years. Most of them use a visual programming approach which makes them particularly appealing to data analysts and users even without a computer science background. These tools are both *commercial* and *academic* softwares.

Some examples of well known commercial tools are `Clementine` [8], which supports the industry-standard CRISP-DM methodology [3] and adopts a *graph* graphical abstraction, `DB2 Intelligent Miner` [9], `Oracle DataMining` [10] and `GhostMiner` [11]. The last three are cases of *tree-like* abstraction. All the

[3] A complete list of commercial and free KDD (and DM) suites can be found in [7].

cited commercial suites provide several DM and exploratory analysis tools, including a wide range of data sources, direct access to datawarehouses and databases, data cleansing, and so on.

On the academic side, `WEKA` (Waikato Environment for Knowledge Analysis) [12] is one of the most widespread software for knowledge discovery. It includes a huge quantity of algorithms and essentially supports all the phases of the KDD process, adopting a graph metaphor on `KnowledgeFlow` environment.

Another free academic system with graph metaphor is `Orange` [13], which represents a real visual programming interface based on visual components. Using `Orange Canvas`, we can develop a real stand-alone KDD application, without writing any code line. It provides a simple visual tool, that even a non-programmer user can employ to develop DM applications.

An example of free software using tree-like structure, coupled with a XML process representation, can be found in `RapidMiner` [14], formerly YALE (Yet Another Leaning Environment). It can be applied to a wide range of new rising DM applications including text, multimedia mining, data stream mining, and distributed data mining, where its rapid prototyping abilities are demonstrated.

Another DM software for academic and research purposes, adopting tree-like process visualization, is represented by `Tanagra` [15].

At the best of our knowledge none of the academic free systems mentioned support meta-execution. An exhaustive analysis and comparison on the cited softwares can be found in [16].

4.2 Final Remarks

In this paper, we presented our experience in the design and development of a Knowledge Discovery Graphical Interface for the KDDML system. The design principles have been motivated by basic requirements concerning the graphical metaphor, a visual programming support and an easy extensibility of the GUI level.

Some choices, such as the adoption of a graph-based metaphor - against a tree-like visualization - have been mainly motivated by design reasons of an XML-based functional language. XML seems appropriate in representing meta-objects, enabling the validation of a KDD process during its construction, without accessing to physical data/models. Finally, a layered architecture also allows easy plugging-in of new models, algorithms and operators making the GUI level independent from these extensions.

KDDML is a prototype with further optimizations and enhancements pursued. The KDDML team at the Department of Computer Science - University of Pisa - is currently involved in the definition of instruments enabling the visualization of PMML models, in style of [9]. A more rigorous definition of an environment capable of dealing with both logical data and meta-knowledge is also underway.

The Flow Designer interface presented in this paper has been recently integrated at the top of the KDDML architecture.

References

1. Romei, A., Ruggieri, S., Turini, F.: KDDML: a middleware language and system for knowledge discovery in databases. Data Knowl. Eng. 57(2), 179–220 (2006)
2. Romei, A., Ruggieri, S., Turini, F.: KDDML: a middleware language and system for knowledge discovery in databases. In: Proceedings of the 13th Italian Symposium on Advanced Database Systems (SEBD 2005), Brixen, Italy, pp. 208–219 (2005)
3. SPSS: CRISP-DM step by step data mining guide. Version 1.0 (2000), www.crisp-dm.org
4. The Data Mining Group The Predictive Model Markup Language (PMML). Version 3.2., http://www.dmg.org/pmml-v3-2.html
5. Bertino, E., Catania, B., Maddalena, A., Skiadopoulos, S., Terrovitis, M., Vassiliadis, P.: Modeling and language support for the management of pattern-bases. In: Proceedings of the 16th International Conference on Scientific and Statistical Database Management (SSDBM 2004), pp. 265–274 (2004)
6. Bernstein, A., Provost, F., Hill, S.: Toward intelligent assistance for a data mining process: An ontology-based approach for cost-sensitive classification. IEEE Transactions on Knowledge and Data Engineering 17(4), 503–518 (2005)
7. KDnuggets - Software Suites for Data Mining and Knowledge Discovery, http://www.kdnuggets.com/software/suites.html
8. Clementine (Home Site), www.spss.com/clementine/index.htm
9. DB2 - Intelligent Miner: Home Site, www-306.ibm.com/software/data/iminer/
10. Oracle Data Mining: Home Site, www.oracle.com/technology/products/bi/odm/
11. GhostMiner: Home Site, www.fqs.pl/business_intelligence/ghostminer
12. Weka3: Data Mining Software in Java, www.cs.waikato.ac.nz/ml/weka/
13. Orange Data Mining Fruitfull&Fun Home Site, http://magix.fri.uni-lj.si/orange/
14. Euler, T., Klinkenberg, R., Mierswa, I., Scholz, M., Wurst, M.: YALE: rapid prototyping for complex data mining tasks. In: KDD 2006: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 935–940. ACM Press, New York (2006)
15. TANAGRA project: Home Site, chirouble.univ-lyon2.fr/~ricco/tanagra/
16. Grossi, V.: Linguaggi grafici per knowledge discovery (in italian only). Master's thesis, University of Pisa (2005)

Inter-frame and Inter-layer Motion Prediction during Fast Block Motion Estimation in MCTF

A.K. Karunakar and M.M. Manohara Pai

Department of Information and Communication Technology
Manipal Institute of Technology
Manipal 576 104 India
karunakar.ak@manipal.edu,mmm.pai@manipal.edu

Abstract. The motion compensated temporal filtering (MCTF) using 5/3 LeGall bi orthogonal wavelet filter requires bidirectional motion estimation (ME) that is computationally intensive. Hence encoder of Scalable Video Coder(SVC) becomes very slow. This paper proposes Motion Prediction technique for Fast Block Based ME during MCTF in SVC. The technique exploits center biased characteristics of motion vectors. It applies coarse and fine search of any fast block based ME, only to the first pair of frames in a group of pictures (GOP). The generated motion vectors (MVs) are supplied to the next consecutive frames and even to subsequent temporal levels, that is used to initiate the fine search. The technique significantly reduces the number of blocks that undergoes ME in a GOP and hence reduces computational complexity of ME in MCTF. The proposed algorithm is implemented in MC-EZBC and there is no visible variations in the quality of the decoded video.

Keywords: Motion Estimation, Motion Compensated Temporal Filtering, MC-EZBC, Scalable Video Coding, Scalable Video Storage.

1 Introduction

The Scalable Video Coding (SVC) is one of the most important features of modern video communication and storage system. Application areas of SVC range from multimedia messaging,video conferencing over mobile TV, Internet video streaming, to storage medias like DVD, Blu-ray disc and HD DVD [1]. For a truly scalable coding, the encoder needs to operate independently from the decoder, while in predictive schemes the encoder has to keep track and use certain information from the decoder's side (typically target bit-rate), in order to operate properly.

The idea of using motion-compensated temporal DWT (i.e. MCTF) was introduced by Ohm [4] and developed by Choi and Woods [5]. Temporal decomposition using any desired motion model and any desired wavelet kernel with finite support is possible using lifting framework for motion compensated temporal DWT (MC TDWT). The results reported in [6][7] indicate superior performance of bi-orthogonal 5/3 wavelet kernel during temporal filtering, compared to conventional Haar wavelet transform.

M. Sebillo, G. Vitiello, and G. Schaefer (Eds.): VISUAL 2008, LNCS 5188, pp. 150–154, 2008.

Motion estimation and Motion compensation is of crucial importance in order to obtain good performance in video compression [2], be it the classical hybrid coding [3] or one of the newer wavelet-based algorithms [4]. There are many fast block based ME algorithms. This paper proposes a novel inter-frame and inter-layer prediction technique that exploits high correlation between frame/levels in a GOP during MCTF.

The rest of the paper is organized as follows. Section 2 discusses proposed technique. The section 3 is experimental results and 4 concludes the paper.

2 Inter-Frame and Inter-layer Motion Prediction

In any video due to its inherent nature, the frames along the temporal directions are highly correlated. Due to center-biased global minimum motion vector distribution characteristics, more than 80% of the blocks can be regarded as stationary or quasi-stationary blocks [9]. The ME in any video coding model is computationally intensive and can consume up to 80% of computational power of the encoder if all possible search positions are evaluated exhaustively (i.e. full search (FS) which gives global minima). Hence various fast block based algorithms (BMA) were proposed, almost all algorithms work in two stages, a low-resolution coarse search followed by a fine-resolution inner search. Even wavelet video coders uses the same ME algorithms as that of hybrid video coder mentioned above. There is a very high correalation among the frames in a GOP of MCTF and that can be well exploited by using suitable prediction technique. The proposed predictive technique in this paper uses HEXBS algorithm as an example and that can be extended to all types of fast ME.

In the proposed technique MVs between the first pair of frames in a GOP have to be found as usual with low-resolution coarse search and fine-resolution inner search of any fast ME technique. During ME between next consecutive frames and even in next temporal levels in a GOP, the previously available MVs are used as a starting search point for fine search. The predictive structure used in this work is shown in Fig. 1(a).

Average gain in search points: In order to simplify the analysis, we assume that the starting search point is always at the $(0, 0)$ position. In the proposed predictive algorithm only inner fine search of fast ME is applied for all the pair of frames except the first in a GOP and the number of search point's remains same (i.e. 4 search points per block) at all positions. Hence average gain in terms of number of search points over the conventional by proposed technique is given by,

$$\overline{G} = \sum_{i=-M}^{M} \sum_{j=-M}^{M} [n_{s,Conv\ *}(i,j) - n_{s,Pred\ *}(i,j)] \times P_s(i,j) \quad (1)$$

where, $(2M+1) \times (2M+1)$ - is the area where we assume that the probability of having MV is nonzero, $n_{s,Conv\ *}(i,j)$ and $n_{s,Pred\ *}(i,j)$- is the number of search points for conventional and predictive algorithms respectively, $*$ - indicates fast

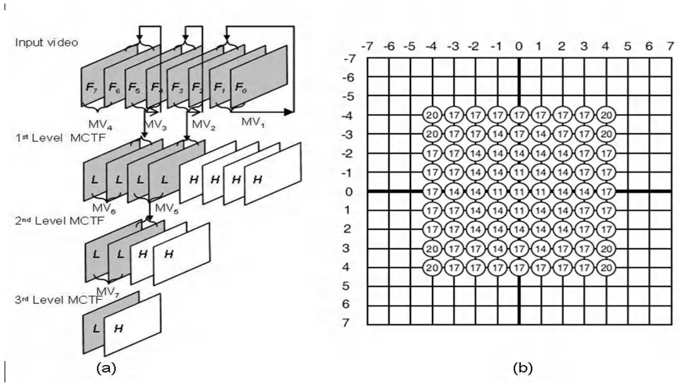


Fig. 1. (a)Predictive ME structure for MCTF. (b) Possible minimum number of search points for Hexagon motion estimation algorithms.

ME algorithm in this case it is HEXBX, $P_s(i, j)$ - is the probability that the optimal MV is located at the (i, j) position.

Since the probability of MVs distributed outside the ±4 region is lower than 8% [10], we analyze minimum number of search points N_s within a region of ±4 pixels about the stationary motion vector (0, 0). To quantify the gain $\overline{G}$ for block motion estimation, we define $P0$ - probability of stationary blocks i.e. motion vector (0, 0) and Pi - probability of quasi-stationary blocks within ±i, but excluding (i-1, i-1).

Since the matching errors are assumed to be monotonically decreasing towards the global minimum point, that can be modeled using Normal distribution. Hence we have computed MVs probabilities to be 0.55046, 0.2724, 0.039915, 0.018638, and 0.0136 at 0, ±1, ±2, ±3 and ±4 respectively. The statistical average gain $\overline{G}$ of search points per block in a MCTF for predictive Hexagon over the conventional Hexagon is,

$$\overline{G} = 7/1 \times P0 + 68/8 \times P1 + 160/16 \times P2 + 294/24 \times P3 + 440/32 \times P4$$
$$\overline{G} = 7 \times 0.595046 + 8.5 \times 0.2724 + 10 \times 0.039915 + 12.25 \times 0.018638 + 13.75 \times 0.0136$$
$$\overline{G} = 7.29515$$

3 Experimental Results

The proposed technique is applied in Motion Compensated Embedded Zero Block Coding (MC-EZBC) [8] a state-of-the-art wavelet based scalable video coding system. During simulation of the proposed technique we have considered fixed block size ME, 5/3 wavelet transform for temporal filtering, Debauchees 9/7 wavelet filter for spatial wavelet transform, window size of 15 × 15 and block size 16 × 16. Standard test sequences like Akiyo, Table Tennis, Foreman, etc., 96 frames of QCIF (176 × 144) resolution at 30 frames per second showing all varieties of motions are considered.

Table 1. Number of search points and quality of the decoded video

Sequence	Technique	Search Points	Rate (kbps)/PSNR(dB)							
			400	800	1200	1600	2000	2400	2800	3200
Akiyo	Hex	93555	46.98	50.35	50.60	-	-	-	-	-
	PredHex	39501	46.98	50.35	50.60	-	-	-	-	-
GrMother	Hex	98010	44.18	46.51	47.41	48.10	48.33	-	-	-
	PredHex	39798	44.15	46.46	47.31	47.94	48.15	-	-	-
Tab Tennis	Hex	115830	29.05	30.00	30.32	30.51	30.58	30.65	30.67	30.71
	PredHex	40986	29.11	30.07	30.39	30.58	30.66	30.72	30.75	30.79
Foreman	Hex	118058	29.71	30.78	31.18	31.39	31.48	31.55	31.57	31.60
	PredHex	41134	29.77	30.83	31.23	31.44	31.54	31.61	31.64	31.66

The number of search points decreased by 2.36 times the original Hexagon algorithm. In case of slow motion video the quality of the decoded video remains same and in case of fast video with camera pan the objective quality (PSNR) decoded video varies as shown in Table-1, but visually there is no effect. Due to the restriction on number of pages the complete results is not given in the paper, but the technique is tested in all possible conditions.

4 Conclusion

This paper proposed a predictive technique to exploit the correlation among the frames in a GOP of MCTF. The technique can be applied to any fast block ME. In case of Hexagon ME algorithm, we observed a gain of approximately 7. The maximum variations in PSNR is 0.2dB that does not have any effect on the subjective quality of the video.

References

1. Schwarz, H., Marpe, D., Wiegand, T.: Overview of Scalable Video Coding Extension of the H.264/AVC Standard. IEEE Transaction on Circuits and Systems for Video Technology 17(9), 1103–1120 (2007)
2. Jain, J., Jain, A.: Displacement measurement and its application in inter frame image coding. IEEE Trans. Commun. 29, 1799–1808 (1981)
3. Girod, B.: The efficiency of motion-compensating prediction for hybrid coding of video sequences. IEEE J. Sel. Areas Commun. 5, 1140–1154 (1987)
4. Ohm, J.-R.: Three dimensional subband coding with motion compensation. IEEE Trans. Image Process. 3(5), 559–571 (1994)
5. Choi, S., Woods, J.: Motion compensated 3d subband coding of video. IEEE Trans. Image Proc. 8, 155–167 (1999)
6. Secker, A., Taubman, D.: Motion-compensated highly scalable video compression using an adaptive 3d wavelet transform based on lifting. In: IEEE Int. conf. Image Proc., pp. 1029–1032 (2001)

7. Secker, A., Taubman, D.: Highly scalable video compression using a lifting-based 3d wavelet transform with deformable mesh motion compensation. In: IEEE Int. conf. Image Proc., pp. 749–752 (2002)
8. Chen, P., Woods, J.W.: Bi-Directional MC-EZBC with lifting implementation. IEEE Transaction of Circuits, Systems and Video Technology 14(10), 1183–1194 (2004)
9. Zhu, C., lin, X., Chau, L.-P.: Hexagon-based search pattern for fast block motion estimation. IEEE Trans. Circuit and Systems for Video Technology 12, 349–355 (2002)
10. Cheung, C.-H., Po, L.-M.: A Novel Cross-Diamond Search Algorithm for Fast Block Motion Estimation. IEEE Trans. Circuits System Video Technology 12(12) (December 2002)

L.U.N.A. Ads – Sustaining Wireless Access for Mobile Users

Alessandro M. Martellone[1], David Tacconi[1], Vincenzo Del Fatto[2], and Giuliana Vitiello[2]

[1] Futur3 srl, via A. Abondi 37, 38100, Trento, Italy
{a.martellone,d.tacconi}@futur3.it
http://www.futur3.it
[2] DMI, Università di Salerno, Italy
{vdelfatto,gvitiello}@unisa.it
http://www.dmi.unisa.it

Abstract. The development of wireless communication technologies offers the possibility to provide new services to the users other than the web surfing. Futur3 is a wireless telecommunication company interested in designing, implementing and managing WMANs (Wireless Metropolitan Area Networks) that has launched a project, called L.U.N.A. (Large Unwired Network Applications), with the objective of covering main locations in the province of Trento (Trentino Alto-Adige, Italy) with a wireless mesh network. The Futur3 business model is based on a very low cost access to the L.U.N.A. network offering users services with tailor made advertising. In this paper we present the L.U.N.A. Ads, a client-side application, able to provide contents together with advertising, considering the usability and accessibility requirements.

Keywords: wireless network, mobile advertising, usability, accessibility, user centric design.

1 Introduction

In the last decade, two phenomena have dramatically changed the society where we live in and the role played by information technology, namely the global diffusion of the Internet with the advent of Web 2.0 and the introduction of innovative mobile handheld devices. These technologies have had a deep impact on everyday life and on the whole modern society, people being now able to connect with anybody everywhere at anytime. We are witnessing an increased need of new access technologies among mobile users, which can create the opportunity of a fusion between these two worlds, i.e. the Web and mobile phones, bringing to reality the concept of *mobile Internet*. The rapid diffusion of wireless technologies, both for residential users (WLANs) and for covering large geographic area (WMANs), and the reduced cost of network implementation and deploy has opened new scenarios for mobile services enhancing user's experience in her everyday activities. The growing success of mobile services, particularly for mobile

M. Sebillo, G. Vitiello, and G. Schaefer (Eds.): VISUAL 2008, LNCS 5188, pp. 155–166, 2008.

advertising, has been due to time sensitiveness, personalization [9] and location awareness [5][26]. In particular, several market studies highlight an increasing interest in mobile commerce applications [10][25]. Furthermore, we are now witnessing a tremendous growth of mobile handheld devices, able to connect to the Web relying on Wi-Fi connectivity and letting the users experience innovative, attractive and usable interfaces. These new ways of communicating stimulated researchers to propose several approaches [20][22] and guidelines [28] for developing innovative interfaces for mobile devices [8][14][19], capable of providing users with minimal and exhaustive visual information, meant to overcome both hardware and software limitations such as small visualization areas [16][27] and interaction modalities permitted by mobile devices [24][29]. In particular, several studies propose methods to improve interfaces in mobile commerce applications [13][15]. Finally, the advent of World Wide Web changed user perception of Quality of services (QoS), which is a rising evaluation issue for mobile service providers, designers and developers [21][23]. The European Telecommunications Standards Institute (ETSI) defines QoS as "...the collective effect of service performances which determine the degree of satisfaction of a user of the service" [4]. Discussing QoS typically deals with system response time [17][18] availability, security, throughput [17] and network performance [12], but also QoS can be considered in terms of the quality of user's experiences [6]. The increasing importance of QoS has encouraged the researchers to propose approaches and methods for quantifying and measuring it [7][11].

Following this analysis, it is a matter of fact that mobile service designers and developers should now take into account this convergence of technologies and services and change the design and development dynamics according to newly introduced challenges. In particular, designers have to face:

– The need for an increased interoperability between different systems and devices
– The fact that the design of services is now easier, faster and cheaper also for non expert users
– The fact that mobility has become a strict requirement, both in the sense of spatial mobility and among different devices.
– New and growing user expectations in terms of Quality of services.

Futur3 acts in this innovative context and proposes itself as a wireless telecommunication company interested in designing, implementing and managing WMANs within a research project funded by the local province and named L.U.N.A.. The goal of the L.U.N.A. project is to cover main locations in the province of Trento with a wireless mesh network and to realize a business model such that the services provided within the L.U.N.A. network are accessible and usable by everybody. The main goal of the Futur3 model is to provide the access to the network at a very low cost and then make it a right for everybody. To reach this objective, Futur3 planning to integrate in its services light advertising interesting for the user, based on her profile, together with the contents she is currently looking at. In particular, the mechanism of users profiling must be drawn from the information currently given by the user in the total respect of her privacy.

As a first step toward this integration, we propose L.U.N.A. Ads, a light, usable and accessible service for Futur3 customers, able to provide contents and advertising in such a way that the user perceives everything as a useful information. We will show in particular the steps taken in designing L.U.N.A. Ads by considering the above described constraints and those imposed by a usability study of the service. The rest of the paper is organized as follows. In Section 2 a comparative analysis of similar applications is presented. In Section 3 we show the details of L.U.N.A. Project and we describe the mechanisms behind L.U.N.A. services and L.U.N.A. Ads. In Section 4 we describe the functional and non functional requirements for the L.U.N.A. Ads and we discuss the accessibility and usability of the application. In section 5 we discuss some envisioned applications scenarios, while section 6 describes a first prototype of L.U.N.A. Ads. Finally, section 7 presents conclusions and some promising directions for future work.

2 Comparative Analysis

In this section we present a comparative analysis of applications similar to L.U.N.A. Ads. Such an evaluation has been accomplished by searching for and testing several applications related to our project objectives, in order to evaluate design ideas and implementative choices and then select best solution for our context. For such a comparative analysis we will refer to the following applications, which have strongly contributed to the L.U.N.A. Ads design:

1. **Google Desktop.** Google Desktop is a popular freeware desktop search application offered by Google, which allows users to search information in local computer, in a similar way to the traditional Google web search. Google Desktop interface is very similar to the Google web search interface, but it also provides additional controls such as Gadgets (such as forecast, email, and so on) and the Google Sidebar, a vertical bar, visualized on the right side of the screen, that helps to organize Google Gadgets. Users can add new functionalities, downloading new Gadgets from Google Desktop web site, organize them on the Sidebar simply with one click. Google offers developers a Software Development Kit for several operating systems, enabling the development of new Gadgets. Google Desktop suggested us a modular architecture and the sidebar position on the side of the screen.
2. **Jing.** Jing is a software, developed by TechSmith [3], that instantly captures and shares images and video from user's desktop. When it is installed, it appears like a little sun that user can move along the sides of her desktop; if the mouse passes over Jing, the interface shows three "rays" that represent the functionalities of application. This solutions suggested us to use a moon metaphor (English translation of Italian word "Luna"), in order to realize an interface for the L.U.N.A. Ads pleasant and recognizable.

3. **Finetune.** Finetune is an online music recommendation service based on Air, an Adobe Flash platform [1]. The Adobe Air runtime lets developers use proven web technologies to build rich Internet applications that deploy to the desktop and run across operating systems; Adobe provide a SDK for Windows and recently a Beta version for Linux platforms. When user runs Finetune a pleasant graphic interface is shown on the desktop, from which users can choose different features: search songs by the favourite artist, create a personal playlist, choose a preferred music category. Finetune has been taken in consideration in our analysis for the tecnologies used for its developement that are able to support several devices and different operative systems.

As it can be easily seen, Google Desktop is interesting for the simplicity of installation of new features and also for supporting many operating systems. Jing makes pleasant and easy to understand its user interface by using a sun metaphor. Finally, Finetune, is developed with interesting web technologies but currently it is a still immature technology. Both Jing and Finetune require other software installed on user's device, such as Microsoft .NET framework, for Jing and an Adobe Virtual Machine for Finetune, thus representing a limit for application deployment both on laptops and mobile devices. As we will show in the following, L.U.N.A. Ads design has started by taking into account these advanced and estabilished desktop applications, trying to take inspiration in designing user interfaces and choosing development technologies. A summary of the three software characteristics is shown in Tab. 1.

Table 1. A comparison among three software in terms of Portability, Installability and Attractiveness

	Portability	Installability	Attractiveness
Google Desktop	X	X	
Jing			X
Finetune			X

3 L.U.N.A. Project

The L.U.N.A. project is a research project institued by the province of Trento, by which Futur3 aims at realizing an wireless network infrastructure able to provide access to everybody, everywhere and at any time. In order to make this possible in a short time and at an affordable cost, Futur3 makes use of emerging technologies in networking, like Wi-fi, Wireless Mesh Networks, Hiperlan and lately WiMax. In particular, Futur3 will provide Wi-Fi coverage in the cities of Trento and Rovereto, cultural, administrative and research centres, and Riva del Garda, a highly tourist area. These areas will be connected through Hiperlan links. Futur3 will provide services for its mobile customers, so that they can enjoy the Wi-Fi access, with solutions connected with the territory. This innovative network will also offer to local actors the opportunity to experiment new technologies, to increase the interest for this area, and to test innovative solutions for the citizens.

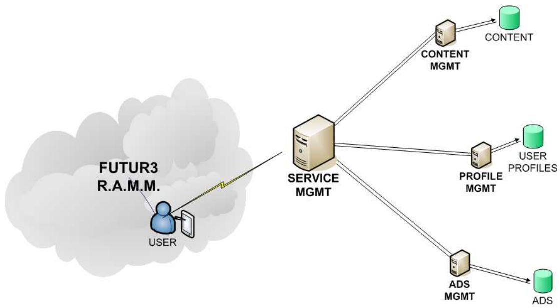


Fig. 1. Service Management Architecture

3.1 L.U.N.A. Services

The services offered through the L.U.N.A. network will be provided to users by the *Service Management* that interact with the *Profile Management*, the *Content Management* and the *Ads Management*, as shown in Fig. 1, which will leverage on the three main pillars of the service architecture:

– **User Profile**, i.e. user interests and user context, including her position within the network (localization).
– **Contents**, i.e. all the information requested by users and the information coming back from the network.
– **Advertisements**, i.e. all the commercial information that can be offered through the network.

The interaction among the management components interact is shown in Fig. 2. When a user connects to the Futur3 network, a request for a *service* is sent toward the *Service Management* (SM). Before performing any other action, the SM will get information about the user from the *Profile Management* (through a $getUP(\cdot)$ request) that will send back the *User Profile* (UP). As a consequence, the SM may search for contents specific for the UP and the Service under consideration (through a $getContents(\cdot)$ request). Then, the SM gets advertising information (ADS) according to the specific service, the UP and the contents to be delivered to the user (through a $getAds(\cdot)$ request). Finally, all the acquired specific information can be used to set the requested service (through a $setService(\cdot)$ function). It has to be noticed here that this cycle can be performed several times also when the service is running, in particular when the user requests some updated information either explicitly (the user sends a request to the SM) or implicitly (the user changes his position or more in general his context).

3.2 L.U.N.A. Ads

L.U.N.A. Ads is a client-side application independent from other running software, allowing the authentication to the L.U.N.A. network and managing all

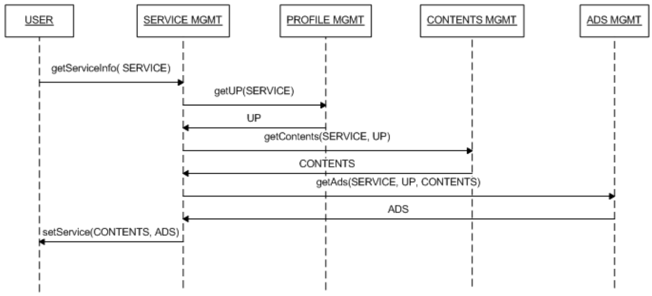


Fig. 2. Service sequence diagram

the personalized services related to the user's profile. The application will be visible on the screen in a strategic position and advertising will be displayed consequently. The application is responsible for keeping the connection alive, so that anytime the user closes the application she loses the connection to the web through the Futur3 network. L.U.N.A. Ads will be compatible with various operating systems and system architectures will be developed in order to allow an easy customization of the application to the user by adding or removing through several plug-in functionalities and services. Personalized services we intend to provide to our users will be mainly based on the user profile and the context she is currently immersed in. The implementation of L.U.N.A. Ads plug-in services will allow the user to continuously interact with the network. By doing that, we will try to understand users experience when using the Futur3 network and her interests either explicitly through periodic questionnaries or implicitly by "capturing" her requests to the network when using L.U.N.A. services. This research of personalization plays a key role in the design of L.U.N.A. Ads and determines an important part of our study.

4 L.U.N.A. Ads Requirements

4.1 Functional Requirements

The main functional requirements we have identified in the analysis of L.U.N.A. Ads reflect the main goals of the L.U.N.A. project. At first, we want the application to be able to automatically identify the L.U.N.A. network and connect the user with pre-entered credentials. The user has then the need for searching information: this has to be supported by a visual map where users can obtain commercial and tourist information or interact with other connected users, based on their position. L.U.N.A. Ads must then provide the functionality for playing multimedia contents like videos. Users will then be able to send sms and email to other registered users, without entering the web browser or other specific software. Finally, L.U.N.A. Ads must be designed in such a way that the installation of additional services, integrated to the application itself, is easy and reliable.

4.2 Nonfunctional Requirements

Nonfunctional requirements come from a variety of sources during the elicitation phase. Below are summarized the main nonfunctional requirements:

- Reliability: The system must have the ability to perform a required function under heavy conditions like a overfully loaded network, bad signal or lost connection.
- Portability: The application must be running on many operating systems: Windows (XP,Vista), MacOS, major Linux distribuitions and on many devices such as laptops, mobile phones, Internet Table.
- Security: The user's informations must not be violated.
- Attractiveness: The interface's appearance must be attractive to facilitate the use.

Furthermore, accessibility and usability are two very important features in the L.U.N.A. project quality plan. In Italy, the main accessibility reference is the Stanca's Law (Law n. 4 of the 9 January 2004) which defines a constitutional right of access to information and services also for people with reduced sense or motion capacity, that are in need of "assisting" technologies or particular configurations. Futur3 has implemented, in L.U.N.A. Ads development process, the accessibility guidelines suggested by "CNIPA" (National Centre for Information Technology in the Pubblic Administration) [2]. For increasing usability in our system design process we referred to the ISO standard 13407 that suggests a framework for increased usability, named Human-centred design process for interactive systems. Moreover, we are planning to involve users and specialists both for testing and verifying the prototype.

The first paper sketch developed, it was hung up in the hall of Futur3's office, according to *hallway and storefront* IBM methodology, and the all Futur3's staff was encouraged to examine the storyboard and issue comments; these have driven the first development of a prototype.

5 Applications Scenarios

In order to design a usable application, we have first detected the user's profiles and the main use scenarios. A profile has been identified based on age, profession and technological knowledge.

5.1 Target Users

Our demographic studies have detected some profiles:

1. Students : they have a good purchasing power, many of them use Internet and mobile phones. Students have been further divided into University and High Schoool ones.
2. Professional category: e.g. businessmen, administrative employees. They need to connect at the their headquarters, synchronize meetings; often these people do not have advanced technology skills.

3. Occasional users: e.g. tourists or museum visitors. They have access to a narrow area, without possibility to connect from outside the local wireless network. Thanks to a proxy configured accordingly, it is possible to selectively allow the surfing on selected websites of public or commercial use.
4. Senior citizens: people that are not familiar with the technologies, but that can be target of some provincial projects (e.g. e-health).

5.2 Scenarios

The scenarios are here depicted with reference to the users' profiles. The services that Futur3 intend to provide depend strongly from mobility and can be deployed and used only under its proprietary network. Some examples could be touristic tours, personal shopping assistant, localization based services. Let us explore in details some examples in the following.

Scenario 1: Social network experience. Luca is a university student recently moved to Trento. While he's walking in the city centre receives on his smartphone a message from his L.U.N.A. Ads runnning application. The application shows to Luca information on a so called "Luna Pub"; the message contains the main contacts of the "Luna Pub", including photos and comments on past parties which have been organized. Luca, after looking the photos and reading the comments, decides to reserve a table for the next night.

Scenario 2: Wine and food route. Anne is an English tourist and she has a Futur3 mobile device, rented at the tourist info point. After visiting the city centre, at the lunch time she wants to search a nearby restaurant; then Anne selects "Wine and food route" function. The L.U.N.A. Ads, shows a search criteria form. Anne fills the form searching for the item "Local cuisine" in the range of 1 km and reachable by foot. L.U.N.A. Ads shows the route on interactive map with information on all restaurants; moreover it plays a video which advertises on restaurant: "Antica trattoria".

Scenario 3: Parking Management. Roberto, a Trento's business man, wants to go to "Piazza Fiera" for an important meeting. He does not like to spend time in looking for a free parking place in the rush hour. Through the L.U.N.A. Ads parking service installed on his smartphone, he requires a parking close to "Piazza Fiera". The application shows an interactive map with a remarkable parking near the target point: the availability (full, free), the number of reserved places and the maximum parking time. Roberto selects one parking and reserves one place for 2 hours. The L.U.N.A. Ads system manages the payment transaction and shows to Roberto the route to reach the parking. In this scenario, we make use of an already existing service for the city of Trento: in multilayer parking places each parking lot is provided with a presence sensor and drivers are able to understand in advance - through public displays- the number of available parking lots.

Scenario 4: Access to healthcare. Olimpia is a senior grandmother with heart problems, and she's testing a new wireless service provided by the Hospital of Trento. Olimpia is under continuos heart monitoring through a mobile device monitor which transmit her ECG data automatically to the hospital data center. It happens that the system highlights a little cardiac problem and alerts automatically the emergency team of the hospital. Here a doctor takes a call, read the data sends and immediatly access to medical records about Olimpia. The doctor does a drug prescription. The drug is directly delivery at the Olimpia's home through a nearest ambulance to her. The actors of the e-health service interact with each other through the L.U.N.A. network and a e-health plug-in installed on the L.U.N.A. Ads application.

6 First Prototype

We present here a first prototype developed in Futur3 laboratory. The basic development idea was the simplicity in: installation, information fruition, personalization. We thought to use the L.U.N.A. Project logo, a stylezed moon, at the base of the deskbar development. The main objective is to give the feeling that the deskbar is like a medium through which the user uses L.U.N.A. services. By now refer to the "Wine and food route" scenario. The interface of the L.U.N.A. Ads application is showed to the user in the upper right corner of the desktop (default mode), in such a way to not disturb the user (see figure 3). When the user moves the cursor on the moon, a transparent window is shown. The window contains the last visited section by the user, when he is connected at the L.U.N.A. network. Furthermore it is shown a login form. To close or open the application's window the user has to move the cursor out of or in the area of the application. The connection at the L.U.N.A. network is displayed through an icon showing in an intuitive way the established connection to the L.U.N.A. network. Figure 4(a) shows the visual interface of a service section. It consists of a tabbed list from where users can recall the functionalities, a working area, based on the running service, and a visualization area, where advertising information are shown.

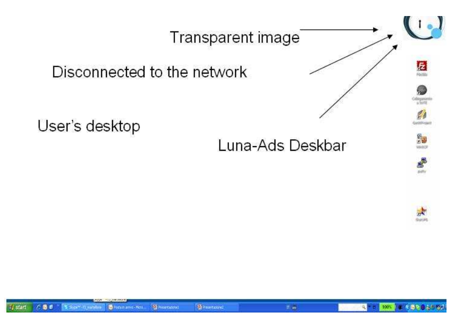
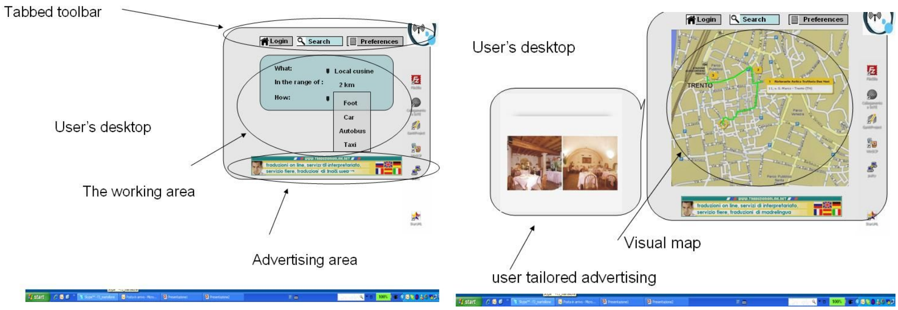


Fig. 3. A screenshot of L.U.N.A. Ads running



(a) The interface of the search section (b) The output of a query search

Fig. 4. Screenshots of L.U.N.A. Ads

A main service of L.U.N.A. Ads is a location-based search. The results of a user's query are visualized on an iteractive map, which allows users to manipulate its graphic components and acquire further informations (e.g. the contacts of a shop) or to improve a query (e.g. the park near a shop). As an example, let us consider the scenario where a user searches a restaurant in a range area and he wants to go there by foot. The service must be previously installed by user, as expected by modularity requirements. L.U.N.A. Ads shows the route on an interactive map (see figure 4)(b)), and for every restaurants found it gives back information and plays a video advertising. In this scenarios we have taken into account many requirements: interactivity, with a search supported by the map, and the visibility with the display of user tailored advertising.

7 Conclusion and Future Work

In this paper we have shown the L.U.N.A. Ads usability study and the design of a first prototype of this client-side application, which allows the authentication to the L.U.N.A. network and the management of the user's profile, necessary to the Advertising Management System on which the L.U.N.A. project is based. In the near future we aim at providing the service to a selected group of users, according to the profiles identified through this paper and let them test and evaluate L.U.N.A. Ads in the respective application scenario,also taking into account QoS issues. By doing this, we will be able to understand the usability of the designed application as well as user perception in terms of system response time, availability, security, throughput and network performances, thus improving accordingly its user interface as well as its functionalities.

Acknowledgments

The work of Alessandro M. Martellone and David Tacconi has been funded by the Luna Project and by the Province of Trento.

References

1. Adobe air, http://www.adobe.com/products/air
2. Centro nazionale per l'informatica nella pubblica amministrazione, http://www.cnipa.gov.it/
3. Jing project, http://www.jingproject.com
4. General aspects of quality of service and network performance in digital networks, including isdn, 1990. ETSI, Network Aspects (NA).
5. Evaluation report on technology and business aspects, usability, user acceptance and user recognition, 2003. ELBA (European Location Based Advertising) Project.
6. Ceaparu, I., Lazar, J., Bessiere, K., Robinson, J., Shneiderman, B.: Determining causes and severity of end-user frustration (2002)
7. Choi, C., Kim, C., Sung, N., Park, Y.: Evaluating the quality of service in mobile business based on fuzzy set theory. In: Fuzzy Systems and Knowledge Discovery FSKD 2007. Fourth International Conference on 2007, vol. 4, pp. 483–487 (2007)
8. Costanza, E., Inverso, S.A., Allen, R.: Toward subtle intimate interfaces for mobile devices using an emg controller. In: CHI 2005: Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 481–489. ACM, New York (2005)
9. DeZoysa, S.: Mobile advertising needs to get personal. Telecommunications International 36(2), 8 (2002)
10. Drossos, D., Giaglis, G.M.: Mobile advertising effectiveness: an exploratory study. In: ICMB 2006: Proceedings of the International Conference on Mobile Business, p. 2. IEEE Computer Society Press, Washington (2006)
11. Ivanovich, M., Fitzpatrick, P., Li, J., Beresford, M., Saliba, A.: Measuring quality of service in an experimental wireless data network
12. Jacko, J.A., Sears, A., Borella, M.S.: The effect of network delay and media on user perceptions of web resources. Behaviour and Information Technology 19(6), 427–439 (2000)
13. Koukia, S., Rigou, M., Sirmakessis, S.: The role of context in m-commerce and the personalization dimension. In: WI-IATW 2006: Proceedings of the 2006 IEEE/WIC/ACM international conference on Web Intelligence and Intelligent Agent Technology, pp. 267–276. IEEE Computer Society Press, Washington (2006)
14. Lee, K.J., Ahn, S.C., Kim, H.-G.: Using a mobile device as an interface tool for hmd-based ar applications. In: ACE 2006: Proceedings of the 2006 ACM SIGCHI international conference on Advances in computer entertainment technology, p. 7. ACM, New York (2006)
15. Lee, Y.E., Benbasat, I.: Interface design for mobile commerce. Commun. ACM 46(12), 48–52 (2003)
16. MacKay, B., Watters, C.: The impact of migration of data to small screens on navigation. IT and Society 1(3), 90–101 (2003)
17. Menascé, D.A.: Qos issues in web services. IEEE Internet Computing 6(6), 72–75 (2002)
18. Palola, M., Jurvansuu, M., Korva, J.: Breaking down the mobile service response time. Networks 1, 31–34 (2004)
19. Paolino, L., Sebillo, M., Tortora, G., Vitiello, G.: Framy: Visualizing geographic data on mobile interfaces. Journal of Location Based Services (2008)
20. Pulli, K., Aarnio, T., Roimela, K., Vaarala, J.: Designing graphics programming interfaces for mobile devices. IEEE Comput. Graph. Appl. 25(6), 66–75 (2005)

21. Saliba, J., Beresford, A., Ivanovich, M., Fitzpatrick, P.: User-perceived quality of service in wireless data networks. Personal Ubiquitous Comput. 9(6), 413–422 (2005)
22. Sánchez, J.A., Starostenko, O., Castillo, E.A., González, M.: Generation of usable interfaces for mobile devices. In: CLIHC 2005: Proceedings of the 2005 Latin American conference on Human-computer interaction, p. 348. ACM, New York (2005)
23. Shneiderman, B.: Designing the User Interface: Strategies for Effective Human-Computer-Interaction, 3rd edn. Addison-Wesley, Reading (1998)
24. Sohn, T., Li, K.A., Griswold, W.G., Hollan, J.D.: A diary study of mobile information needs. In: CHI 2008: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems, pp. 433–442. ACM, New York (2008)
25. Tripathi, A.K., Nair, S.K.: Mobile advertising in capacitated wireless networks. IEEE Trans. on Knowl. and Data Eng. 18(9), 1284–1296 (2006)
26. Varshney, U.: Location management for mobile commerce applications in wireless internet environment. ACM Trans. Interet Technol. 3(3), 236–255 (2003)
27. Wang, L., Sajeev, A.S.M.: Roller interface for mobile device applications. In: AUIC '07: Proceedings of the eight Australasian conference on User interface, pp. 7–13. Australian Computer Society, Inc, Darlinghurst, Australia (2007)
28. Weiss, S.: Handled Usability. John Wiley, Chichester (2002)
29. Zhou, L., Shaikh, M., Zhang, D.: Natural language interface to mobile devices, pp. 283–286 (2005)

MoViSys – A Visualization System for Geo-Referenced Information on Mobile Devices

Maria Beatriz Carmo, Ana Paula Afonso, Paulo Pombinho de Matos, and Ana Vaz

Departamento de Informática, Faculdade de Ciências, Universidade de Lisboa
Edifício C6, Piso 3, Campo Grande, 1749-016 Lisboa, Portugal
{bc,apa}@di.fc.ul.pt, {ppombinho,afvaz}@xldb.di.fc.ul.pt

Abstract. Visualization of geo-referenced information on mobile devices can be useful in diverse areas. These applications require a user-friendly interface to enable querying the data interactively and the generation of intelligible images. Moreover, they must deal with the limitations of mobile devices, such as small screen size, and the context of mobility. This paper describes MoViSys, a visualization system for geo-referenced data, organized in several categories with multiple attributes, on mobile devices. This system explores filtering mechanisms to present the result of interactive queries specified through an adaptive interface.

Keywords: Geo-referenced information visualization, mobile visual information systems, filtering mechanisms, adaptive user interfaces.

1 Introduction

The portability of mobile devices and the evolution of its capabilities, like, its increasing processing power and functionality capacity, and the proliferation of wireless communication infrastructures have contributed to the increased research effort on design and development of methods and systems for visualization of information conceived for this specific type of devices.

However, compared to desktop computers, mobile devices have some restrictions that continue to pose difficulties for user interaction. Moreover, some of these restrictions are unlikely to disappear in the near future, because the device itself has the requirement of being small. Consequently, these limitations and the mobility context impose severe usability and visualization problems.

To overcome these limitations and problems, visualization applications for mobile devices have to adopt efficient mechanisms to select the relevant information and include appropriate interactive interfaces that integrate context information and user preferences.

Research on visualization of geo-referenced information on mobile devices has been mostly focused on mechanisms that better represent and interact with this information on mobile maps [3]. There are already several commercial geo-referenced visualization systems for mobile devices, such as TomTom Navigator System [15] and Google Maps Mobile [8]. However, these systems present some limitations. Systems like TomTom are excessively focused on the navigation task and, consequently, do not

M. Sebillo, G. Vitiello, and G. Schaefer (Eds.): VISUAL 2008, LNCS 5188, pp. 167–178, 2008.

usually offer the best graphic quality, presenting maps with an excess of information and difficult to read. The last problem also occurs with Google Maps Mobile as it may produce cluttered images when a lot of results are displayed. Many of these systems have limited interfaces for query specification, limiting the user to a set of categories, not allowing the definition of attributes in each category (e.g., category hotel can have various attributes like stars of the hotel, parking conditions and room information) and do not present a personalized query interface based on user preferences and longer or shorter time usage histories.

We are developing MoViSys (Mobile Visualization System), a visualization system for geo-referenced data organized in several categories with multiple attributes through mobile devices. This system includes filtering mechanisms to control the density of the information to be displayed [5] and an adaptive interface for user query specification based on selection of categories and attributes.

This paper is organized as follows. Section 2 presents the technical limitations of small devices and the characteristics of mobile environment that must be taken into account in the design of mobile applications. Section 3 presents our visualization system describing its fundamental features. Section 4 describes the interface of queries specification. Section 5 points out some conclusions and future work.

2 Mobile Design Issues

A comparative analysis of the characteristics of mobile devices and desktop computers shows that the first ones present several restrictions that must be considered in the design of applications for mobile devices, namely for mobile visualization applications.

Screen size

The screen display on mobile devices exhibits a drastically reduced area when compared to the screens of desktop or laptop computers. The size of screens strongly constrains the amount of information displayed: the number of pages necessary to visualize the same information is much higher when compared to personal computers.

Zooming and panning are among the techniques that have emerged to enlarge virtually the screen size. However, these techniques are considered complex, disorientating and tedious from a cognitive point of view [3]. Other techniques can be explored to solve this problem, namely: dialogue boxes, windows, tabs, pull-down menus and pop-up menus. However, some of these techniques are more suited to small-screens devices than others [17].

Other proposals have been formulated, namely, from the mobile computing area. The exploration of information access models that incorporate knowledge of user's context, for example, geographic location, permits to adapt the displayed information to user's preferences. This is one of the main methods that significantly reduce the amount of textual or graphical information displayed.

Resolution and colors

Another aspect that should be considered in the design of mobile applications is the resolution and colors that are available. These devices generally present low resolutions and few colors, which limits the amount and quality of information simultaneously visible on a screen.

Processing and storage capacities
The storage and processing power capacities, incomparably smaller when compared with the capacities of desktop computers, limit the quantity and quality of the information accessed, particularly, by avoiding the use of computationally demanding algorithms. For example, the quality of zoom techniques on small screens is still far from ideal, namely because of the processor speed of these devices.

Methods of interaction
The methods of interaction on mobile devices are still quite far from being considered as natural interfaces. The nature of user interaction plays a more important role in the design of applications for mobile devices than in the design of applications for desktop computers, because the input peripherals among different devices can vary greatly. Different alternatives should be considered in the design of interfaces to interactive applications for these devices. For example, the use of pens and touch screens, advise that the elements for navigation and interaction, are placed at the bottom of the screen to avoid that this becomes obstructed during the interaction.

Environment
The environment of mobility constitutes another factor that affects the design of interactive mobile applications. The physical conditions of use of these devices, for example, luminosity and noise, can affect the perception of colors, maps and the reception of sound alerts.

The characteristics of the mobile communication networks and the limitations of bandwidth and reliability make impracticable its use in similar conditions to the communication systems and computation that operates on fixed networks. These characteristics affect the interaction and the usability of the applications and should also be considered in the design of interactive applications.

Autonomy
Another important issue in mobile devices is their power supply through batteries. Although they have been improved recently, they have not followed the evolution of other components in mobile devices. Despite the evolution of the autonomy of these devices that can match up to the values of a laptop computer, the limitations of energy continue to determine the behavior of users' work and is one of the factors that can affect the design options of mobile applications.

The main focus in the MoViSys project is to explore techniques able to mitigate the limitations of small screens in the design of mobile visualization applications, particularly, visualization mechanisms that guarantee a straightforward and easy access to relevant information, and an interactive interface that integrates context information and user preferences.

3 Visualization Issues

Our goal is to visualize geo-referenced data organized in several categories with multiple attributes. Data elements that satisfy a query made by the user are represented with icons superimposed on a map. Selecting appropriate symbols to convey the information and organize them in a clear way are important issues to achieve meaningful representations.

As the screen size of mobile devices is very limited, the representation of all the icons that satisfy a certain query may lead to a confusing image. Therefore, to obtain intelligible images there is a need to control the number of icons displayed on the map. This means that we have to reduce the number of icons displayed when the number of points of interest that satisfy the user' s query are too many or too close to each other. To achieve this purpose, we use a combination of filtering mechanisms, eliminating less relevant results, and aggregation of elements, grouping results which are close to one another.

3.1 Filtering Mechanisms

To filter the points of interest that should be represented we use a function to quantify the degree of interest of a given point. The expression "degree of interest" was introduced by Furnas [7] to name a function that quantifies the interest of a point given a *focus* of interest. The value of this function in a point x depends on the *a priori* importance of the point, API(x), and on the distance between x and the current *focus*, y: DoI(x|y)=API(x)-D(x,y).

Other authors have also defined functions to quantify the interest of a data element in a particular query. Keim and Kriegel calculate the relevance of each element of a database in a query [10]. The relevance factor of an element is determined by calculating distances between the values of its attributes and the values of the selection predicates. The distance functions used depend on the attributes' data types. As the values calculated by the distance functions may be in completely different orders of magnitude, these values are normalized in a fixed range. The normalized distance values are combined in a single distance value using weighting factors to control the importance of each selection predicate.

Reichenbacher defined a relevance function that is applied to the visualization of event queries in mobile environments [13]. This function depends on a spatial distance, a topical distance and a time distance. The spatial distance is the Euclidean distance. The time distance is the difference in minutes between the time of the event and the current time. The topical distance takes the following values: 0, if the event does not belong to the category searched in the query; 0.5, if the event belongs to the category searched in the query but the type of the event does not match; 1, if the category and type of the event match the query. The total relevance of each event is the sum of the normalized values of each function (spatial, temporal and topical).

Our goal is to calculate for each point of interest a value that measures its relevance in a given query. We have adopted an approach similar to [10] and [13] as we define a function that calculates the distance between the conditions defined by the user and the values that occur in the points of interest. Comparing with Furnas' function, we consider several *foci* of interest in each query and do not give an *a priori* importance to each point.

We have defined a base DoI function that quantifies the degree of interest of the user on a given point of interest p_j as the average of the user's interest (UI) in specific values of k different attributes a_i, i=1,2,…,k:

$$DoI(p_j) = \frac{\sum_{i=i}^{k} UI(a_i, p_{ji})}{k} \in [0,1]$$

Function UI(a_i,p_{ji}) depends on the distance between the value selected by the user for the attribute a_i and the actual value of that attribute in the point of interest p_{ji}. We adopted the following distance functions:

– For an attribute of nominal type

$$Dist(a_i, p_{ji}) = \begin{cases} 0, \text{ if } a_i = p_{ji} \\ 1, \text{ if } a_i \neq p_{ji} \end{cases}$$

– For an attribute of numerical type

$$Dist(a_i, p_{ji}) = \left| \frac{a_i - p_{ji}}{\max_i - \min_i} \right|$$

where $\max_i$ and $\min_i$ correspond, respectively, to the maximum and minimum known values of the attribute.

– For the geographic distance we use a normalized Euclidean distance

$$Dist(a_i, p_{ji}) = \sqrt{\left(\frac{x_a - x_{p_i}}{\max_x - \min_x} \right)^2 + \left(\frac{y_a - y_{p_i}}{\max_y - \min_y} \right)^2}$$

where (x_a, y_a) and (x_{pi}, y_{pi}), correspond, respectively, to the position of interest to the user and the actual location of the point of interest. If the position of interest defined by the user is in the visible area, the distance is normalized using the maximum and minimum coordinates of this area. If the position of interest is defined outside the visible area, the maximum and minimum coordinates will be the edges of a square centered in the visible area and extended to include the position of interest.

Having calculated the distance we can then determine the UI:

$$UI(a_i, p_{ji}) = 1 - Dist(a_i, p_{ji}) \times w_i,\ w_i \in [0,1]$$

where w_i is the weight of the attribute a_i, which can be defined by the user to specify how important that attribute is in the query.

3.2 Extended DoI

Although the use of weights in attributes allows the DoI function to properly distinguish between important attributes and not so important ones, it does not take into account differences of importance between the categories themselves. To solve this problem, an extension to the base DoI function was defined.

To make a distinction on the importance between the different categories, a category weight was added to the DoI function:

$$DoI(p_j) = \frac{\sum_{i=i}^{k} UI(a_i, p_{ji})}{k} \times w_{cat} \in [0,1]$$

where w_{cat} is the weight defined for the p_j's category.

Furthermore, the base DoI function only allows the user to select a single value for each attribute (e.g. restaurant type: Chinese). However, in some cases, it is useful to the user to be able to select multiple values (e.g. restaurant type: Chinese, Italian …), or a range of values (e.g. hotel stars: 3 to 5).

To solve this limitation, the nominal and numerical distance functions were also extended:

– For an attribute of nominal type with l alternate values

$$Dist(a_i, p_{ji}) = \begin{cases} 0, \ if \ a_{i1} = p_{ji} \vee a_{i2} = p_{ji} \vee \ldots \vee a_{il} = p_{ji} \\ 1, \ if \ a_{i1} \neq p_{ji} \wedge a_{i2} \neq p_{ji} \wedge \ldots \wedge a_{il} \neq p_{ji} \end{cases}$$

– For an attribute of numerical type, with l alternate values

$$Dist(a_i, p_{ji}) = \min\left\{ \left| \frac{a_{i1} - p_{ji}}{\max_i - \min_i} \right|, \left| \frac{a_{i2} - p_{ji}}{\max_i - \min_i} \right|, \ldots, \left| \frac{a_{il} - p_{ji}}{\max_i - \min_i} \right| \right\}$$

– For an attribute of numerical type, with a range of values between a_{i1} and a_{i2}

$$Dist(a_i, p_{ji}) = \begin{cases} 0, \ if \ a_{i1} \leq p_{ji} \leq a_{i2} \\ \left| \frac{a_{i1} - p_{ji}}{\max_i - \min_i} \right|, \ if \ p_{ji} < a_{i1} \\ \left| \frac{a_{i2} - p_{ji}}{\max_i - \min_i} \right|, \ if \ p_{ji} > a_{i2} \end{cases}$$

In the queries where the user only selects one value, $l = 1$ the extended functions correspond to the base distance functions.

The DoI function quantifies the relevance of each point of interest. If a predefined number of icons are allowed, only the points of interest with the higher degree of interest value are displayed. This means that less relevant points of interest are omitted.

3.3 Aggregation

Even using a DoI function, when the points of interest are not uniformly distributed over the screen, we can still have a cluttered image. Aggregation is used in other systems, for instance MetaCarta system [11], to reduce this problem. An aggregation symbol is used when several objects are associated with the same coordinates. For this reason, this technique is referred as a generalization operator in cartographic literature [6]. We have considered several aggregation symbols to convey more information.

To decide whether to aggregate objects, it is not enough to count the total number of objects to display, because they may not be evenly distributed on the visualization area. A wiser approach is to superimpose a regular grid on the visualization area (as proposed in [16]), and determine for each cell the number of icons presented. An aggregation will be made when the number of icons in a cell exceeds a predefined number (Figure 1). In this case, if the grid was attached to screen coordinates it would lead to visual discontinuity in pan operations. This would occur because the number of objects in each cell of the grid would vary as the displayed area is changing. For

(a) (b)

Fig. 1. Example of an aggregation: (a) grid superimposed on the map; (b) aggregations based on the number of elements on each cell of the grid

this reason, points of interest that were once inside an aggregation may cease to be, and single ones may be aggregated. To avoid this problem, we associate the regular grid with the visible area in user's coordinates. When a pan operation is executed, the grid moves together with the visualized area, avoiding visual discontinuities.

3.4 Symbology

Selecting the adequate symbol for each category leads to the generation of meaningful representations. For instance, symbols can be adapted according to user's current activity, language, age group and time of year/day [12]. Moreover symbols can express object's relevance: using different levels of opacity [13], including different complexity levels [4], or attaching a vertical bar whose height represents how much the object satisfies the user's query [3].

Considering the dimension of the symbols, they must be large enough to allow interactive selection to get details on demand, but as small as possible so that they do not hide the underlying map.

From a semantic point of view, we have considered different levels of detail expressing different meanings: to identify a category, to express the value of an important attribute of a category and to represent aggregations either with all objects of the same category or with objects of different categories. The aggregation symbol may also show if it includes at least one object with the higher level of interest. This is done by adding a '+' marker to the aggregation symbol. Some examples of symbols are shown in Figure 2.

To represent an aggregation with objects of different categories we designed a symbol consisting of small symbols. That is, the square area of the symbol is tiled with four reduced symbols. If the aggregation includes objects from more than four categories, only the small symbols corresponding to the four categories with more objects will be displayed.

In Figure 3 we show the hierarchy we have defined to organize all the symbols.

(a) (b) (c) (d) (e)

Fig. 2. Visualization of points of interest with: (a) individual symbol; (b) and (c) individual symbols with detail; (d) aggregation; (e) multiple category aggregation

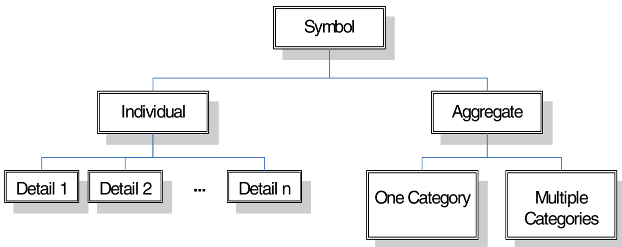


Fig. 3. Symbol Hierarchy

4 Query Specification Interface

The MoViSys interface uses the concept of dynamic queries [14] as an easy method for query specification and visualization of query results. Dynamic queries continuously update the data that is filtered from a database and visualized. By direct manipulation of sliders, check buttons, or other input widgets, called query devices [1], users can easily specify the attribute values for the desired categories and explore the map results.

As a result of several limitations presented by mobile devices, such as, small displays, inadequate input mechanisms for complex tasks, and considering also the amount and type of attributes that may be associated with geographic data [2], we propose a dynamic and consistent interface based on the set of data stored in a database. There are two distinct parts in our interface: an initial configuration interface and a query specification interface.

The configuration interface allows the definition of the center of the geographic area to visualize: a previous stored location, the current user location or a specific one defined by its geographic coordinates. It is also possible to specify some visualization parameters, such as, the size of the icons, the maximum number of icons visualized on a map at the same time, the minimum level of zoom with aggregations, among many others.

Figure 4 shows the organization of our query specification interface in two main areas: map area and the query specification area.

In the map area the user performs pan and zoom operations. On the upper right corner of the screen there is a button to increase or decrease the size of the icons according to the user state (in motion or stopped).

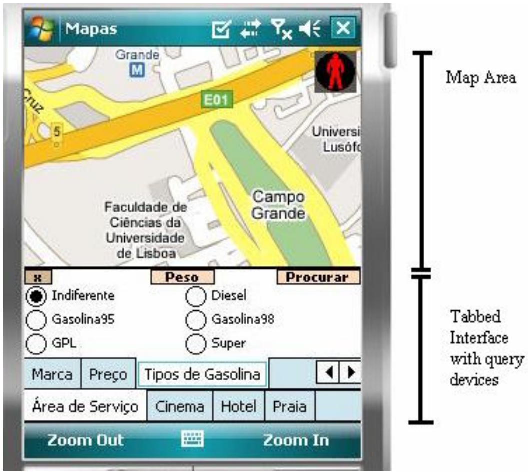


Fig. 4. Interface areas

(a) (b)

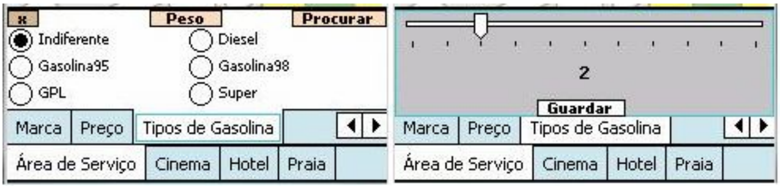
Fig. 5. Query device based on (a) numericUpDown control for range selection values (b) checkboxes for multiple choice queries

In the query specification area, the user selects the categories he intends to visualize, limited to seven options, as suggested in [17]. For the sake of comfort and operation quickness, the selection of categories and attributes is made using a *leafing* technique. These selected categories are shown using a double tabbed interface: a bottom tab line with the categories, another tab line with the attributes of the selected category and an area with query devices for the values of the selected attribute.

The user can interactively change the query specification based on the visualized results.

For the definition of attribute query values, we adapted the query devices proposed in [1]. Four query devices types have been defined. The first one is based on a numericUpDown control that allows the selection of a range of values (e.g., the schedule of a cinema). The maximum and minimum values are defined according to the values stored in the database, but the user can specify a custom range. An example is shown in Figure 5 (a) where this widget is used to select a schedule for a cinema.

The second query device is based on checkboxes usually used when there are a list of options for a pair (*category, attribute*) and the user may select any number of choices. Figure 5 (b) shows an example of this query device type. For the category



(a) (b)

Fig. 6. Query devices: (a) Selection of exclusive values (b) Selection for ordinal values

"Gas station service area" ("Área de serviço") and the attribute "Type of services" ("Extras") the user may select any number of options, such as, "having a restaurant" or "having a repair-shop".

The third query device is based on radio buttons used to select exactly one choice in a list of two or more options that are mutually exclusive. An example is shown in Figure 6 (a), where for the attribute "Type of gas station" several options are displayed.

The forth query device is based on a trackBar (slider) to select ordinal values (illustrated in Figure 6 (b)).

The query device is automatically selected according to the semantic of each attribute. The interface design is adapted to the user selection history. The first categories that are showed in the tabbed interface are the ones selected by the user during the last session. The remaining categories are arranged by alphabetical order. The categories attributes are sorted taking into account the number of selections of each attribute.

The lack of space implies that one of the most challenging aspects associated with the design process is the dynamic organization of space. We use a double horizontal tabbed interface to organize the categories and the respective attributes. This space organization maintains always visible the categories and the attributes enabling an easy selection and navigation for the user. Another technique that we have used to work in a limited space is *leafing*. When the screen navigation is performed using a stylus, scrolling is more complicated than *leafing*. Additionally, the directly manipulation of query devices avoids the use of a virtual keyboard. This is particularly important in small devices because it neither occupies any space nor overlaps important information.

Our visualization system is being developed for the Pocket PC, with the Windows Mobile 5.0 operating system and uses the .Net Compact Framework 2.0. The current position of the user is automatically obtained through an embedded GPS device. The client application communicates with the Google Maps server to retrieve the map images for the desired locations and zoom level [9]. Additionally, the client application obtains information from an SQL Server about the points of interest inside the visible area. These data is processed to calculate the degree of interest function for each point of interest selected. These points of interest are then sorted and filtered. Finally, using only the most relevant points, the grid is computed and all overlapping icons are aggregated. The image is then processed to include all the symbols and is drawn on the screen.

5 Conclusions and Future Work

This paper presents a visualization system for mobile devices, MoViSys, that combines visualization techniques to display query results of geo-referenced information organized by categories with an intuitive and adaptive interface for query specification. To obtain intelligible representations the number of objects displayed is controlled using filtering mechanisms, like a degree of interest function and aggregations. The interface based on dynamic visual queries allows the specification of queries and user's preferences.

We have made some evaluations on particular components of our system, namely, on symbols for the representation of categories and on the design of our interface, following a user-centred approach. We are doing more complete usability tests to evaluate the effectiveness of filtering mechanisms and user satisfaction.

Until now, our concern in map layout was restricted to icons overlapping, when a large number of objects are retrieved. We intend to explore also the overlapping of cartographical objects. To achieve this goal we are studying the inclusion of the proposed techniques in a Geographical Information Systems (GIS) for mobile devices.

Another important ongoing work is the interface refining to include multi-lingual support and the design of interface solutions to represent more complex queries that can deal with environmental changes in the user's proximity, temporal user requirements (e.g. find a theatre that is *t* minutes from user current location) and spatial user preferences (e.g. find hotels around *y* Km from the user current position). We also intend to improve the query specification interface to represent user longer or shorter time usage histories in the selection of categories.

Acknowledgments

This work was supported by FCT, through project PTDC/EIA/69765/2006 and through the Multiannual Funding Programme.

References

1. Ahlberg, C., Truvé, S.: Exploring terra incognita in the design space of query devices. In: Proceedings Work. Conference on Engineering for HCI, pp. 305–321 (1995)
2. Burigat, S., Chittaro, L.: Visualizing the Results of Interactive Queries for Geographic Data on Mobile Devices. In: Proceedings of ACM GIS 2005, pp. 277–284 (2005)
3. Burigat, S., Chittaro, L.: Geographical Data Visualization on Mobile Devices for User's Navigation and Decision Support Activities. In: Spatial Data on the Web – Modelling and Management, pp. 261–284. Springer, Heidelberg (2007)
4. Carmo, M.B., Freitas, S., Afonso, A.P., Cláudio, A.P.: Filtering Mechanisms for the Visualization of Geo-Referenced Information. In: GIR 2005, pp. 1–4 (2005)
5. Carmo, M.B., Afonso, A.P., Pombinho, P.: Visualization of Geographic Query Results for Small Screen Devices. In: Proceedings of ACM-GIR-2007, pp. 63–64 (2007)
6. Edwardes, A., Burghardt, D., et al.: Portrayal and Generalisation of Point Maps for Mobile Information Services. In: Map-based Mobile Services – Theories, Methods and Implementations, pp. 11–30. Springer, Heidelberg (2005)
7. Furnas, G.: Generalized Fisheye Views. In: ACM CHI 1986, pp 16–23 (1986)

8. Google Mobile Maps, http://www.google.com/gmm
9. Google Maps, http://maps.google.com
10. Keim, D.A., Kriegel, H.-P.: VisDB: Database Exploration Using Multidimensional Visualization. IEEE Computer Graphics and Applications 14(5) (1994)
11. MetaCarta Products, http://www.metacarta.com
12. Nivala, A.-M., Sarjakoski, L.T.: User Aspects of Adaptive Visualisation for Mobile Maps. Cartography and Geographic Information Science, Towards Ubiquitous Cartography 34(4), 275–284 (2007)
13. Reichenbacher, T.: Mobile Cartography – Adaptative Visualization of Geographic Information on Mobile Devices. PhD Thesis. Verlag Dr. Hut, München (2004)
14. Shneiderman, B.: Dynamic Queries for Visual Information Seeking. IEEE Software 11(6), 70–77 (1994)
15. TomTom Car navigation System, http://www.tomtom.com
16. Woodruff, A., Landay, J., Stonebraker, M.: Constant density visualizations of non-uniform distributions of data. In: ACM UIST 1998, pp. 19–28 (1998)
17. Zwick, C., Schmitz, B., Kuhl, K.: Designing for Small Screens. AVA Academia Publishing (2005)

A New Approach Toward a Modular Multimodal Interface for PDAs and Smartphones

Giovanni Frattini, Federico Ceccarini, Fabio Corvino, Ivano De Furio, Francesco Gaudino, Pierpaolo Petriccione, Roberto Russo, Vladimiro Scotto di Carlo, and Gianluca Supino

ENGINEERING.IT S.p.A. ,Via Antiniana 2/A,
80078, Pozzuoli (NA), Italy
{giovanni.frattini,federico.ceccarini,fabio.corvino,
ivano.defurio,francesco.gaudino,pierpaolo.petriccione,
roberto.russo,vladimiro.scottodicarlo,gianluca.supino}@eng.it

Abstract. In this paper we present an innovative approach for exploiting the handheld device hardware increasing power. These results have been achieved in the framework of a larger project aimed to the definition of a platform to build mobile multimodal applications. Such project, called CHAT, has been co-funded by Italian Ministry of Research. Focusing our attention on the system client side, we will show how it is possible to use a thin client approach for building high modular client interfaces, managing asynchronous interactions originated on the server-side, eventually pushing contents using different criteria (e.g. change of the environmental context). We have studied how to build mobile multimodal clients, adopting an innovative approach based on the aggregation of what we have called "multimodal objects" that are able to exploit the benefits of both telecommunication and web protocols to manage input/output processes. Furthermore, we will report some restrictions and constraints coming from the adoption of two different environments: .NET and J2ME. We will highlight our solid experiences in this area and some preliminary results of our research.

1 Introduction

The penetration of mobile device in western countries is still increasing. The Italian case is really surprising: every single Italian citizen has more than one mobile terminal. Thus, considering this large potential audience, and the increasing power of new generation terminals, a large space for innovation exists. In this context, usable multimodal services could have an unexpected impact on the market. Nevertheless, the research community should be able to propose a framework for building generic multimodal services, covering all their lifecycle. We are currently working on a platform for building end-to-end coordinated simultaneous multimodal applications. In this context, we believe that a special effort for enabling a mass diffusion of mobile multimodal services should be spent on the client side, where the situation is still evolving.

The CHAT project aims to partially cover this gap. We are designing and developing a platform for building multimodal services dedicated to cultural heritage

M. Sebillo, G. Vitiello, and G. Schaefer (Eds.): VISUAL 2008, LNCS 5188, pp. 179–191, 2008.

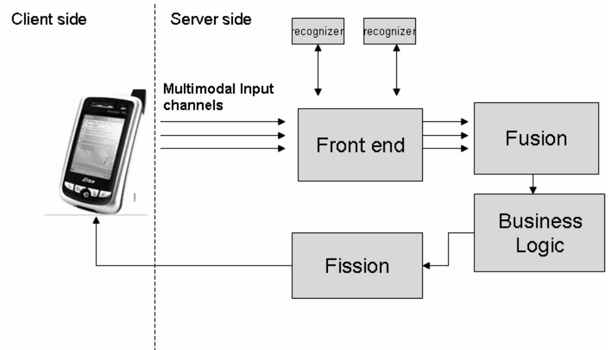


Fig. 1. System architecture

fruition (museums, archeological parks) and e-learning [1]. The ambition is to offer synergic multimodality in a mobile context using common devices [2]. Next figure shows a very high-level description of CHAT architecture.

On the server side are located the following modules:

- Front end: to collect the input coming from the clients and send them to specific recognizers (speech, sketch, handwriting recognizers);
- Fusion module: to merge inputs coming from different channels;
- Business logic module: to select the contents to send;
- Fission module: to send content to final users.

This architecture comes from well known patterns [3]; the main innovation we have introduced concerns the usage of a mix of web 2.0 and telecommunication technologies and it has produced interesting results. Nevertheless, this approach has different impacts on how software on terminals must be designed.

In this paper we will discuss our achievement in building mobile multimodal user interfaces using thin clients that are able to use available network resources for interacting with servers. To build this kind of user interfaces we have selected common commercial mobile devices such as PDAs or smartphones. We have developed a software framework for building multimodal clients that, using the network and a mix of different standard protocols, are able to effectively interact with the local environment and remote servers and for acquiring personalized contents.

As we are talking about a multimodal system, such interface should offer the following main features:

- Collect input from different channels (multimodal input) ;
- Manage outputs using different content formats (multimedia output).
- Exploit local resources to create an appropriate context to the service fruition.

Thus, a client satisfying these requirements must be able to manage the output channels and to collect the inputs coming and from users redirecting them toward the recognizers. Possibly, such a client must be thought for reducing as much as possible complex user interaction that are generally difficult and frustrating on small mobile terminals. We call such application MMUIManager (MultiModal User Interface

Manager). To offer such capabilities there are two possible approaches: thin client and fat client; depending on the volume of data processing that is performed locally on the terminal. The thin client approach is far more flexible, allowing the reuse of the same software in different contexts. Nevertheless, a simple browser is not enough. Considering that a standard markup language for synergic mobile multimodal interactions does not exist, a client side framework should be thought to be flexible and adhere to new upcoming standards in the future. The idea is to work to a framework for supporting a tag language for aggregating any type of content, considering the potential combination of modes as the composition of multimodal objects. In addition, such an approach could be useful for building mobile multimodal web 2.0 applications being a mix of traditional mark-up based rendering technologies and resident applications. MMUIManager is something similar to a browser that receives from the server the information describing the appearance and the behavior of the interface and load locally the software needed for interacting with the user.

This approach is someway similar to what happens with Mozilla and the language XUL [4]. For the same reason we have developed our own XML language for aggregating multimodal objects. We have called it LIDIM (from the Italian LInguaggio di Definizione Interfaccia Multimodale, language for designing multimodal interface). We have implemented a LIDIM-based MMUIManager for two different terminal technologies J2ME and .NET Compact Framework. In this paper we report some of the result we have achieved, analyzing strengths and weaknesses of both platforms.

Obviously the thin client approach implies that the MMUIManager has to be connected to a network for interacting with the server. Furthermore, for avoiding to buffer locally input signals (eventually continuous) an efficient real time transport protocol must be used. This is again a very good reason for adopting Telecommunication protocols like SIP/RTP and for developing special multimodal objects for using streaming protocols for managing both input/output.

2 Related Works

Let us start considering commercial products. Among the others, Kirusa [5] is a relatively small company that produces multimodal software for Telecommunication operators. Kirusa core product is a solution for Voice SMS, which is a relatively simple multimodal application.

Nuance [6], a well known brand in speech recognition, has recently released a framework for building multimodal mobile applications. This framework is based on standard like X+V and, thus, limited to alternate multimodality. It is worth noting that Nuance proposes a distributed recognition approach limiting the processing on the client side.

In [1] it is discussed some of the topic that, in the meantime, the CHAT project has faced. Many issues reported in that paper were related to terminals. It is interesting that, while for alternate multimodality several mark-up languages have been proposed and adopted industrially (namely X+V) [7], for synergic multimodality does not exist a standard language, even if several researches have addressed the problem proposing possible solutions [8]. After having studied how to extend existing languages (X+V, SALT, SMIL), we have decided to implement our (lightweight) language called LIDIM. This decision comes also on having considered other experiences. For example, the MONA project [8] produced a multimodal presentation server, which main features is to support the deploy of device independent applications combining graphical user interface and speech input/output. We have found particularly interesting the MONA capabilities to push pages on mobile devices using UIML, a markup language for describing user interfaces. MONA is based on a client side engine, which allows the presentation of a non standard mark-up language and allows the pushing of pages on the mobile device (while for low-end phone they use WAP push messages). Differently from the MONA project, in CHAT we have developed multimodal objects that are able to manage telecommunication protocol for pushing contents. More specifically, we use the IP signaling protocol (SIP) that can be used for sending unsolicited for triggering multimodal data retrieval.

On the same wave, we found other works. In [9] the authors face the problem of managing outputs, generally defining a special markup for coordinating the different modes (generally graphics and speech). We have considered these experience because it is hard to find a solution for adapting the input/output considering all the possible constraint: user attitudes and preferences, device limitations, data to be presented, etc. These are well known problems [10] that are amplified when considering the need to support different devices, different communication capabilities and general purpose interfaces. Anyway such browser-like approach needs a specific language to describe appearance, input and output modalities. In [2] we discussed an embryonic prototype of the J2ME mobile multimodal interface that we present in this paper.

Finally, we have considered experiences available on internet as open source packages. It has been of fundamental importance the Piccolo framework [11]. Piccolo is an open source library for graph visualization. The idea of using graphs and their representation as objects has inspired us to work around the idea of a more general framework for managing multimodal input/output based on this concept. Such a framework could have been feasible using the Piccolo approach. Furthermore, using the Piccolo example we have defined objects having "multimodal" features using J2ME. Continuing with our tribute to open source, the study of the thinlet and Zap frameworks [12, 13] has inspired us in finding a solution for aggregating multimodal objects in a single UI. Unfortunately Piccolo is not written for java-enabled phones while the thinlet J2ME implementation is still very light (but useful).

3 Multimodal Mobile Interface: The Thin Client Approach

As discussed previously, for the MMUIManager we have followed a thin client approach. So we have a relatively light application on the client side, while all the heavy weight processes (e.g. recognition processes, fusion etc.) resides on the server side.

The main MMUIManager tasks are:

- To build the graphic interface using an XML user interface language;
- To manage the input channels on the device;
- To manage the output channels on the device.

The UI has to be contemporary able to collect multimodal input and to show multimedia output. So client application has a presentation and an acquisition role and it is able to build an interface with the wished features according to the indication coming from the server.

Such choice allows a more general approach to the problem. The same client application can be used for all the possible application contexts (different cultural heritage fruition, e-learning); content or logic update will affect only the server side.

4 A Framework to Build a Multimodal Modular Interface

According to our goals, the MMUIManager must be based on an engine for interpreting an XML language for creating multimodal/multimedia interface on the fly. Considering this, we have designed and developed a framework for aggregating multimodal objects, which is a framework for managing objects having multimodal user interaction capabilities. A composition of multimodal objects creates a complete multimodal user interface.

The target platform choice (J2ME, .NET CF, Symbian, etc.) is not a simple task. Each platform has its strengths and weaknesses. Up to now, we have developed prototypes for Java phones and Windows Mobile PDAs, using respectively J2ME and .NET compact framework (ver 2.0). Such initial choice comes from two different and someway opposite demands:

- The possibility to run on almost every commercial device;
- The necessity to exploit all the hardware and OS capabilities.

Using Java (J2ME) meets the first requirement; while .NET CF meets the second one. So our J2ME implementation is less powerful but runs everywhere, while our .NET implementation works better with more features, but runs only on Windows Mobile terminals.

For the .NET CF implementation we started from the experience of an existing software package: the Carnegie Mellon Piccolo framework [11], which helped us to move forward. Multimodal objects can be thought as an extension of the Piccolo objects. They can be, as their Piccolo ancestors do, eventually linked one to the other in a graph, for creating a complete interface for both collecting inputs and managing outputs. For J2ME case, since Piccolo is not available for such platform, we started from scratch. In both cases we have used the same approach: the interface is a composition of different objects of the framework.

The interaction modes enabled by our multimodal framework are, at the moment, the following: point on specific buttons (e.g. a “Back” button), point on an object, draw/handwrite on the screen, speak. Output media supported are: image, text, Html, audio, video. The framework components are classified according to their capabilities to manage some of the available input/output channels. So the framework will contain objects like:

- “decorative panels” (just art, with no input/output feature),
- “input panels” (that are able to acquire input commands)
- “output panels” (that are able to show/play something to final user).
- “input/output panels” able to manage both features.

Audio acquisition should be continuous and should be sent in real time to the recognizers. This implies the necessity to manage streaming protocols (RTP) between MMUIManager and speech recognizer. Such a choice guarantees the absence of dead times, and a more natural interaction as the acquired audio stream is automatically sent while acquiring.

Stylus (sketch/handwriting) acquisition should take place in InkML format according to the W3C indication [3]. Ink traces are saved in InkML format.

J2ME implementation has some limitations. Actual limits are related to the fact that the basic J2ME libraries, are not able to:

- Support uplink streaming protocols;
- Support most of the audio/video formats.

Such limitations make our J2ME implementation less powerful than the .NET one because some features (continuous speech acquisition, output audio streaming) will work only on specific Java devices or won't work at all. It must be said that the situation could change soon as players like Google or SUN are investing resources in developing new and more powerful Java frameworks for mobile devices [14,15].

Fig. 2. Screen sample of MMUIManager .NET interface

5 Input Channels

As discussed, the MMUIManager is able to manage different kind of protocols. More specifically it is able to interact with server using web services and SIP/RTP. Coherently with our approach, we have focused our attention on how it is possible to use recognizers efficiently considering the fact that they are distributed over a network. An optimal solution to this problem must consider all the restrictions due to the current technology evolution and the available standard interface. First of all we have considered the Automatic Speech Recognizer (ASR) case. The best solution we have found for integrating a thin mobile client with and ASR is depicted in Fig. 3.

The need for a negotiation module comes from the fact that modern ASR interfaces are based on an emerging standard, the Media Resource Control Protocol (MRCP).

MRCP is a standard proposed to the IETF for enabling an optimal ASR resource allocation while managing service requests coming from many clients. MRCP does not define a transport protocol: it uses RTSP, SIP or a simple TCP connection. Thus for masking to our MMUIManager, as well as any other thin mobile client, the protocol complexities and keeping it as light as possible, we have designed a gateway module that convert HTTP requests into SIP/MRCP messages. At the end of the negotiation process, MMUIManager is able to open/close a streaming session (using the RTP protocol).

Once opened the stream, the user can send speech commands to the server. Apart the complexities of interpreting such commands in a multimodal context, which are not in the scope of this paper, we would like to discuss some of the limits/potentiality of such a user interface. In order to trigger the recognition process we have introduced a special keyword that makes the recognition start. The recognition ends after a pause in user speech. A trigger word could be for example "computer" and a valid user command "computer, give me information about this room". So if a user says: "give me information about this room", recognition does not start.

As mentioned several times in this paper, our goal is a system for managing synchronous coordinate synergic multimodality. Thus, the client has to collect many different input channels at the same time. At the actual state of development the channels that can be combined are: speak, point, sketch, handwriting.

The acquisition of "point" is simple: each time a user touch a sensible object on the screen, an HTTP call containing the object identifier is sent toward the server for being eventually processed together with other concurrent modes.

The acquisition of sketch/handwriting is more complex. The special multimodal objects is sensible to stylus inputs and traces are buffered locally and sent to the server using a standard format (InkML) and, after a pre-defined time interval, if the user does not add additional signs, the input are sent toward the front-end server via HTTP. Front-end server, in turn, is configured for routing the input acquired to a recognizer and to the fusion module.

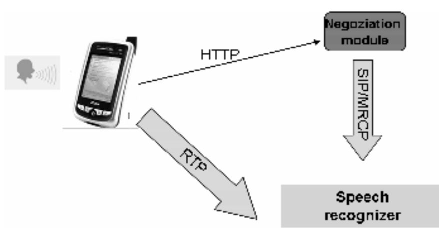


Fig. 3. Audio input channel

At the present time, Windows Mobile clients are based on an open source SIP/RTP stack called PJSIP [16] while for J2ME phones we have implemented a library for uploading speech signals using HTTP that, however, it is not suitable for transferring continuous speech signals. As said in the near future we expect substantial evolutions of the mobile Java world and a native support for both up and down stream.

The MMUIManager input modules are integrated with the following recogniser at the moment.

- ASR Loquendo [17] for speech;
- Jarnal for handwriting [18];
- Software based on original algorithms by the IRPPS (Istituto di Ricerche sulla Popolazione e le Politiche Sociali, Rome) for sketch recognition [19,20].

At least for speech recognisers, the solution described above should allow an easy replacement with equivalent products. For the sketch and handwriting recognisers, the integration strategy is not based on standards and, thus, tailored on the selected tools. This fact makes evident that a lot of work must be done for standardising recognisers and making them working in an open multimodal environments. Even the streaming of signs is not supported natively by streaming protocols: this is a domain of research that could lead to interesting results.

6 Output Channels

As discussed previously, the contents are not returned to the application in a single session: the server trigger the application to download a LIDIM file that contains instructions for every output channel managed by the MMUIManager.

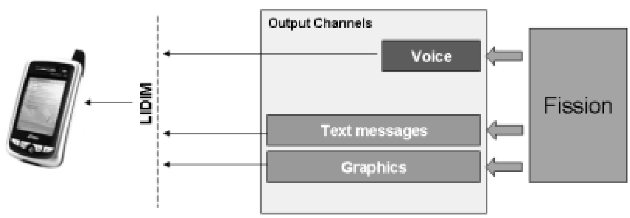


Fig. 4. Output channels schema

The output channels that the MMUIManager can process are depicted in fig. 4. The so-called Voice channel is a voice output channels that allows playing either pre-recorded audio resources or text sources which must be converted in voice (TTS) by a specific engine. According to our light-weight approach, the LIDIM file contains just the information about resource location (an address). Thus, it is MMUIManager responsibility to retrieve the remote resources using HTTP or RTP. For text resources, a TTS engine has the responsibility to transform text to speech according to the MMUIManager request and to stream it on the client.

The Text Message channel is able to show popup text messages over the interface after the page loading. Text messages can be used also for asking user confirmations.

The Graphics channel is the channel that contains information about the graphical interface layout. Notice that such output is related to text, audio or even video resources that the MMUIMAnager loads accordingly with user commands.

All the data on output channels can be included in a LIDIM page. Essentially Fission module has the task to “push” LIDIM page toward MMUIManager.

It is worth notice that the fission module is completely asynchronous. This is an important feature: input and output are completely detached and, thus, the server can take the initiative to modify the interface shown to final users even without an explicit request from the user himself. Essentially the server side has the capability to "push" content on devices depending on different criteria.

Such a feature is of fundamental importance: real adaptive application must be able to react to context changes even without an explicit request coming from the user (for example according to the usage context changes: physical, positional, etc…). In the following we will discuss the key elements of the fission module architecture.

Our multimodal fission module distributes information over one or more output channels selecting the output mode combination according to [21]; so the Fission module tasks are classified into three categories:

- *Content selection and structuring*: the presented content must be selected and arranged into an overall structure.
- *Modality selection*: the optimal modalities is determined based on the current situation of the environment.
- *Output coordination*: the output on each of the channels should be coordinated so that the resulting output forms a coherent presentation.

The first two tasks are delegated to the business logic module, which has the responsibility to process intelligently the information concerning the user context. We will not describe this module in this paper. The third task concerns the interaction with the clients and is under the fission module and MMUIManager responsibility.

Thus, the Fission module must be to locate the client, format opportunely a mark-up file following the business logic directives, and trigger the client content retrieval. Furthermore, the fission module must keep track of the contents sent. The MMUIManager, in turn, has the responsibility to load the multimodal objects, retrieve the contents and synchronize them for an optimal overall rendering.

As discussed previously, the fission module has been designed for allowing asynchronous interactions with the clients. More specifically, the fission module notifies the MMUIManager module using the INVITE SIP command (Session Initiation Protocol). This notification contains the URI of the LIDIM file that must be downloaded.

The complete Fission architecture is illustrated in the following figure:

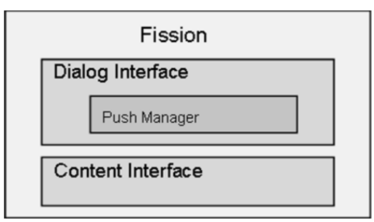


Fig. 5. Fission module internal architecture

This figure shows the following components:

- *Dialog Interface* is the module through which the business logic layer talks with the Fission module for the notification of new contents.

- *Push Manager* is responsible for the notification at the MMUIManager of the presence of new content for a specific client.
- *Content Interface* is the component through which MMUIManager module communicates with the Fission module for the final content retrieval.

As mentioned earlier, notifications to the MMUIManager, are based on the SIP protocol. This is possible only after a MMUIManager registration to the SIP server and, thus, after an explicit client identity declaration that is used all along the application lifetime. In Fig. 7 we show the operation sequence for enabling the SIP communication among the architecture modules.

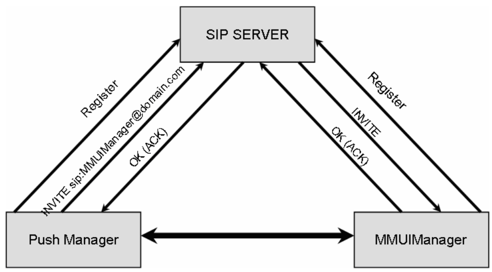


Fig. 6. Asynchronous push

We are now using the Asterisk SIP proxy using their API's for integrating it [22] which intermediary among the push manager and the client.

7 A Multimodal Output Language: LIDIM

The XUL-like approach we have followed for the building of client interface implies the need of a specific multimodal interface description language. The need for such a language comes from the evidence that a standard language for describing synergic multimodal interface does not exist.

A first possibility would have been to adapt an existing language to our requirements: synergic coordinate multimodality, multimedia output, modular interface. This proved to be difficult: SMIL [23] have been created for output and it is not simple adapt it to multimodal inputs, while X+V and SALT have not been created for synergic multimodality or to manage many input channels at the same time.

So we decided to create our own language. The result of our efforts is LIDIM.

LIDIM is able to manage the channels described in previous sections: graphic modular panel interface, audio (pre-recorded sources and TTS), pop-up text messages.

Let us see how LIDIM is able to describe information about all these channels.

The construction of the modular graphical interface is done using the tag <object>. For each object of the modular interface it can be specified one more output media (image, audio, video, text) and, eventually, one more input modality (point, sketch).

In the following sample we show how it is possible to instruct the MMUIManager to load a multimodal object for managing text, image, and audio resources and for accepting point and sketch inputs.

```
<object id="78678687">
      <output_media>
      <text uri="http://...."/>
      <image uri=""/>
      <audio uri=""/>
    </output_media>
    <input_mode>
      <point isActive="true"/>
      <sketch isActive="true"/>
    </input_mode>
</object>
```

Obviously LIDIM as only the goal to describe the output interface. the MMUIManager engine interpret it and aggregate the multimodal objects for creating the interface.

Text message are sent using the tag <message>.

```
  <message text="Are you sure you want to exit the application?" isDialog="true">
```

More or less the same for managing Voice and TTS channels (tag <voice>).

```
<voice src="Welcome to CHAT guide" mode="TTS" />

<voice src="http://..." mode="VoiceMail" />
```

Like HTML, LIDIM does not include the resources (images, text, audio/video resources) shown in the interface; it just contains the addresses where MMUIManager can get such resources. At the moment we are using LIDIM just for describing multimodal object composition and not for describing the screen layout: the MMUIManager disposes objects on screen according to its configuration. Furthermore at the time of writing, we are working on content synchronization on the client, considering all the possible synchronisation scenarios.

8 Conclusion and Future Works

We have presented an approach for developing multimodal/multimedia enabled mobile UI. This paper treats only a part of the CHAT project which goal is to make available a complete client/server platform for multimodal mobile services.

We have introduced the concept of multimodal objects. These objects are designed to be able to gather vocal, sketch and stylus inputs and to render multimedia output (text, graphics, audio and video). As a part of such approach we have introduced a language to describe this kind of UI. We have called this language LIDIM.

Apart from these results, we have built prototypes of client application, MMUIManager, able to render such UI. We have developed these applications with

.NET CF and J2ME. J2ME prototype has limited features as we are still unable to manage continuous speech acquisition on such platform and for limited standardization in audio/video format. .NET implementation is more complete but it can run only on Windows Mobile terminals, while J2ME implementation can run practically everywhere. Anyway J2ME rapid growth and the introduction of new mobile platform based on Java (JavaFX, Android) let us be optimistic in having soon a multiplatform implementation of MMUIManager.

Our next step will be to integrate context acquisition in the MMUIManager to have a real adaptivity according to the physical localization and context. To implement localization, we are working with technologies like GPS (for outdoor location), Bluetooth and Zigbee (for indoor location). To acquire the physical context data (luminosity, noise, etc.) we are now integrating a wireless sensor network. The idea that inspired us is that every single handset in the environment could be considered as a sensor and produce interesting data for determining the user context.

References

1. Frattini, G., Romano, L., di Scotto Carlo, V., Petriccione, P., Supino, G., Leone, G., Autiero, C.: Multimodal Architectures: Issues and Experiences. In: Meersman, R., Tari, Z., Herrero, P. (eds.) OTM 2006 Workshops. LNCS, vol. 4277, pp. 974–983. Springer, Heidelberg (2006)
2. Frattini, G., Gaudino, F., Scotto di Carlo, V.: Mobile multimodal applications on mass-market devices: experiences - 18th International Workshop on Database and Expert Systems Applications (DEXA 2007). In: Wagner, R., Revell, N., Pernul, G. (eds.) DEXA 2007. LNCS, vol. 4653, pp. 89–93. Springer, Heidelberg (2007)
3. http://www.w3.org/2002/mmi/
4. XUL (Last accessed December 12, 2007), http://www.mozilla.org/projects/xul/
5. Kirusa (Last accessed March 20, 2008), http://www.kirusa.com
6. Nuance (Last accessed March 20, 2008), http://www.nuance.com
7. X+V for the Next Generation Web, http://www.voicexml.org/specs/multimodal/
8. Niklfeld, G., Anegg, H., Pucher, M., Schatz, R., Simon, R., Wegscheider, F., Gassner, A., Jank, M., Pospischil, G.: Device independent mobile multimodal user interfaces with the MONA Multimodal Presentation Server. In: Proceedings of Eurescom Summit 2005, Heidelberg, Germany, April 27-29 (2005)
9. Reitter, D., Panttaja, E., Cummins, F.: UI on the fly: Generating a multimodal user interface. In: Proceedings of Human Language Technology conference 2004 / North American chapter of the Association for Computational Linguistics (HLT/NAACL 2004) (2004)
10. Filippo, F., Krebs, A., Marsic, I.: A framework for rapid development of multimodal interfaces. In: Proceedings of the 5th international conference on Multimodal interfaces, November 5-7, Vancouver, British Columbia, Canada (2003)
11. Piccolo (last accessed February 20, 2008), http://www.cs.umd.edu/hcil/jazz/index.shtml
12. Thinlet classic (Last accessed March 20, 2007), http://www.thinlet.com/
13. Zap. The mozilla SIP client (Last accessed April 13, 2007), http://croczilla.com/zap

14. Android (last accessed March 7, 2007), http://code.google.com/android/
15. JavaFX (last accessed March 7, 2007), http://www.sun.com/software/javafx/mobile/
16. PJSIP (last accessed March 7, 2007), http://www.pjsip.org/
17. Loquendo vocal technology and services, http://www.loquendo.com
18. Jarnal (Last accessed March 5, 2008), http://www.dklevine.co/general/software/tc1000/jarnal.htm
19. Avola, D., Ferri, F., Grifoni, P.: Formalizing Recognition of Sketching Styles in Human Centered Systems. In: Apolloni, B., Howlett, R.J., Jain, L. (eds.) KES 2007, Part II. LNCS (LNAI), vol. 4693. Springer. Heidelberg (2007)
20. Avola, D., Caschera, M.C., Ferri, F., Grifoni, P.: Ambiguities in Sketch-Based Interfaces. In: Proceedings of the Hawaii International Conference on System Sciences (HICSS 2007), vol. 290. IEEE Computer Society, Los Alamitos (2007)
21. Foster, M.E.: State of the art review: Multimodal fission (2002), http://www.hcrc.ed.ac.uk/comic/documents/deliverables/Del6-1.pdf
22. Asterisk SIP Server (last accessed March 7,2007), http://www.asterisk.org/
23. SMIL Specification, last accessed March 7th (2007), http://www.w3.org/TR/REC-smil/

Learning Rules for Semantic Video Event Annotation

Marco Bertini, Alberto Del Bimbo, and Giuseppe Serra

Media Integration and Communication Center
University of Florence
Firenze, Italy
{bertini,delbimbo,serra}@dsi.unifi.it

Abstract. Automatic semantic annotation of video events has received a large attention from the scientific community in the latest years, since event recognition is an important task in many applications. Events can be defined by spatio-temporal relations and properties of objects and entities, that change over time; some events can be described by a set of patterns.

In this paper we present a framework for semantic video event annotation that exploits an ontology model, referred to as Pictorially Enriched Ontology, and ontology reasoning based on rules. The proposed ontology model includes: high-level concepts, concept properties and concept relations, used to define the semantic context of the examined domain; concept instances, with their visual descriptors, enrich the video semantic annotation. The ontology is defined using the Web Ontology Language (OWL) standard. Events are recognized using patterns defined using rules, that take into account high-level concepts and concept instances. In our approach we propose an adaptation of the First Order Inductive Learner (FOIL) technique to the Semantic Web Rule Language (SWRL) standard to learn rules. We validate our approach on the TRECVID 2005 broadcast news collection, to detect events related to airplanes, such as taxiing, flying, landing and taking off. The promising experimental performance demonstrates the effectiveness of the proposed framework.

1 Introduction and Previous Work

Video archives have grown steadily in the recent years. There is therefore the necessity to develop effective and efficient methods for automatic annotation and retrieval of information. Indexing of these archives, based on low-level visual features like color and texture, often does not meet the user's information needs due to the semantic gap between the information that can be extracted from the visual data and the interpretation of the same visual data by a user in a given context. Recently ontologies have been regarded as an appropriate tool to overcome this semantic gap. An ontology consists of concepts, concept proprieties, and their relationships and provides a common vocabulary that overcome semantic heterogeneity of information. Ontology Web Language (OWL) and Semantic Web Rule Language (SWRL) have been approved by W3C as language standards for representing ontologies and performing reasoning using rules, respectively.

M. Sebillo, G. Vitiello, and G. Schaefer (Eds.): VISUAL 2008, LNCS 5188, pp. 192–203, 2008.

Recently several EC projects have addressed the problem of using ontologies for semantic annotation and retrieval by content from audio-visual digital libraries, among them AceMedia [1], Aim@Shape [2], Boemie [3] and VidiVideo [4].

Many researchers have built integrated system where the ontology provides the conceptual view of the domain at the schema level, and appropriate classifiers play the role of entities detectors. Once the observations are classified, the ontology is exploited to have an organized semantic annotation, establishing links between concepts and disambiguating the results of classification [5,6].

Other researches have directly included in the ontology an explicit representation of the visual knowledge to perform reasoning not only at the schema level but also at the data level. Staab et al. [7] defined three separate ontologies that respectively modeled the application domain, the visual data and the abstract concepts, to perform the interpretation of video scenes. Automatically segmented image regions were modeled through low-level visual descriptors and associated to semantic concepts using manually labeled regions as training set. Kompatsiaris et al. [8] included in the ontology instances of visual objects that were used as references to perform the classification of the entities observed in video clips. They used as descriptors low-level perceptual features like color homogeneity, components distribution, and spatial relations. A similar solution was presented by Bertini et al. in [9], using generic and domain specific descriptors and introducing mechanisms for updating the prototypes of the visual concepts of the ontology, as new instances of visual concepts are added to the ontology; the prototypes are used to classify the events and objects observed in video sequences.

For event recognition several authors have exploited the ontology schema using temporal reasoning over objects and events. Snoek et al. [10] performed annotation of sport highlights using rules that exploited face detection results, superimposed captions, teletext and excited speech recognition, and Allen's logic to model temporal relations between the concepts in the ontology. Francois et al. [11] defined a special formal language to define ontologies of events and used Allen's logic to model the relations between the temporal intervals of elementary events, so as to be able to assess complex events in video surveillance. Haghi et al. [12] proposed to use temporal RDF to model temporal relationships in the ontology and provided examples of simple queries with temporal relationships between events. Bai et al. [13] applied temporal reasoning with temporal description logic to perform event annotation in soccer video, using a soccer ontology. All of these methods defined rules, used to describe events, that were created by human experts; thus, these approaches are not practical for the definition of a large set of actions.

To overcome this problem some researchers have studied techniques to learn automatically a set of rules. Dorado et al. [14] performed video annotation based on learned rules that infer high-level concepts from low-level features using decision tree technique. Shyu et al. [15] proposed a method to annotate rare events and concepts based on set of rules that use low-level and middle-level features. A decision tree algorithm is applied to the rule learning process. Moreover they addressed the imbalance problem of positive and negative examples in the case of rare event/concept using data mining techniques. Liu et al. [16] proposed a method to enhance accuracy of semantic concepts detection, using association mining techniques to imply the presence of a concept from the co-occurrence of other high-level concepts. None of these three works is based on ontologies.

These methods that learn a set of rules by exploiting decision tree algorithms and low-level features, or simple junctions of high-level concepts, are not enough expressive to describe complex events. For example consider the event *A person enters in a secured area*. This event can not be described using only the low-level descriptors of the person and of the area, or using the co-occurrence of the high-level concepts *person* and *secured area* since the person may stay outside of it, or may have always been inside it; instead it is required to take into account the temporal evolution of the characteristics and features of the objects and entities. This event can be fully described and modelled using first-order logic. A sentence that describes the events is: IF a person is outside of the secured area in the time interval t_1 AND the same person is in the secured area in the time interval t_2 AND t_1 is before t_2 THEN that person has entered the secured area; this sentence can be translated in the following fragment of first-order logic language:

$$
\begin{aligned}
&IF\ person(p)\ \wedge\ personOutsideOfSecuredArea(p, t_1) \wedge \\
&personIsInSecuredArea(p, t_2) \wedge before(t_1, t_2) \\
&THEN\ personEntersSecuredArea(p)
\end{aligned}
$$

where p is a variable that can be bound to any person and *t1* and *t2* are variables that are used to represent time intervals.

In this paper we propose a framework for video event annotation that exploits the Pictorially Enriched Ontology model [9] that includes concepts and their visual descriptors, and a method to learn sets of first-order logic rules that describe events defined in the ontology. Events can be described by spatio-temporal relations and properties of objects and entities, that change over time. The learned rules, defined using the SWRL, are applicable directly to an ontology defined using the OWL. The proposed learning method is an adaptation of the First Order Inductive Learner technique (FOIL [17]) to the Semantic Web technologies; for convenience this method will be referenced in the following as FOILS. This approach permits to create an ontology structure that allows to perform automatic semantic annotation of video sequences matching visual descriptors and recognizing events described using automatically learned rules. Moreover the learning approach used is more expressive than the previous methods because it defines rules through the first-order logic theory. To demonstrate the applicability to the automatic event annotation we show several events that can be recognized by automatically learned rules. In particular our tests are performed on the definition of some events related to airplane entities, defined in the LSCOM ontology.

2 Automatic Video Annotation Framework

Our proposed framework, shown in Fig. 1, consists of three major components. In the shot segmentation and feature extraction component, the video is divided in syntactic units, low-levels features are extracted and objects and entities are identified. The Pictorially Enriched Ontology component includes a formal definition of a specific video domain and video structure. Rule-based reasoning is performed on high-level concepts and concepts instances for the automatic annotation of events. Rules are directly learned from the ontology using the FOILS algorithm.

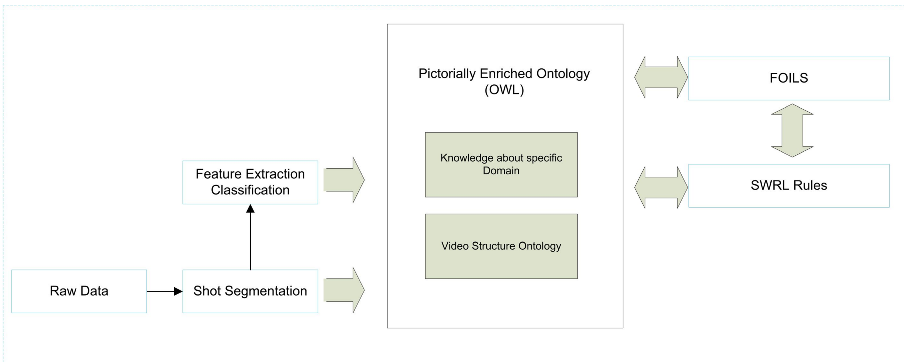


Fig. 1. Framework of the system

2.1 Video Parsing and Feature Extraction

Video segmentation involves temporal partitioning of the video into units which serve as the basis for descriptor extraction and semantic annotation. In this work, shots are adopted as the basic syntactic unit, while video clips (video sequences possibly composed by more than one shot) are used as annotation units. For each shot visual descriptors such as color histograms, edge maps, etc. are extracted to perform a rough segmentation of each frame. Appropriate classifiers are applied to identify objects or entities. The feature extractors are used to provide the visual descriptors associated to the visual concepts of the ontology. These descriptors, that may be generic or domain specific, are then used to characterize concepts instances; this characterization allows to select the most representative concepts as visual prototypes of a concept, and allow to perform reasoning based on the visual appearance of a concept.

2.2 Pictorially Enriched Ontology

The Pictorially Enriched Ontology defines formally the domain of interest. In Fig. 2 is shown a simplified view of the main concepts used to represent the events related to airplanes, as studied in the use case. Video structure and visual

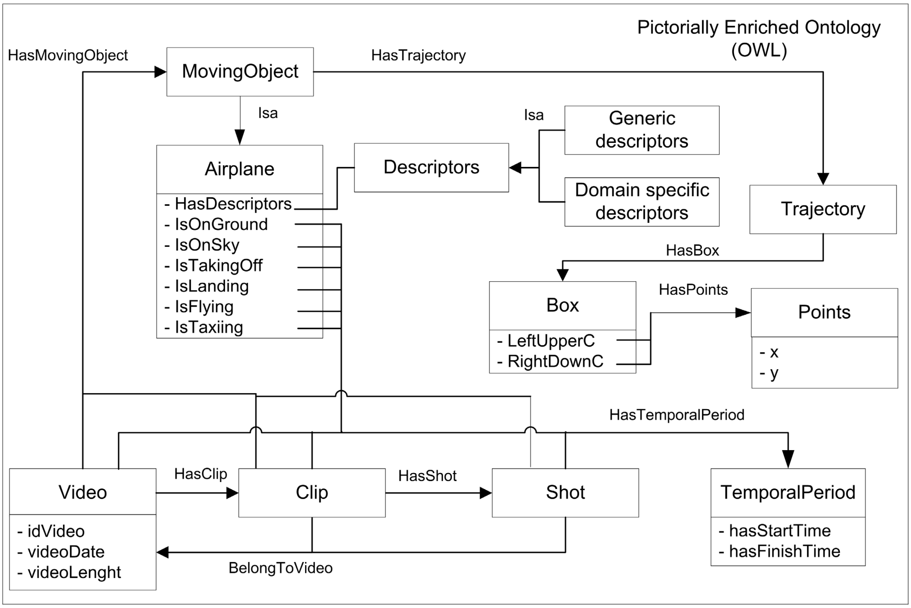


Fig. 2. Main concepts, relations and properties of the airplane events ontology

descriptors associated to the visual concepts are stored in the ontology according to the features extraction and classifier detection results. The concept instances that are associated to visual descriptors can be used as matching references for the entities that have to be annotated. In our experiments the airplane concept is associated with color histograms, that are used by the tracker to identify the instances of the detected airplanes in a video sequence.

2.3 First-Order Rule Learning

Terminology. To describe correctly the algorithm for learning sets of first-order rules, let us introduce some basic terminology from formal logic. All expressions are composed of constants (e.g. *Airplane1*, *Boeing-747*), variables (e.g. x, y), predicate symbols (e.g. *HasTrajectory*, *GreaterThan*) and function symbols (e.g. *duration*). The difference between predicates and functions is that predicates have value of *True* or *False*, whereas functions may have any constant as their value. In the following we will use lowercase for functions and capitalized symbols for predicates. A term is any constant, any variable, or any function applied to any term. A literal is any predicate or its negation applied to any term. If a literal contains a negation symbol (¬), we call it *negative literal*, otherwise *positive literal*. A *clause* is any disjunction of literals, where all variables are assumed to be universally quantified. A *Horn clause* is a clause containing at most one positive literal, as shown in the following:

$$H \vee \neg L_1 \vee \neg L_2 \ldots \vee \neg L_n$$

where H is the positive literal, and $\neg L_1 \vee \neg L_2 \ldots \vee \neg L_n$ are negative literals. It is equivalent to:

$$(L_1 \wedge L_2 \ldots \wedge L_n) \rightarrow H$$

which is equivalent to the following:

$$IF\ (L_1 \wedge L_2 \ldots \wedge L_n)\ THEN\ H$$

The Horn clause precondition $L_1 \wedge L_2 \ldots \wedge L_n$ is called *clause body*; the literal H that forms the post-condition is called the *clause head*.

First-Order Inductive Learner for SWRL technique. FOILS, first-order inductive learner for SWRL technique is an adaptation of the FOIL algorithm to the SWRL standard. The hypotheses learned by FOILS, similarly to FOIL, are sets of first-order rules, where each rule is similar to a Horn clause with the limitation that literals are not permitted to contain function symbols, in order to reduce the complexity of the hypothesis space search. At the beginning the algorithm starts with the *head* that we want to find in the rule and an empty or initial *body*. The algorithm iterates searching the new literals that have to be added to the body of the rule. The search is a general-to-specific search through the space of hypotheses, beginning with the most general preconditions possible (the empty or initial precondition), and adding literals one at a time to specialize the rule until it avoids all negative examples, or when no more negative examples are excluded for a certain number of loops. Two issues have to be addressed: the generation of hypothesis candidates and the choice of the most promising candidate.

Generating hypothesis candidates. Suppose that the current rule being considered is:

$$(L_1 \wedge L_2 \ldots \wedge L_n) \rightarrow P(x_1, x_2, \ldots, x_k)$$

where $(L_1 \wedge L_2 \ldots \wedge L_n)$ are literals forming the current rule preconditions and where $P(x_1, x_2, \ldots, x_k)$ is the literal that form the rule *head*. FOILS generates candidate specializations of this rule by considering new literals L_{n+1} that fit one of the following forms:

- $Q(v_1, \ldots, v_r)$ where Q is any predicate name occurring in *Predicates* and where the v_i are either a new variable or a variable already present in the rule. At least one of the v_i in the created literal must already exist as a variable in the rule.
- $Equal(x_j, x_k)$ where x_j and x_k are variables already present in the rule.

We observe that in FOIL there is another rule for generation of new candidates: it is the negation of either the above form of literals. This rule can not be exploited in our algorithm because it is not permitted by SWRL.

Most promising literal. To select the most promising literal from the candidates generated at each step, FOILS, similarly to FOIL, considers the performance of the rule over the training data. The evaluation function used to estimate the utility of adding a new literal is based on the number of positive and negative bindings covered before and after adding the new literal. More precisely consider some rule R, and a candidate literal L that might be added to the body of R. Let R' be the rule created by adding the literal L to rule R. The value of adding L to R is defined as:

$$Foil\_Gain(L,R) \equiv t\left(log_2 \frac{p_1}{p_1+n_1} - log_2 \frac{p_0}{p_0+n_0}\right)$$

where p_0 is the number of positive bindings of rule R, n_0 is the number of negative bindings of R, p_1 is the number of positive bindings of rule R' and n_1 is the number of negative bindings of R'. Finally, t is the number of positive binding of rule R that are still covered after adding literal L to R. When a new variable is introduced into R by adding L, then any original binding is considered to be covered as long as some binding extending it is present in the bindings of R'.

3 Use Case

We have applied the automatic video annotation framework to the detection of events related to airplanes, selecting them from the revised list of LSCOM events/activities [18]. Four events related to an airplane concept are analyzed: airplane flying, airplane takeoff, airplane landing, airplane taxiing. In Fig. 2 a simplified schematization of the ontology defined for these events is shown; for the sake of simplicity the visual descriptors associated to the airplane concept are not reported.

These events can be detected using airplane, sky and ground detector and the temporal relationship between these concepts. For example, the evolution of an *airplane takeoff* event video is composed by a view of the airplane moving on the ground followed by a view of airplane on sky. An airplane detector has been created using the Viola&Jones object detector. The positive and negative examples used to train the detector have been selected from standard image datasets such as Caltech, VOC2005 and VOC2006. The negative examples used are images of man-made objects (e.g. other vehicles like cars, buses and motorcycles), outdoor scenes, animals and persons, various objects. The sky and ground detectors implemented are not used to classify all the parts and segments of each frame, but only locally, next to the airplane position, because it is enough to know if the airplane is on ground or in sky. The sky/ground detector evaluates statistical parameters of the luminance of the blobs around the detected airplane. Finally using a tracker, based on an improved version of the particle filter [19], we can determine the temporal evolution of the trajectory of airplane. The detected airplane, its bounding box trajectory and sky and ground detection are inserted in the ontology. Using learned SWRL rules the airplane takeoff, airplane landing, airplane flying and airplane taxiing events are identified. In Fig. 3 two examples of landing and take-off events are shown. These rules are learned from a set of positive and negative examples stored in the ontology using the FOILS technique, described in the previous section. To illustrate how the FOILS algorithm

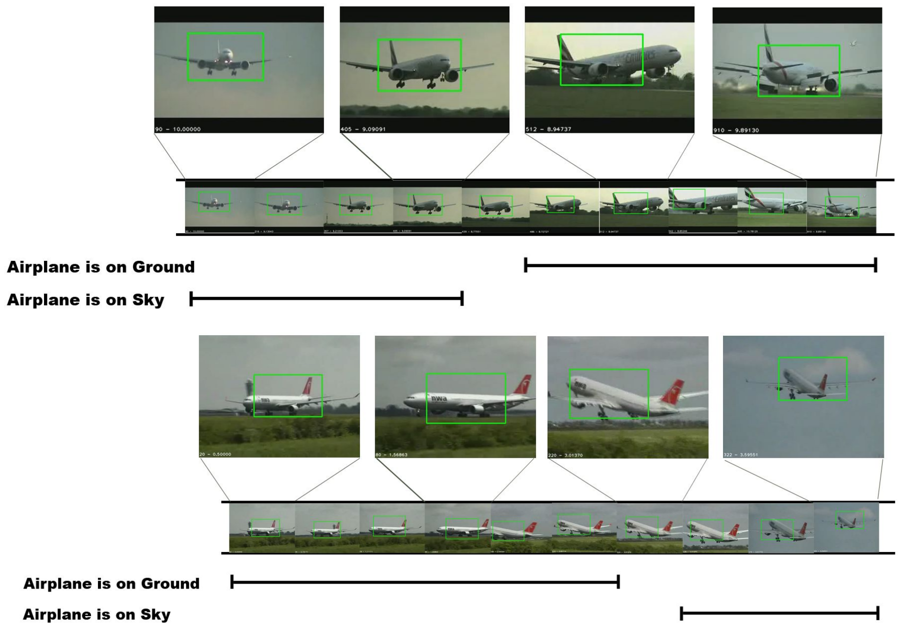


Fig. 3. Examples of airplane landing and take-off events. For each event the results of airplane detection and tracking are shown, along with a temporal model of the event.

works we consider, for example, the target literal *AirplaneIsTakingOff*. The process starts with an initial rule written in SWRL. that models the take-off of airplane a within the video clip c:

$$Airplane(?a) \wedge Clip(?c) \rightarrow AirplaneIsTakingOff(?a, ?c)$$

The initial candidates are all the classes and properties defined in the ontology domain and temporal properties used to encode Allen's logic. At each step the most promising literal is added, considering the performance of the rules over the training data.

4 Experimental Results

In the first part of the experiment we evaluate the performance of the airplane detector. We have trained five different detectors, using five configurations, with different numbers of positive and negative examples, image window sizes, and learning steps. Results are reported in Tab. 1. To train the fifth detector the number of positive examples of airplanes has been increased, adding more images of frontal and rear views of airplanes. The first three detectors did not provide an acceptable performance in terms of precision, as shown in the table. The decrease of the precision value between the fourth and fifth detector is mainly due to the fact that the detector may provide multiple detections for the same airplane,

Table 1. Precision and recall of airplane detector

N. detector	N. steps	Neg. examples	Pos. examples	Window size	Precision	Recall
1	17	3000	800	50×30	0.20	0.74
2	18	1500	800	50×30	0.19	0.83
3	20	1500	800	50×30	0.32	0.65
4	20	1500	800	25×10	0.75	0.55
5	22	1500	1040	50×30	0.41	0.66

whose bounding boxes are overlapping, and these multiple detections have been counted as falses; without considering this overlapping effect the precision is comparable with that of the fourth detector. Considering this fact, the fifth detector has been selected and used in the following experiment.

To test the effectiveness of the learned rules we have used them to recognize events in a large dataset, that comprises 100 videos containing airplane events taken from the web[1] and 65 Trecvid 2005 videos.

The set of videos selected from the web video sharing sites (called on the following as Web Dataset) is available online, along with the airplane detector[2]. The Trecvid videos were selected from those reported in the LSCOM development set as containing the concepts *airplane_takeoff*, *airplane_landing* and *airplane_flying*, after a manual inspection that eliminated some errors of the ground truth (e.g. videos that contained rockets or helicopters instead of airplanes). Since the concept *airplane_taxiing* is not defined in LSCOM we inspected the videos annotated as containing *airplane* to select some videos that contained this event.

We have used an implementation of the FOILS algorithm, described in Sect. 2.3, to learn the SWRL rules that model the airplane events. The videos of the Web Dataset have been used to learn the rules. For each event that we want to learn we randomly select one third of the videos containing that event as positive examples, and one third of the videos of the other events as negative examples. In Tab. 2 the learned rules are shown. For each rule we present the initial rule and the final rule obtained using FOILS. The learned rules recognize events within clips; this allows to cope with the case in which an event is shown using more than one shot. In some cases we can observe that FOILS learns some literals that are not necessary for the event definition, however this does not affect negatively the performance of the rule. This fact may happen since FOILS does not take into account the structure of the ontology; an example is the $MovingObject(?p)$ literal in the landing and taking-off rules, that is not necessary due to the fact that in our ontology this concept is an hypernym of airplane.

We have then applied the rules to the videos, evaluating the results, in term of precision and recall, for Web Dataset and Trecvid 2005 video separately and together, as shown in Tab. 3. As it can be observed the overall results for all the

[1] YouTube (http://www.youtube.com), Alice Video (http://dailymotion.alice.it), PlanesTV (http://www.planestv.com/planestv.html), Yahoo! Video (http://it.video.yahoo.com)

[2] http://www.micc.unifi.it/dome

Table 2. Rules for airplane events recognition, obtained using FOILS

Rule: Airplane TakingOff
Initial rule:
$Airplane(?p) \wedge Clip(?c) \rightarrow IsTakingOff(?p, ?c)$
Result rule:
$Airplane(?p) \wedge Clip(?c) \wedge IsOnSky(?p, ?g1) \wedge IsOnGround(?p, ?g2) \wedge$
$Temporal : after(?g1, ?g2) \wedge HasTemporalPeriod(?c, ?g3) \wedge Temporal : contains(?g3, ?g1) \wedge$
$Temporal : contains(?g3, ?g2) \wedge MovingObject(?p) \rightarrow IsTakingOff(?p, ?c)$

Rule: Airplane Landing
Initial rule:
$Airplane(?p) \wedge Clip(?c) \rightarrow IsLanding(?p, ?c)$
Result rule:
$Airplane(?p) \wedge Clip(?c) \wedge IsOnSky(?p, ?g1) \wedge IsOnGround(?p, ?g2) \wedge$
$Temporal : notafter(?g1, ?g2) \wedge HasTemporalPeriod(?c, ?g3) \wedge Temporal : contains(?g3, ?g1) \wedge$
$Temporal : contains(?g3, ?g2) \wedge MovingObject(?p) \rightarrow IsLanding(?p, ?c)$

Rule: Airplane Flying
Initial rule:
$Airplane(?p) \wedge Clip(?c) \rightarrow AirplaneFlying(?p, ?c)$
Result rule:
$Airplane(?p) \wedge Clip(?c) \wedge IsOnSky(?p, ?g1) \wedge$
$HasTemporalPeriod(?c, ?g2) \wedge Temporal : contains(?g2, ?g1) \rightarrow IsFlying(?p, ?c)$

Rule: Airplane Taxiing
Initial rule:
$Airplane(?p) \wedge Clip(?c) \rightarrow IsTaxiing(?p, ?c)$
Result rule:
$Airplane(?p) \wedge Clip(?c) \wedge IsOnGround(?p, ?g1) \wedge$
$HasTemporalPeriod(?c, ?g2) \wedge Temporal : contains(?g2, ?g1) \rightarrow IsTaxiing(?p, ?c)$

Table 3. Precision and recall of Airplane flying, Airplane takeoff, Airplane landing, Airplane taxiing for different datasets

Data Set	**Airplane Action**	**Precision**	**Recall**
Web Dataset	Airplane flying	0.96	0.94
Web Dataset	Airplane takeoff	0.76	0.80
Web Dataset	Airplane landing	0.84	0.96
Web Dataset	Airplane taxiing	1	0.76
TRECVID 2005	Airplane flying	0.94	0.5
TRECVID 2005	Airplane takeoff	0.3	0.42
TRECVID 2005	Airplane landing	0.66	0.66
TRECVID 2005	Airplane taxiing	1	0.76
Web Dataset + TRECVID 2005	Airplane flying	0.96	0.71
Web Dataset + TRECVID 2005	Airplane takeoff	0.61	0.70
Web Dataset + TRECVID 2005	Airplane landing	0.90	0.90
Web Dataset + TRECVID 2005	Airplane taxiing	0.94	0.84

rules are extremely promising. Since the rules that describe flying and landing are more simple, their performance is better than that of the rules that model landing and taking-off. The main difference in the performance results between the two datasets is related to the quality of the images and to the presence of superimposed graphics, that were present only in the Trecvid news videos. Since the performance of the rules is dependent on the performance of the detectors and tracker we have investigated the cases in which the rules failed. The main cause of failure is due to the performance of the simple sky/ground detector, that uses only the luminance information. In a few cases the fault was the airplane detector, especially when superimposed graphics and text covered the appearance of the airplane.

5 Conclusions

In this paper a framework for automatic event video annotation has been presented. A Pictorially Enriched Ontology has been defined, to perform automatic semantic annotation of videos, and a set of rules, used to describe events, has been learned from positive and negative video examples, using an adaptation of the First Order Inductive Learner technique to the Semantic Web Rule Language. The performance has been tested using different datasets to demonstrate the effectiveness of the proposed approach. Our future work will investigate techniques to incorporate learning of constants and function symbols in our framework, to permit to insert numerical temporal specifications in the event description.

Acknowledgment. This work was partially supported by the Information Society Technologies (IST) Program of the European Commission VidiVideo Project (Contract FP6-045547).

References

1. aceMedia project (IST EC-FP6): Integrating knowledge, semantics and content for user-centered intelligent media services, `http://www.acemedia.org/`
2. Aim@Shape project (IST EC-FP6): Advanced and innovative models and tools for the development of semantic-based systems for handling, acquiring, and processing knowledge embedded in multidimensional digital objects, `http://www.aimatshape.net/`
3. Boemie project (IST EC-FP6): Knowledge acquisition from multimedia content, `http://www.boemie.org/`
4. VidiVideo project (IST EC-FP6): Improving the accessibility of video, `http://www.vidivideo.info/`
5. Snoek, C., Huurnink, B., Hollink, L., de Rijke, M., Schreiber, G., Worring, M.: Adding semantics to detectors for video retrieval. IEEE Transactions on Multimedia 9(5), 975–986 (2007)
6. Zha, Z.J., Mei, T., Wang, Z., Hua, X.S.: Building a comprehensive ontology to refine video concept detection. In: Proc. of ACM Int'l Workshop on Multimedia Information Retrieval, Augsburg, Germany, pp. 227–236 (September 2007)
7. Simou, N., Saathoff, C., Dasiopoulou, S., Spyrou, E., Voisine, N., Tzouvaras, V., Kompatsiaris, I., Avrithis, Y., Staab, S.: An ontology infrastructure for multimedia reasoning. In: Atzori, L., Giusto, D.D., Leonardi, R., Pereira, F. (eds.) VLBV 2005. LNCS, vol. 3893. Springer, Heidelberg (2006)

8. Dasiopoulou, S., Mezaris, V., Kompatsiaris, I., Papastathis, V.K., Strintzis, M.G.: Knowledge-assisted semantic video object detection. IEEE Transaction on Circuits and Systems for Video Technology 15(10), 1210–1224 (2005)
9. Bertini, M., Del Bimbo, A., Torniai, C., Cucchiara, R., Grana, C.: Dynamic pictorial ontologies for video digital libraries annotation. In: Proc. ACM Int'l Workshop on the Many Faces of Multimedia Semantics, Augsburg, Germany, pp. 47–56 (2007)
10. Snoek, C., Worring, M.: Multimedia event-based video indexing multimedia event-based video indexing using time intervals. IEEE Transactions on Multimedia 7(4), 638–647 (2005)
11. Francois, A., Nevatia, R., Hobbs, J., Bolles, R., Smith, J.: VERL: an ontology framework for representing and annotating video events. IEEE Multimedia 12(4), 76–86 (2005)
12. Qasemizadeh, B., Haghi, H., Kangavari, M.: A framework for temporal content modeling of video data using an ontological infrastructure. In: Proc. Semantics, Knowledge and Grid, Guilin, China, November 2006, p. 38 (2006)
13. Bai, L., Lao, S., Jones, G., Smeaton, A.F.: Video semantic content analysis based on ontology. In: Proc. of Int'l Machine Vision and Image Processing Conference, Maynooth, Ireland, pp. 117–124 (2007)
14. Dorado, A., Calic, J., Izquierdo, E.: A rule-based video annotation system. Circuits and Systems for Video Technology. IEEE Transactions 14(5), 622–633 (2004)
15. Shyu, M.L., Xie, Z., Chen, M., Chen, S.C.: Video semantic event/concept detection using a subspace-based multimedia data mining framework. Multimedia, IEEE Transactions 10(2), 252–259 (2008)
16. Liu, K.H., Weng, M.F., Tseng, C.Y., Chuang, Y.Y., Chen, M.S.: Association and temporal rule mining for post-filtering of semantic concept detection in video. Multimedia, IEEE Transactions 10(2), 240–251 (2008)
17. Quinlan, J.R.: Learning logical definitions from relations. Mach. Learn. 5(3), 239–266 (1990)
18. Kennedy, L.: Revision of LSCOM event/activity annotations, DTO challenge workshop on large scale concept ontology for multimedia. Advent technical report #221-2006-7, Columbia University (December 2006)
19. Bagdanov, A.D., Del Bimbo, A., Dini, F., Nunziati, W.: Improving the robustness of particle filter-based visual trackers using online parameter adaptation. In: Proc. of IEEE Int'l Conference on AVSS, London, UK, pp. 218–223 (September 2007)

Two Step Relevance Feedback for Semantic Disambiguation in Image Retrieval

Daniel Heesch[1] and Stefan Rüger[2]

[1] Pixsta Research
London
United Kingdom
daniel.heesch@pixsta.com
[2] Knowledge Media Institute
The Open University
Milton Keynes
United Kingdom
s.rueger@open.ac.uk

Abstract. This paper presents a new approach to the problem of feature weighting for content based image retrieval. If a query image admits to multiple interpretations, user feedback on the set of returned images can be an effective tool to improve retrieval performance in subsequent rounds. For this to work, however, the first results set has to include representatives of the semantic facet of interest. We will argue that relevance feedback techniques that fix the distance metric for the first retrieval round are semantically biased and may fail to distil relevant semantic facets thus limiting the scope of relevance feedback. Our approach is based on the notion of the NN^k of a query image, defined as the set of images that are nearest neighbours of the query under *some* instantiation of a parametrised distance metric. Different neighbours may be viewed as representing different meanings of the query. By associating each NN^k with the parameters for which it was ranked closest to the query, the selection of relevant NN^k by a user provides us with parameters for the second retrieval round. We evaluate this two step relevance feedback technique on two collections and compare it to an alternative relevance feedback method and to an oracle for which the optimal parameter values are known.

1 Introduction

The question of how to infer semantic similarity between objects from their feature representations is central to content-based information retrieval. Not all features are equally well suited to pick up relevant properties of an object and often it is combinations of features that work best. How to combine evidence from different features depends on the particular object and is not known a priori. The problem is even more acute when objects are semantically rich and admit to a number of different interpretations, such as often holds for images.

M. Sebillo, G. Vitiello, and G. Schaefer (Eds.): VISUAL 2008, LNCS 5188, pp. 204–215, 2008.

Fig. 1. An image (leftmost) and the most similar images under different features

The optimal feature combination then also depends on the particular semantic facet the user is interested in.

As an illustration of semantic ambiguity or *polysemy* consider Figure 1. The three images to the right are visually closest to the image on the left under one of three different visual features (from left to right, texture, local colour distribution, global colour distribution). By ignoring colour content altogether the texture feature does not pick up the autumnal character of the query but captures well the distinct texture of leafy vegetation. The global colour feature is insensitive to how colours are distributed locally and ranks an image high if the overall proportions are similar as in a close-up view of fallen leaves. We may view the three images as representing different semantic facets of the query: "tree-like vegetation", "trees with autumn foliage", and "autumn colours". Which of these a user is interested in, and thus, which feature is to be given greater weight, cannot be decided on the basis of the query image alone. In such circumstances, some form of user feedback on the relevance of retrieved images becomes a methodological imperative. Indeed, relevance feedback features in many early research systems (e.g. [12], [14]), and has been researched extensively in more recent times (e.g. [5], [7], [10], [11], [15]). A large proportion of relevance feedback techniques are concerned with optimising parameters of a distance metric, often these are weights associated with feature-specific similarity scores. A widely overlooked issue is that of initialising feature weights. The standard practice of assigning uniform weights for the purpose of the first retrieval step renders the system semantically biased simply because any particular weight setting will favour some semantic facets over others. If users are interested in an aspect for which the initial setting is ill-suited, the scope for positive relevance feedback may be greatly compromised as few or no relevant items are brought to the surface in the first round. This limitation is bound to become more apparent as the image collections grow in size and users' information needs become more specific.

We explicitly address the problem of parameter initialisation by employing the idea of an image's NN^k [6]. This is the set of all images that are most similar to that image under *some* feature combination. Instead of retrieving with one weight set, we compute the top-ranked image under all possible weight vectors. The resulting set of images provide us with a pictorial representation of at least parts of the semantic spectrum of the query image. We forego high precision among the set of NN^k but thereby increase the chance that at least one of the returned images is relevant to a particular user. The set of NN^k now gives us a powerful way to perform relevance feedback and to lift precision in the second step. The key is to associate each NN^k with the weight set under which it has been retrieved. By selecting NN^k, users thus implicitly select optimal distance metrics under which the chosen image is returned top. Following relevance feedback, images in the collection can be ranked according to this particular distance metric. The strength of this two step relevance feedback method therefore derives from two properties: it is semantically unbiased in the first step and it provides an effective way of associating images with optimal distance metrics to raise performance in the second step.

We tested the proposed method on two image collections of 8,200 and 32,000 images, respectively, and compared its performance with the relevance feedback method described in [12] as well as an oracle that knows the best distance metric for individual queries.

The paper is organised as follows. In Section 2 we describe related work. In Section 3 we define the set of NN^k and analyse some of their properties. Section 4 describes our relevance feedback technique and presents details of the evaluation. Section 5 reports results and Section 6 concludes the paper.

2 Related Work

In [1] relevance feedback is given on a set of images that are created on the fly by modifying segments of the query image in terms of shape, size and colour. Each modification can be viewed as representing a particular parameter setting under which the modified query is close to the original query. Positive relevance feedback is then utilised to update feature weights as in [13]. Although its motivation is different, the method is operationally similar to our approach. Instead of retrieving with a fixed parameter set, the user is presented with the spectrum of possible modifications. It is through interaction with these that parameters are initialised for the second step. The major limitations of the method are its dependence on good segmentation results and the small number of useful modifications that can be applied to the segments. When an image cannot meaningfully be segmented, users may find it difficult to judge the relevance of the modified images.

The idea of Bayesian sets [3, 8] shares the same motivation with ours but is otherwise very different. Based on several items exemplifying a certain concept, or, in our context, a certain information need, the system scores other items according to how well they fit into a set containing the exemplars. The requirement of submitting several items is key as the semantic facet can then be divined without recourse to relevance feedback.

3 NN^k and Their Properties

In [6], we define an image p to be an NN^k of image q if and only if there exists at least one convex combination of feature-specific distance functions $d_f(\cdot, q)$ for which p has minimal distance to q. Formally, p is an NN^k of q if and only if

$$\arg\min_i \left(\sum_{f=1}^{k} w_f d_f(i,q) \right) = p \tag{1}$$

for some $w = (w_1, w_2, \ldots, w_k)$ with $w_f \geq 0$ and $\sum w_f = 1$. k denotes the number of features. Note that we impose a convexity constraint on the weights without which *every* image would be an NN^k for *some* weight combination. Because weights are allowed to vary continuously, we cannot practically compute the NN^k for every weight set but need to discretise the space of all w that satisfy the convexity constraint and compute the weighted sum at a finite number of points. For $n+1$ grid points along each dimension of the weight space and k features, the total number of grid points can be shown to be $\binom{n+k-1}{n}$ [6].

3.1 Computational Speed-Up

We may reduce the computational cost inflicted at every grid point as follows. Recall that each vertex corresponds to a particular weight vector w. The nearest neighbour of the query for that weight vector is the image with the smallest distance to the query. As we recurse through the grid, most pairs of consecutive vertices differ in only two components of the weight vector. Instead of computing the distance D as $\sum_{f=1}^{k} w_f d_f$ involving k multiplications and $k-1$ additions, we can exploit the fact that if two successive vertices only differ with respect to two dimensions, $k-2$ of the terms of the product have already been computed at the first vertex. In such cases we can obtain the distance at the new vertex D^j by subtracting a correction term from the distance at the previous vertex D^{j-1}. Let F be the set of indices whose weights differ between the two vertices. The update is then

$$D^j = D^{j-1} + \sum_{f:w_f \in F} (w_f^j - w_f^{j-1}) d_f.$$

For a collection of 8,200 images the performance benefits are quite appreciable and become greater as we increase the resolution of the grid. With eight features and six grid points per dimension, the update method speeds up performance by almost 40%.

3.2 Iso-NN^k Regions

We define an iso-NN^k region as the region in $\mathbb{R}^k$ that contains all the weight vectors for which the NN^k is the same. Since each point of the subspace is associated with exactly one NN^k, the set of NN^k partition the weight space. The boundaries of the partitions can in principle be determined arbitrarily close by increasing the resolution n. For $k=3$, the set of permissible weight vectors lie in the two-dimensional triangular subspace of $\mathbb{R}^3$. Figure 2 shows the partitioning of the weight space into iso-NN^k regions for a particular query from the Corel collection, three features and varying resolution.

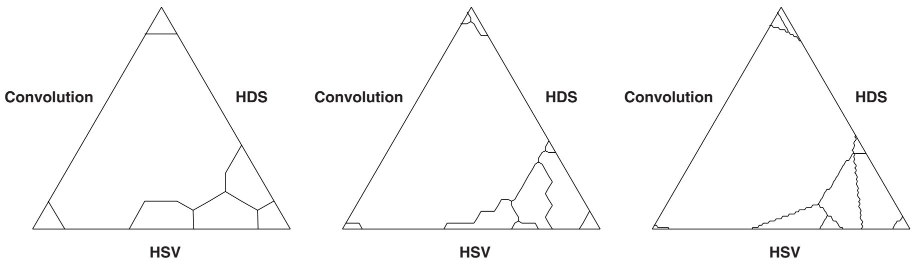


Fig. 2. Partitioning of the weight space for $n = 5, 10, 100$. Each region is associated with a particular image which is one of the NN^k.

3.3 Representative Weights

Let us define the *support* of an NN^k as the relative size of its iso-NNk region. The more feature combinations there are for which the image is ranked top, the greater its support. We may view the support $S(p, q)$ of an NN^k p as a measure of its relevance to the query q.

For the purpose of relevance feedback we wish to associate each NN^k with a characteristic weight set for which the image is closer to the query than is any other image. A natural way to define this is as the k-dimensional centre of mass of the iso-NN^k region. This can be approximated by the arithmetic average of all the weight vectors sampled within that region.

4 NN^k Relevance Feedback

4.1 Weight Estimation

Each NN^k is associated with the weight vector under which it is most similar to the query. It can be viewed as the optimal weight vector for the particular semantic facet represented by that NN^k. We should hope that more relevant images will surface if we retrieve more than just the top-ranked image (the selected NN^k) under that weight vector. The proposed relevance feedback method thus consists of two steps. In the first step, NN^k are computed for a particular query image using the techniques described in Section 3. In the second step, following relevance feedback, images are retrieved under the chosen weight vector. If users select only one image from among the NN^k, the one ranked list induced by the corresponding weight vector constitutes our final ranked list. If more than one image is selected we need to merge the ranked lists obtained from each of the weight vectors. Below we evaluate several different merging techniques.

The following sections investigate the effects of NN^k search on retrieval performance. As is customary for this purpose, we employ manually annotated image collections and automate the feedback process so that an evaluation can be carried out on a large scale.

4.2 Image Collections, Features and Performance Measures

Our evaluation is based on two collections. The first is a subset of the Corel 380,000 Photo Gallery with 31,997 images. The pre-assignment of images to categories by means of their position in the directory hierarchy greatly facilitates automatic evaluation as an image may be considered relevant to a query if it is from the same directory. By not allowing memberships to more than one category, however, the Corel classification fails to encode polysemy. To overcome this limitation we use the textual annotation that accompanies each image to form an alternative classification with overlapping classes. The joint vocabulary of the 31,997 images comprises 20,250 terms of which we only keep those that are associated with at least 20 and at most 100 images. Of the resulting 530 classes we discard those with too little visual coherence.

We also built a second collection of a more diverse kind with a richer and more consistent annotation than Corel. We downloaded 8,202 medium-resolution photographs from the Getty Image Archive (http://www.gettyimages.com) along with the annotations assigned by the Getty staff. The selection of photographs was obtained by submitting the query "photography, image, not composite, not enhancement, not 'studio setting', not people" to the Getty website with the additional search option to exclude illustrations. We thereby hoped to obtain a random selection of photographs that would exclude pictures with no photographic content, digitally composed or enhanced photos and any photos taken in a studio setting. Because the resulting dataset contains pictures from a number of different photo vendors we hope to reduce the chance of unrealistic correlations between images. The Getty collection contains 8,202 images from which classes are constructed as before. The joint vocabulary of 8,551 terms is reduced by retaining only those terms that are associated with at least 20 and at most 100 images. We again discard terms that we consider either too difficult for visual retrieval such as 'freshness' or too easy such as 'blue'. After these pruning steps the vocabulary has shrunk to 100 terms which we treat as class labels.

With images belonging to several classes, the same image may form more than one query. Because of the extensive reduction of the vocabulary, only a subset of the images are annotated with one of the remaining class labels and it is only these that we can employ as queries. For each collection we choose a random subset of 500 for evaluation purposes. A summary of various collection statistics is shown in Table 1.

Table 1. The two collections used for evaluation

	Corel	Getty
Number of images	31,997	8,202
Number of classes	191	100
Class size (avg)	49	44
Class size (range)	20–99	20–99
Avg polysemy	1.1	1.4
Max polysemy	5	7

A standard measure of retrieval performance in information retrieval is mean average precision. For the purpose of evaluating relevance feedback techniques, however, arguments based solely on mean average precision must be taken with caution. For when the set of relevant images is low, even a small shift of relevant images in the ranked list may result in substantial gains in average precision without necessarily leading to an increase in the number of relevant images displayed to the user. In addition to average precision we therefore measure performance in terms of precision at 50, Pr(50). We use eight texture and colour features including Tamura features, Gabor wavelets and local HSV colour histograms. More details of these can be found in [6].

4.3 Reference Performance

Because absolute performance tends to be very sensitive to the choice of features, the particular collection as well as the set of query images, systems are best evaluated in direct comparison with other methods under the same experimental conditions. In this paper, we benchmark NN^k search against three other methods: The first method weighs all features equally and provides us with a baseline. The second method, an oracle, weighs features according to an optimal weight vector that has been determined empirically for each individual query. No system that retrieves with only one weight vector can do better than this. The third method is an implementation of the weight update technique proposed by [12], a relevance feedback technique that aims to optimises a single weight vector. The three methods will be described in turn.

Baseline performance: The baseline performance in our experiments is obtained by assigning every feature the same weight. This is our best guess if we have no information about the relative performance of individual features, and is the initialisation strategy employed in the great majority of relevance feedback methods.

Oracle performance: We can define an *upper* performance bound by empirically determining the best weight vector for each individual query. This is an important benchmark because a large number of relevance feedback techniques are concerned with finding such a single weight vector. For such methods, the oracle provides the best possible performance. Our method could in principle exceed this bound because we may retrieve with as many weight vectors as there are NN^k selected by the user.

Owing to the discrete nature of ranks, an infinitesimal change in the weights may lead to a finite change in performance. As a result the functions mapping weights to average precision and Pr(50) are discontinuous. One way to optimise performance under such conditions is to provide a stochastic formulation of the objective function thereby rendering it differentiable, see for example [4]. Alternatively, we may carry out a grid search of the weight space and improve the grid-based estimate by a subsequent gradient ascent step. This is the method adopted here.

Figure 3 illustrates how retrieval performance varies for a particular query as we change the weight of a single feature. When performance is measured in

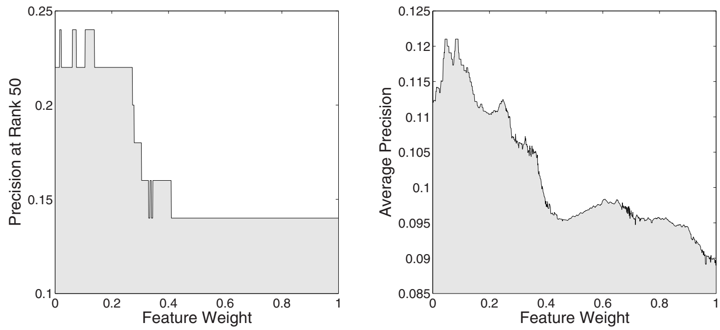


Fig. 3. Variation in retrieval performance as we vary one feature weight

terms of Pr(50) (left graph) the variation is visibly discontinuous. This is much less so for average precision (right graph). Indeed, we should be able to obtain an approximation of the slope of the average precision curve for the purpose of our gradient ascent step. Let j index the feature the weight of which we vary by a small amount h. An estimate of the jth component of the gradient is then given by

$$\frac{P(w) - P(w + \Delta w)}{h}, \tag{2}$$

where Δw is a vector whose jth element is h. Because the weights are sum-normalised, any increase in one feature weight must be accompanied by a corresponding decrease in the sum of all other weights. Strictly speaking, therefore, increasing the jth weight by h requires us to multiply all others by some common factor that ensures that they again sum to 1. We find the other components of Δw to be given by

$$\Delta w_i = \left(\frac{h}{w_j - 1} \right) w_i. \tag{3}$$

After identifying the weight and the direction (the sign of h) along which performance increases most, we perform a search along the line given by the set

$$\left\{ w \;\middle|\; w_i = 1 - \frac{t w_i^0}{1 - w_j^0},\; w_j = w_j^0 + t,\; t \in [-w_j^0, 1 - w_j^0] \right\}, \tag{4}$$

where j indicates the weight we vary, and w^0 denotes the starting weight. The w that maximises performance locally is found by moving along the line in small intervals, Δw, starting at w^0 until performance decreases. In Figure 3 the line search begins at $w_j^0 = 0$ and moves at intervals of 0.01 to the right. The line search terminates at the first peak $(0.05, 0.121)$ which also happens to be the global maximum in that direction. Subsequent iterations find the new direction along which performance can be increased further, and so on until performance decreases in every feature direction. In this case performance can be increased to above 0.14 with this method. Although we are not guaranteed to find the global optimum, the final estimate generally comes very close to it as suggested by grid searches with very high resolution.

Weight update according to Rui: [12] derive an optimal solution for the feature weights w that minimises the summed distances between relevant images and the query with the distance of each relevant image being weighted by its relevance score v_i. Given N positive examples, the optimal w is found to satisfy

$$w_j \propto \left(\sum_{i=1}^{N} v_i d(p_{ij}, q_j) \right)^{-\frac{1}{2}}, \tag{5}$$

where we sum over weighted distances between the query and all relevant images under feature j. In our experiments we choose at random a maximum of ten relevant images and set their relevance scores to one. Note that since the aggregation formula for the feature-specific distances is linear in the feature weights, the proportionality constant in Equation 5 can be chosen arbitrarily. In our application of this method, relevance feedback is initially given on the results of a baseline run. We then allow for another two rounds of relevance feedback before measuring retrieval performance.

5 Results

The aim of the first step of the NN^k relevance feedback method is not to achieve high precision but rather to gather an eclectic set of images that represent a range of semantic facets of the query. Queries for which a uniform parameter initialisation happens to be appropriate are therefore expected to give better results under the baseline run than the first step of our method. Indeed, it turns out that the total number of relevant images returned from the first step of NN^k search is smaller than for a baseline search. However, by increasing the chance of containing at least one relevant image, our method may be expected to outperform the baseline search in the second step, the alternative update method by [12], and possibly even the performance attained by the oracle.

The main results are compiled in Table 2. It displays the absolute performance averaged over 500 queries for the baseline search, the optimised search, Rui's method, and NN^k search with three different merging methods. Round Robin (RR) merges the lists according to the highest rank achieved by an item in either of the lists. BordaFuse (BF) is another rank based method that has the effect of averaging over individual ranks [2]. CombSum (CS) ranks according to the average score.

Performance figures are very similar across collections. This may at first surprise as the Corel collection contains more than three times as many images as the Getty collection with roughly the same number of images per class. It appears, therefore, that the greater variability of images within Getty classes makes up for the smaller size and that it does indeed constitute a more challenging collection for retrieval.

Table 2. Retrieval Performance of NN^k search

	Corel		Getty	
	MAP (%)	Pr(50)(%)	MAP (%)	Pr(50) (%)
Baseline	2.29	1.58	2.95	1.64
Oracle	5.26	2.81	6.41	2.92
RUI	2.49	1.66	3.17	1.71
NN^k (RR)	4.45	2.12	5.76	2.28
NN^k (BF)	3.43	1.83	4.55	2.02
NN^k (CS)	3.79	1.95	4.91	2.14

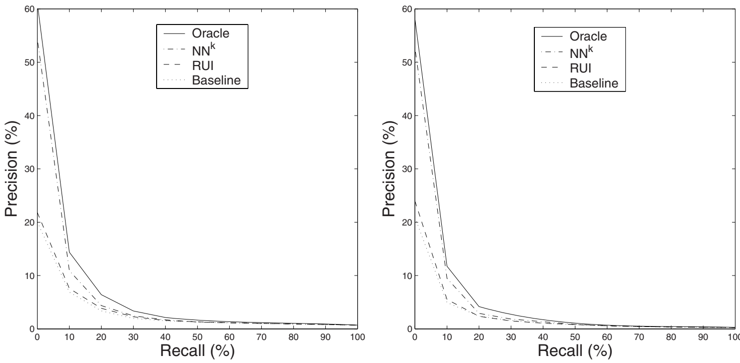


Fig. 4. Precision-recall graphs for 500 queries for Getty (left) and Corel (right)

From among the three merging techniques, the Round Robin method consistently achieves the greatest performance gains. The reason for the superior performance of Round Robin can be seen in the fact that, unlike Borda Fuse and CombSum, the final rank assigned by Round Robin to an object is only very weakly correlated with the average rank it occupies in the different lists to be merged. Averaging ranks or scores, which is the effect of CombSum and Borda Fuse, appears sensible in situations in which we do not want extreme individual ranks to dominate the final ranking. That they are less helpful in the context of image retrieval may result from the observation that relevance classes tend to be composed of several groups of visually similar images with little visual coherence between groups [9]. If the selected NN^k happen to come from different groups, we expect the different weight sets to be particularly good at retrieving more relevant images from the respective groups. Relevant images, therefore, are expected to be ranked high for some weight sets and low for others. Round Robin ensures that such images still receive a high overall rank.

Irrespective of the fusion technique, the performance of our NN^k-based method lies markedly above the baseline and reaches close to the performance achieved under an optimised single weight set (oracle). The differences to the baseline is statistically significant, but so is its difference to the optimal performance (for both we find $p < 0.001$ under a paired t-test).

As further confirmation of the previous observation, we note that our method does considerably better than the update method by [12]. In fact, the weight update according to Equation 5 does not seem to lift performance much above the baseline and, indeed, with $p = 0.066$ the difference in Pr(50) is not significant under a paired t-test. This is perhaps to be expected of semantically biased relevance feedback methods when evaluated on realistic image collections. Visual confirmation of these claims comes from the precision against recall curves obtained by averaging over the 500 queries (Figure 4).

6 Conclusions

We have proposed a new relevance feedback technique for content-based image retrieval that addresses the problem of parameter initialisation and weight optimisation for image retrieval. We propose a solution in which parameters are not initialised at all for the first retrieval step. Instead of retrieving with some fixed parameter setting, we compute the images that are most similar to the query under *at least one* parameter setting. These images represent pictorially the range of semantic facets users may have had in mind when posing the query. In addition to covering a wider range of possible image interpretations, and thus being more likely to capture the semantic facet that matters to the user, this first step makes it possible to associate each of the images with a characteristic weight set, which is the average weight set under which that image is closest to the query. By selecting relevant images from among the retrieved set, users thus implicitly select weight sets which can subsequently be used for retrieval. The method has been tested on two image collections for both of which retrieval performance gets close to the optimum performance and consistently outperforms an alternative relevance feedback method.

References

[1] Aggarwal, G., Ashwin, T., Ghosal, S.: An image retrieval system with automatic query modification. IEEE Trans Multimedia, 4(2), 201–213 (2002)

[2] Aslam, J., Montague, M.: Models for metasearch. In: Proc. Int'l ACM SIGIR, pp. 276–284 (2001)

[3] Ghahramani, Z., Heller, K.: Bayesian sets. In: Proc. NIPS (2005)

[4] Goldberger, J., Roweis, S., Hinton, G., Salakhutdinov, R.: Neighbourhood component analysis. In: Proc. NIPS (2005)

[5] Gosselin, P., Cord, M.: Active learning methods for interactive image retrieval. IEEE Trans. Image Processing 17(7), 1200–1211 (2008)

[6] Heesch, D.: The NN^k technique for image searching and browsing. PhD thesis, Imperial College London (2005)

[7] Heesch, D., Rüger, S.: Interaction models and relevance feedback in content-based image retrieval. In: Zhang, Y.-J. (ed.) Semantic-Based Visual Information Retrieval, pp. 160–186. Idea-Group (2006)

[8] Heller, K., Ghahramani, Z.: A simple Bayesian framework for content-based image retrieval. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, vol. 2, pp. 2110–2117 (2006)

[9] Kim, D.-H., Chung, C.-W.: Qcluster: relevance feedback using adaptive clustering for content-based image retrieval. In: Proc. ACM SIGMOD Int'l Conf. Management of Data, pp. 599–610 (2003)

[10] Nguyen, G., Worring, M., Smeulders, A.: Interactive search by direct manipulation of dissimilarity space. IEEE Trans. Multimedia 9(7), 1404–1415 (2007)

[11] Qin, T., Zhang, X.-D., Liu, T.-Y., Wang, D.-S., Ma, W.-Y., Zhang, H.-J.: An active feedback framework for image retrieval. Pattern Recognition Letters 29, 637–646 (2008)

[12] Rui, Y., Huang, T.: Optimizing learning in image retrieval. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 236–243 (2000)

[13] Rui, Y., Huang, T., Ortega, M., Mehrotra, S.: Relevance feedback: A power tool for interactive content-based image retrieval. IEEE Trans. Circuits, Systems and Video Technology 8(5), 644–655 (1998)

[14] Sclaroff, S., Taycher, L., La Cascia, M.: ImageRover: A content-based image browser for the WWW. Technical report, Boston University (1997)

[15] Urban, J., Jose, J.: Evidence combination for multi-point query learning in content-based image retrieval. In: Proc. IEEE Int'l Symposium Multimedia Software Engineering, pp. 583–586 (2004)

An Architectural Paradigm for Collaborative Semantic Indexing of Multimedia Data Objects

Clement H.C. Leung[1], Jiming Liu[1], Alice W.S. Chan[1], and Alfredo Milani[2]

[1] Department of Computer Science,
Hong Kong Baptist University,
Kowloon Tong, Hong Kong
{clement,jiming,milani,wschan}@comp.hkbu.edu.hk
[2] Department of Mathematics & Computer Science,
University of Perugia,
Perugia, Italy
milani@unipg.it

Abstract. A challenging problem facing the semantic search of multimedia data objects is the ability to index them. Here, we present an architectural paradigm for collaborative semantic indexing, which makes use of a dynamic evolutionary approach. By capturing, analyzing and interpreting user response and query behavior, the patterns of searching and finding multimedia data objects may be established. Within the present architectural paradigm, the semantic index may be dynamically constructed, validated, and built-up, where the performance of the system will increase as time progresses. Our system also incorporates a high degree of robustness and fault-tolerance whereby inappropriate index terms will be gradually eliminated from the index, while appropriate ones will be reinforced. We also incorporate genetic variations into the design to allow objects which may otherwise be hidden to be discovered. Experimental results indicate that the present approach is able to confer significant performance benefits in the semantic searching and discovery of a wide variety of multimedia data objects.

Keywords: Collaborative indexing, Concept-based Search, Genetic Algorithms, Index Set, Multimedia Data Objects, Ranking, Relevance Feedback, Semantics Retrieval.

1 Introduction

As the creation of multimedia information rapidly expands, their search becomes an important yet less than effective activity [6], [7]. Most of the today's search engines apply the PageRank algorithm for ranking query results based on scoring and the link structure of the Web [3], [4], [9], [10]. By applying the PageRank algorithm, the relative importance of hyperlinked set of documents can be measured. However, multimedia information search is far more difficult than searching text-based documents since the content of text-based documents can be extracted automatically while the content of multimedia objects cannot be automatically determined.

M. Sebillo, G. Vitiello, and G. Schaefer (Eds.): VISUAL 2008, LNCS 5188, pp. 216–226, 2008.

Many different methods and techniques have been proposed for retrieving images. They are mainly classified into two main categories: "concept-based" image retrieval, and "content-based" image retrieval [1], [2], [13], [14], [17]. The former focuses on using words to retrieve images (e.g. title, keywords, and caption), while the latter focuses on the visual features of the image (e.g. size, colours, and textures). In an effective "concept-based" multimedia retrieval system, efficient and meaningful indexing is necessary [6], [8]. Due to the current technological limitations, it is impossible to extract the semantic content of multimedia data objects automatically [16], [15], [19]. Meanwhile, the discovery and insertion of new indexing terms are always costly and time-consuming. Therefore, novel indexing mechanisms are required to support their search and retrieval.

With the development of Web 2.0, user-generated contents and users' expert knowledge are extensively exploited [20]. In our method, we adopt the spirit of Web 2.0 and develop a collaborative indexing architectural paradigm. Through the continuous and extensive application of multimedia search within this architecture, users will be able to provide meaningful indexing and visual judgment for the multimedia data objects. Consequently, these objects would be optimally indexed such that semantic search on them would become possible.

2 The Collaborative Indexing Approach

In our approach, we focus on the indexing of semantic contents of multimedia objects and exclude metadata [13], since the indexing of metadata is relatively straightforward and less meaningful than semantic contents as perceived by humans. For example, indexing and retrieving a song by its characteristics (e.g. style) is more challenging than indexing its metadata (e.g. artist).

We consider a set of multimedia data objects $\{O_j\}$ such as images, video, or music where their semantic characteristics and contents cannot be extracted automatically. Every O_j links with an index set I_j, which consists of a number of elements e_{j1}, e_{j2}, …, e_{jM}. Each index element e is a triple, which is composed of an index term ID t_{jk}, an index score s_{jk} associated with t_{jk}, and the object ID o_j. The higher the index score, the more important is the index term to the object. Conversely, the lower the index score, the less important is the index term to the object. The relationship of t_{jk}, s_{jk}, and O_j can be represented in the following relations, where the primary keys are underlined.

IndexTerm (index_id, index_term)
MultimediaObject (object_id, object_name, object_description)
IndexTable (index_id, object_id, score)

In particular, the index table links the index terms and multimedia objects in a many-to-many (M:M) relationship.

The index hierarchy refers to the index sets of all the objects stored in the database. By partitioning the value of score s_{jk}, it can be divided into N levels $L_1, L_2, \ldots, L_N$ with a set of parameters $P_1, P_2, \ldots, P_N$. By considering the score value x of a given index term, the index term would be placed in level L_i when the value of x is inclusively lower bounded with a parameter value P_i and exclusively upper bounded with another parameter value P_{i+1}. All values that are larger than or equal to a parameter P_N would

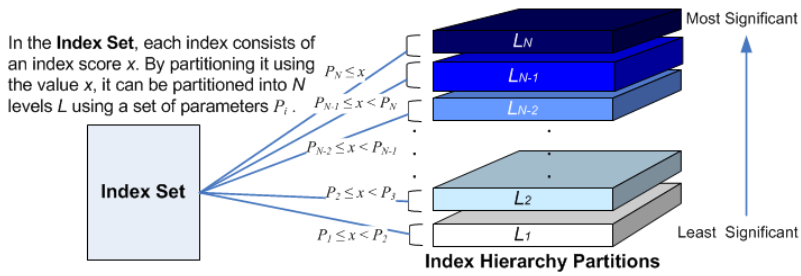


Fig. 1. Partitioning of the Index Set

be placed at the top level of the hierarchy. Fig. 1 shows the structure for the index set partitioning.

In this index hierarchy, the higher the level, the more significant it is. Hence, multimedia data would be searched from the top level first. In some cases, we may disregard the lower levels. When an object *O* is minimally indexed, it means *O* has only a single index term *T*, where *T* is a single word. The multimedia objects should initially be minimally indexed in order to be searched, since the multimedia objects are located by search term(s) from a search query input by user.

As our approach is concerned with indexing rather than tagging, it is useful to draw a distinction between them. Indexing is widely applied in today's search engines, while social tagging is also commonly used in Web 2.0 searching websites, such as Flickr, Gmail, YouTube, etc. An index is a systematic hierarchy that enables fast search and accurate data retrieval, while a tag is a term or keyword associated with a piece of information, such as a document, a picture, a video, etc. Although indexing and tagging both allow users to search information by "words" in queries, the content of the "words" may be different. In the indexing approach, the index term can be related keywords, a term, or a phrase that can have semantic meanings, while stop-words, such as a, an, the, etc., would not be indexed [11], [12]. In the tagging approach, most of the tags tend to be isolated terms without semantic interconnections. By comparing their structure, indexing maintains a more well-defined hierarchy levels and structure while tagging has no hierarchy and tags are unstructured and unconnected [18]. In relation to their installation mechanism, tagging is generally performed manually while indexing can be done automatically as well as input manually.

3 Index Score Updating Algorithms

Stored procedures in the query level are responsible for all database related processes which include multimedia object retrieval, search result ranking, index score updating, index insertion, etc.

When user input a series of search terms T_1, T_2, …, T_n in a search query $Q(T_1, T_2, \ldots, T_n)$, the query score $S(Q|O_j)$ for an object O_j can be extracted by:

SELECT score
FROM IndexTable
WHERE object_id = $O_{j\text{-}ID}$ and index_id = $T_{i\text{-}ID}$;

This score implies the relative importance of an index term T_i to the corresponding object O_j. Thus, the multimedia objects in the query result should be ordered by score in a descending order. The query result can be obtained by:

```
SELECT object_id, SUM(score) AS s
FROM IndexTable
WHERE index_id in (T1-ID, T2-ID, …, Tn-ID)
GROUP BY object_id ORDER BY s DESC
```

The index score is directly affected by user search behavior, such as result selection and relevance feedback. By the continuous use of the search system, user search behavior can be collected and analyzed. Considering an example of a user input search query $Q(T_1, T_2)$, suppose N multimedia objects $O_1, O_2,\ldots,O_n$ are returned in the query result and ordered by the corresponding score $S_1, S_2, \ldots, S_n$ in descending order. The related index scores on T_1 and T_2 for the desired object O_x would be increased when the user selects O_x in the query result list, or when the user provides positive feedback on O_x. These two cases would increase the related index scores on T_1 and T_2 for the desired object O_x by a predefined valueΔ_x, where the predefined value for these two cases can be different.

In contrast, the related index scores on T_1 and T_2 for the desired object O_x would be decreased when the user provides negative feedback on O_x. Furthermore, the related index scores on T_1 and T_2 for all objects $O_1, O_2,\ldots,O_n$ in the query result would be decreased when the user does not select any object on the query result list. These two cases would decrease the related index scores by a predefined valueΔ_y, where the predefined value for these two cases can again be different.

4 Installing New Index Terms

In the index growth approach, consider an object K that is minimally indexed with a term T_1. K can be searched by a user query which contains T_1 and many objects may be returned in the query result since many objects are indexed with T_1. Among these returned objects, a user can distinguish objects by adding another index term T_2 to K. Thus, the user can search the desired object by entering both index terms in the search query [13]. Consider the searching of a song “Für Elise”, which is a piece of music composed by Beethoven, and we assume that the audio object is minimally indexed with the term “Beethoven” initially. Users can search this song by the term “Beethoven”. Sometimes, some user query would be more specific, with both search terms “Beethoven” and “Für Elise” used. The same multimedia object would be returned in the result when searching by the term “Beethoven”, since the term “Für Elise” is not indexed yet. Eventually, the user would select the audio object “Für Elise” and this suggests that a new index term, “Für Elise”, may be useful for indexing this data object. Thus, the new index term would be included in the lowest level of the index hierarchy for this audio object. For every query that specifies both terms, “Beethoven” and “Für Elise”, the user on selecting this audio object will cause an increase in the score of the index terms for that object. Thus, the score of the index would be gradually increased and the new index would be properly installed. Through progressive usage, the indexing of multimedia objects would be enriched (Fig. 2).

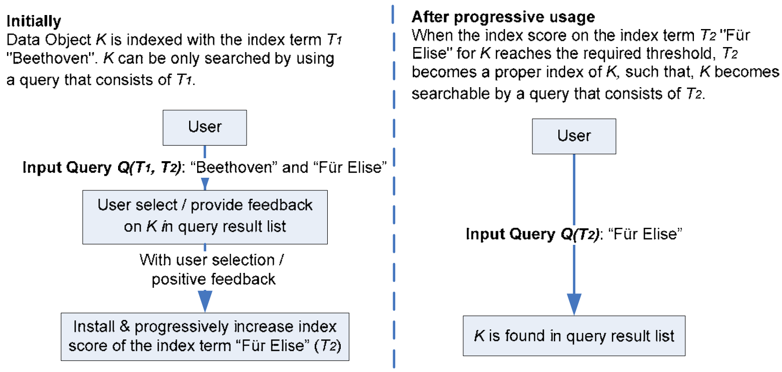


Fig. 2. Installing a New Index Term for a Multimedia Data Object

5 Introducing Genetic Variations for Object Discovery

After considerable usage, the high scored multimedia objects always rank near the top of the search results. Generally, users tend to be interested in the highly ranked objects. Consequently, the high scored objects always have greater chance of increasing the score while the chance for the newly added objects appearing in the result list would be reduced. Thus, the new added objects would be ranked very low and nearly "hidden".

In order to optimize the search results, our search system would introduce small degrees of perturbations so as to allow constructive variations in the result. When we consider a large number of multimedia objects returned by a query result, the users may not reach their desired object since the target object always rank very low and nearly hidden. By exploiting the random variations of Genetic Algorithms (GA), those "hidden" objects would have a chance of being promoted to a higher ranking position and discovered eventually.

We apply the GA in discovering those "hidden" objects. With GA, the object ranking for each object O_j that with corresponding index score S_j in the query result can be made to be determined by a probability:

$$P_j = f(S_j)\,, \qquad j = 1, 2, \ldots, M,$$

where $f(.)$ is a monotonically increasing function, and M is the number of objects in the search result, so that the higher the score, the greater is this probability. A particularly simple form for this function is

$$P_j = \frac{S_j}{\sum_{i=1}^{n} S_i}\,, \text{ where } j = 1, 2, \ldots, M. \tag{1}$$

Since this probability is directly proportional to the index score S_j, the object O_j with the higher index score S_j would have a higher probability value P_i. Consequently, the object with the higher probability value would have greater chance of getting a higher rank in the query result list.

Table 1. Unsorted Query Result List

Object ID (O_j)	**Query Score (S_j)**	**Probability value (P_j)**
O_1	S_1	P_1
O_2	S_2	P_2
O_3	S_3	P_3
…	…	…
O_M	S_M	P_M

Table 2. Sorted Query Result List (without GA)

Object ID (O_j)	**Query Score (S_j)**
O_1	S_1
O_2	S_2
O_3	S_3
…	…
O_M	S_M

Table 3. Query Result List (with GA)

Object ID (O_j)	**Query Score (S_j)**
O_1'	S_1'
O_2'	S_2'
O_3'	S_3'
…	…
O_M'	S_M'

More precisely, we assume there are M expected objects that fulfill a search query (Table 1). Moreover, the probability value of each object O_j can be determined by its corresponding query score S_j by equation (1). Suppose $S_1 > S_2 > \ldots > S_M$, then we can deduce that $P_1 > P_2 > \ldots > P_M$. Without GA, objects are strictly ranked by the query score S_j in a descending order (Table 2). After applying GA, an object with a low query score S_j would also have a chance to be ranked at the top of the query result list since the object ranking is determined by the randomness of GA. Table 3 shows a sample query result that applied GA. In this result, it is not always true that $S_1' > S_2' > \ldots > S_M'$, and it is possible to have an object O_k' ranked higher than object with O_m', with $S_k' < S_m'$, where $m > k$.

6 Architecture Overview

Our proposed architecture consists of different levels for supporting the multimedia search. It includes user level, interface level, and database level. The database level contains two components, query component and index component. The overview of the architectural structure is shown in Fig. 3.

In the architectural model, the users interact with the system via the interface. They can search multimedia objects by submitting query with query terms. After the initial search process is completed, they can provide feedback to the system. The interface (see Appendix I) acts as a bridge between user and the database. It captures query input, search result feedback and selection from the user.

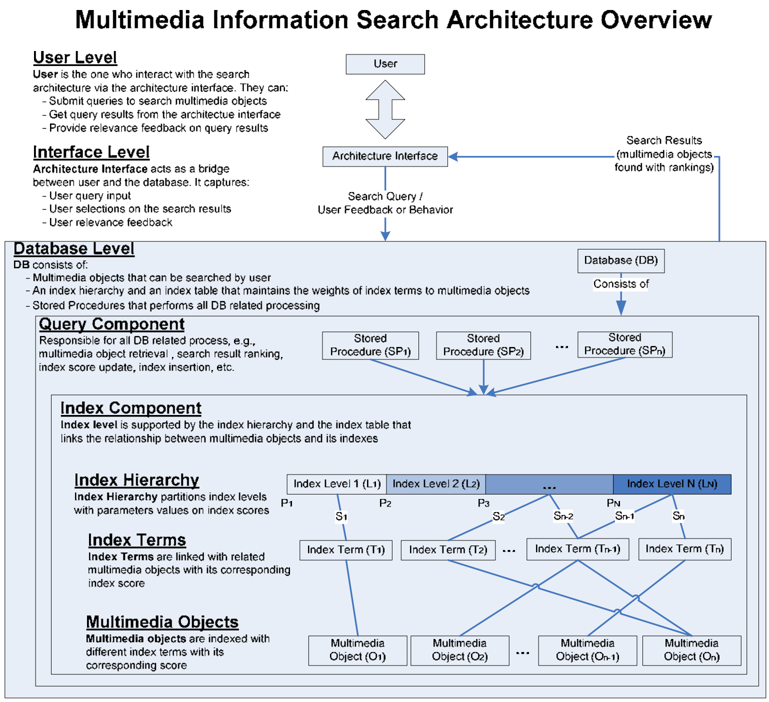


Fig. 3. Multimedia Information Search Architecture

The database component is the core of this architecture. It consists of a query component and an index component. The query component consists of the following subsystems:

1. Index insertion subsystem
2. Index score updating subsystem
3. Multimedia object ranking and retrieval subsystem

The index level consists of an index table, with an index hierarchy, which links multimedia objects and index terms with the corresponding index score, which indicates the importance of the index term that relates to the data object.

7 Experimental Results

We have carried out experiments to measure the effectiveness of our approach.

Index Growth
Initially, a hundred multimedia objects are stored in the database. Each object is initially indexed with one index term. The index table is grown through sixty queries, input by the user. Throughout the queries, the user add new index term(s) to the query results. Fig. 4 shows the relationship between the number of queries and the number

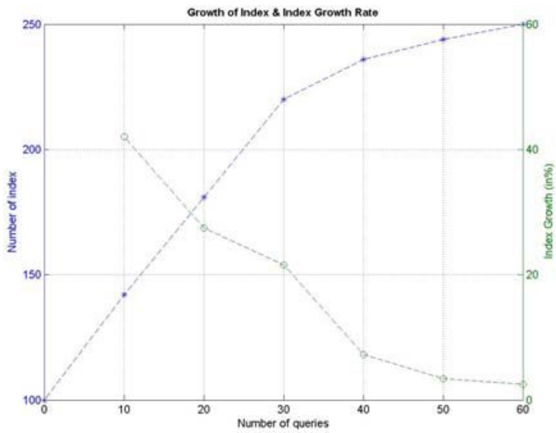


Fig. 4. Index Growth

of index terms (on the left-y-axis). As the number of queries increases, the number of index terms also increases as expected. At the initial state, the number of index terms grows faster since there is more scope for users to add new ones. The number of index terms keeps on increasing while the scope for user to add new ones is reduced. Thus, it grows slower and the number of index terms tends towards an equilibrium level. In addition, Fig. 4 shows the index growth rate (on the right-y-axis). We observe that the index growth rate is the highest at the first ten queries, and then, it drops rapidly. As the number of queries increases, it drops continuously and more gradually. The drop slows down until it reaches an equilibrium level.

The Effect of Indexing

In this experiment, a total of thirty queries are used. These queries are performed before indexing with fifteen queries performed before usage, and an identical set of queries performed after the usage. Initially, all the multimedia objects are minimally indexed with one index term. As users interact with the system, the index table grows. The multimedia data objects become not only searchable with its initial index term, it is also searchable by other index term(s) that are added by users. The new index terms that are meaningful will gradually be promoted to the most significant level of the index hierarchy.

Fig. 5 shows the recall rate. In the first 5 queries of our standard query set, we use the initial index terms for searching the desired multimedia object. From the experimental results of these queries, all the targeted multimedia objects can all be found. By comparing performance before and after indexing through user queries, the targeted objects are ranked significantly higher (Fig. 6). Then, we use new index terms as the input queries for the remaining queries in the query set. Although none of these can be found with the original index terms, they can be found after collaborative indexing through sixty queries. Assuming a cutoff rank of ten, the improvement in search result rankings after indexing is shown in Fig. 7.

We thus see that the effect of collaborative indexing is quite pronounced and it consistently produces significant improvement in search performance. While it takes

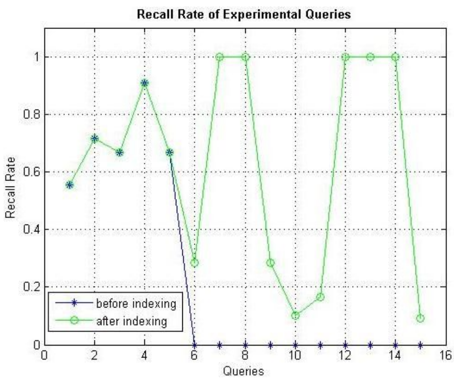


Fig. 5. Recall Rate of the Experimental Queries

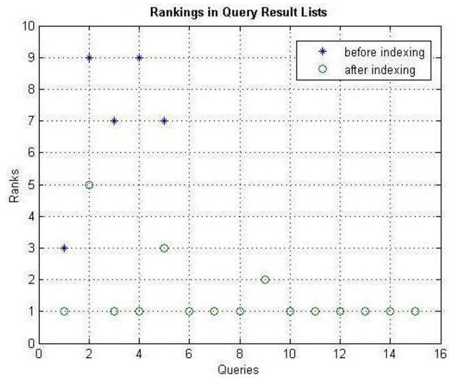


Fig. 6. Rankings in Query Result Lists

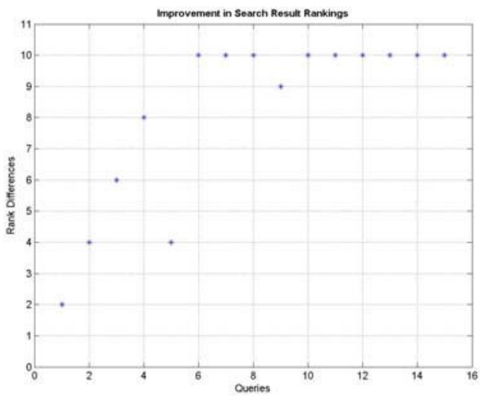


Fig. 7. Improvement in Search Results Based on Cutoff at 10

time for the index to evolve to an optimal performance level, search benefits will start to accrue as soon as the index starts to grow, and search efficiency tends to grow fastest at the initial stages of usage.

8 Conclusions

Dedicated intensive manual indexing of semantic contents tends to be costly, laborious and time-consuming. Even when one is willing to accept the high cost, the relatively slow speed of such a process is unlikely to be able to keep up with the explosive rate by which new multimedia data objects are created – it is estimated that by 2010, over a billion digital images will be created each day [5], not counting other forms of multimedia data objects. Such a mismatch of creation and indexing rates necessitates a radically new approach to solve the multimedia data object retrieval problem.

An architectural paradigm for collaborative semantic indexing has been presented, which makes use of a dynamic evolutionary approach. The participative element together with collaborative judgement based on human intelligence further demonstrates the importance of collective web intelligence within a Web 2.0 framework. By capturing, analyzing and interpreting user response and query behavior, the patterns of searching and finding multimedia data objects may be established. Within the present architectural paradigm, the semantic index may be dynamically constructed, validated, and built-up, and the performance of the system will tend to increase as time progresses. Our system also incorporates a high degree of robustness and fault-tolerance whereby inappropriate index terms will be gradually eliminated from the index, while appropriate ones will be reinforced. We also incorporate genetic variations into the design to allow objects which may otherwise be hidden to be discovered. Experimental results indicate that the present approach is able to confer significant performance benefits in the semantic searching and discovery of a wide variety of multimedia data objects.

References

[1] Azzam, I., Leung, C. H. C., Horwood, J.: A fuzzy expert system for concept-based image indexing and retrieval. In: Proc. of the IEEE International Conf. on Multi-media Modeling, Melbourne, Australia, pp. 452–457 (2005)

[2] Azzam, I., Leung, C. H. C., Horwood, J.: Implicit concept-based image indexing and retrieval. In: Proc. of the IEEE Int'l Conf. on Multi-media Modeling, Brisbane, Australia, pp. 354–359 (2004)

[3] Diligenti, M., Gori, M., and Maggini, M.: Web page scoring systems for horizontal and vertical search. In: Proc. of the 11th Int'l World Wide Web Conf (WWW 2002), Hawaii, USA, pp. 508–516 (2002)

[4] Dwork, C., Kumar, R., Naor, M., and Sivakumar, D.: Rank Aggregation Methods for the Web. In: Proc. of the 10th Int'l World Wide Web Conf., Hong Kong, pp. 613–622 (2001)

[5] Gantz, J.F., et. al.: The expanding digital universe: a forecast of worldwide information growth through 2010, IDC White Paper (March 2007)

[6] Gomez, J., Vicedo, J.L.: Next-Generation Multimedia Database Retrieval. IEEE Multimedia 14(3), 106–107 (2007)

[7] Goth, G.: Multimedia search: ready or not? IEEE Distributed Systems Online 5(7) (2004)

[8] Gros, P., Delakis, M., and Gravier, G.: Multimedia Indexing: The Multimedia Challenge. In: ERCIM News No. 62 (2005)

[9] Haveliwala, T.H.: Topic-Sensitive PageRank. In: Proc. of the 11th Int'l World Wide Web Conf., Hawaii, USA (2002)

[10] Haveliwala, T.H.: Topic-Sensitive PageRank: A Context-Sensitive Ranking Algorithm for Web Search. IEEE Transactions on Knowledge and Data Engineering 15(4), 784–796 (2003)

[11] Hoppenbrouwers, J.: Social Tagging/indexing. In: Workshop at the 31st ELAG Conf., Barcelona (2007)

[12] Kipp, M.E.I., Campbell, D.: Patterns and Inconsistencies in Collaborative Tagging Systems: An Examination of Tagging Practices. In: Proc. of the 2006 Annual Meeting of the American Society for Information Science and Technology, Austin, Texas (2006)

[13] Leung, C., Liu, J.: Multimedia data mining and searching through dynamic index evolution. In: Qiu, G., Leung, C., Xue, X.-Y., Laurini, R. (eds.) VISUAL 2007. LNCS, vol. 4781, pp. 298–309. Springer, Heidelberg (2007)

[14] Over, P., Leung, C.H.C., Ip, H., Grubinger, M.: Multimedia retrieval benchmarks. IEEE Multimedia 11(2), 80–84 (2004)

[15] Rafferty, P., Hidderley, R.: Indexing Multimedia and Creative Works: The Problems of Meaning and Interpretation, Aldershot: Ashgate, pp. 43–48 (2005)

[16] Snoek, C.G.M., Worring, M., Gemert, J.C., Van, G.J.M., Smeulders, A.W.M.: The Challenge Problem for Automated Detection of 101 Semantic Concepts in Multimedia. In: Proc. of the 14th annual ACM Int'l Conf. on Multimedia, CA, USA, pp. 421–430 (2006)

[17] Tam, A., Leung, C.H.C.: Structured natural-language descriptions for semantic content retrieval of visual materials. Journal of the American Society for Information Science and Technology, pp. 930-937, 2001.

[18] Voss, J.: Collaborative thesaurus tagging the Wikipedia way. Wikimetrics research papers 1(1) (2006)

[19] Yang, B., Hurson, A.R.: Ad hoc image retrieval using hierarchical semantic-based index. In: Proc of the 19th Int'l Conf. on Advance Information Networking and Applications (AINAW 2005), pp.629–634 (2005)

[20] Yesilada, Y., Harper, S.: Web 2.0 and the Semantic Web: Hindrance or Opportunity? In: Proc. of WA4 International Cross-Disciplinary Conf. on Web Accessibility 2007, (90), pp. 19–31 (2008)

Appendix I

A System Interface

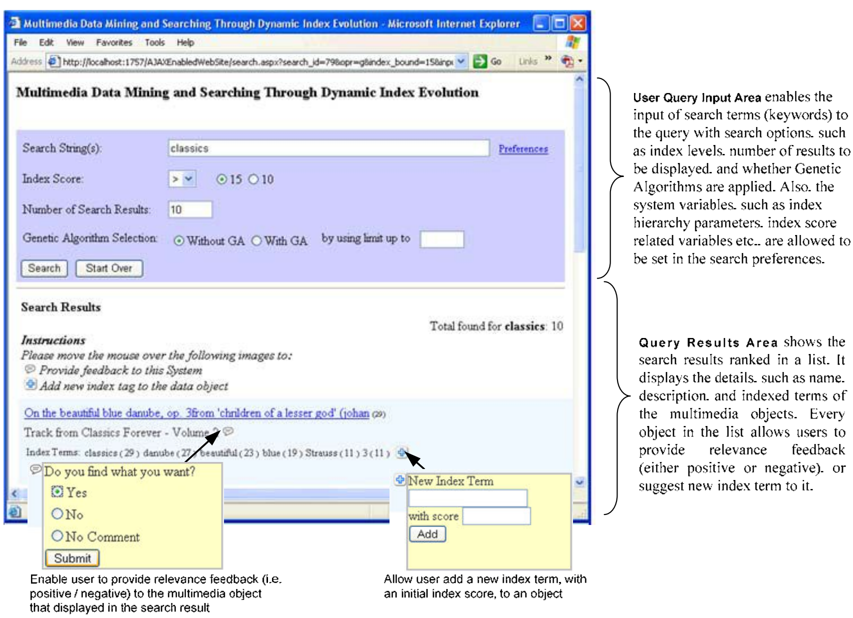


Fig. 8. System Interface

Evaluation of Quality Retaining Diagnostic Credibility for Surgery Video Recordings

Mariusz Duplaga[1], Mikołaj Leszczuk[2], Zdzisław Papir[2], and Artur Przelaskowski[3]

[1] Jagiellonian University, Collegium Medicum
[2] AGH University of Science and Technology, Department of Telecommunications
[3] Warsaw University of Technology, Institute of Radioelectronics

Abstract. The paper proves that it is possible to meet contradictory requirements of effective compression and diagnostic credibility, for long-lasting and low motion bronchoscopic recordings at compression ratios close to 100. As the main supporting assumption, it has been accepted that the content can be compressed as far as clinicians are not able to sense a loss of video diagnostic fidelity. Different market codecs were inspected by means of the combined subjective and objective tests.

Keywords: video, video compression, quality evaluation, bronchoscopy.

1 Introduction

Development efforts reveal a significant potential behind platforms allowing to access digital surgery video recordings. Video compression problems are the mostly widespread ones among those pertinent to such platforms. With no doubt medical application adds a new dimension as lossy compression techniques to be used therein have to be both resource-effective and medically credible.

The presented paper targets video-bronchoscopy, being a subclass of surgery recordings. In such moving images, the image can sustain almost motionless for quite long periods of time. This predisposes such recordings to compression.

Usage of lossless codecs results in low Compression Ratios (CRs) currently achieved when compressing and storing video sequences. Therefore these codecs implicate waste requirements on both an archive memory and a throughput of a streaming network. On the other hand, lossy codecs generally produce desirable narrow-band video streams. Unfortunately in the case of using visually impaired images for diagnostic purposes, a serious danger of an impermissible influence on the diagnosis does exist. In conclusion, are we doomed to a contradictory choice between either narrow-band streams without a diagnostic potential of voluminous file storage or transfers being diagnostically credible?

The approach states that it is possible to use images in which lossy compression has not caused any distortion visible to a panel of physicians. Therefore, it seems reasonable to assume, according to [1], that **it is possible to use lossy compressed images and video sequences for diagnostic purposes provided the compression has not introduced any quality impairment visible to a panel of physicians**.

M. Sebillo, G. Vitiello, and G. Schaefer (Eds.): VISUAL 2008, LNCS 5188, pp. 227–230, 2008.

2 Subjective and Objective Evaluation of Lossy Codecs

Premises revealed in the previous section determine the maximum CR value for different available types of codecs allowing for diagnostically credible compression. The experimental approach has been defined as follows:

1. To specify the CR_{max} (still yielding in a visually lossless compression) for some selected, reference video codec using subjective evaluations supported by a physician panel.
2. To specify numerically the corresponding objective distortion metrics for the same video sequence compressed with the reference codec at the CR_{max}.
3. To estimate CR_{max} values for other codecs under consideration at the same objective distortion values as for the reference codec.

In order to determine the CR_{max} value of a codec three recordings of bronchoscopic examinations have been used. Each video sequence contained a recording with a different disease type.

2.1 Subjective Evaluation of a Reference Codec

The tested [2] subjective method for qualifying lossy compressed still images to the subset of visually undistorted ones is based on ordering compressed images by their quality. The same approach has been adopted in the undergoing investigation of video sequences. A pulmonologist is randomly presented with several video sequences — the original one and seven others compressed with various CR values. As the result of ordering, it is usually possible for an experiment supervisor to distinguish two subsets of video sequences [2]. The first one consists of the highest quality video sequences in the random order. The rest of the video sequences appear in the second lowest quality subset. Only the video sequences belonging to the first subset are considered to have the quality which makes them suitable for diagnostic purposes.

As the first step, a subjective evaluation of the video sequences compressed at various CRs has been performed. The MPEG-4 codec has been selected as the reference codec as it is a modern, open, and still widely used solution. The codec has been successfully applied for surgery video compression as well [1]. In order to obtain CR_{max}^{MPEG-4}, for which a clinician cannot distinguish between the original video sequence and the compressed video sequence, the test based on the above-mentioned quality-based ordering method has been executed. Each of the three original (uncompressed) video sequences has been complemented with a few compressed video sequences having the same content, thus constituting three investigated video sequences. The compression has been done with the CR equal to approximately 16, 32, 64, 96, 128, 256 and 512.

Eight clinicians have been asked independently to order the video sequences in each of the sets following the well-known bubble sort algorithm. The only information gathered was order of the video sequences. For the purpose of ordering, the specialists have used a software developed specially for this purpose The software has been run at a regular personal computer located at the clinic. No time restrictions for evaluations have been applied.

2.2 Objective Evaluation of a Reference Codec

Most objective methods are designed to evaluate still images. In order to use them for moving pictures results are commonly averaged over all frames of a given video sequence. Such an approach does not touch object motion smoothness. This is a weak restriction as investigated recordings are rather motionless.

Two objective metrics have been selected for the experiments as being prospectively suitable for bronchoscopic recordings. The metrics were the Hybrid Vector Measure (HVM) [2] and the graphical Hosaka plots [3]. The HVM and Hosaka plots are universal metrics since one can deploy them to compare images compressed using various codecs [2,3]. The HVM metric is optimized with a subjective test and its strong correlation with the diagnostic value of medical images has been proved [2]. The HVM approach produces a vector $\overrightarrow{HVM}[V_1, \ldots, V_6]$; each of its six coefficients measures some kind of an image depreciation. The Hosaka metric has been chosen for comparison purposes as a case study of a non-medical metric that is not optimized for medical images. The Hosaka approach delivers a polygon P_H whose area is a measure of picture impairment. Both metrics have been thoroughly presented in [2] and [3].

Once the feasible reference CR_{max}^{MPEG-4} has been obtained objective tests under the HVM [2] and the Hosaka approach [3] have been executed mapping $\overrightarrow{HVM}$ and P_H metrics to feasible CR_{max}^{MPEG-4}. The results of both evaluation approaches are presented in Section 3.

2.3 Objective Evaluation of Other Codecs

For the last stage of the experiment, six popular, modern video codecs (MJPEG2000, MPEG-2, H.263+, H.264, WMV9, RV10) have been chosen. Both lossy and lossless (if available) compression modes have been selected.

After specifying CR_{max}^{MPEG-4} and corresponding HVM and Hosaka metrics, objective tests for other selected codecs were repeated. In order to determine for each codec its specific feasible CR_{max} a compression ratio was changed as long as $\overrightarrow{HVM}$ (all coefficients) and P_H metrics became equal to the ones obtained for the MPEG-4 codec.

3 Results of Codec Evaluation

For the most of video sequences and physicians, a clear-cut partition on a subset of video sequences visually lossless compressed and a subset of lower quality ones have been identified. It has been identified and decided to set $CR_{max}^{MPEG-4} \approx 96$ as the obtained result in subjective evaluations. The $CR_{max}^{MPEG-4} \approx 96$ threshold is associated with corresponding $\overrightarrow{HVM}$ and P_H metrics.

For objective tests, the simplest task is to compare lossless compression standards as they do not introduce any distortions in compressed video sequences. Therefore, the only comparison criterion is the CR, usually achieving low values. Definitely, the best results within the family of lossless standards have been achieved using the H.264 lossless standard ($CR = 7$).

The MJPEG2000 codec was the single compression standard without an inter-frame compression. In the lossless version compression is running only with a low $CR = 2$; however, in the lossy version it is possible to achieve $CR = 21$ while preserving visually lossless compression. Nevertheless, it is still a low CR if compared to the one achieved for the MPEG-4 standard compression ($CR \approx 96$).

Compression by the MPEG-2 codec with $CR \approx 96$ allowed a relatively low amount of error level, except the maximum pixel error — V_2 and the integral square with frequency weighting defined by CCIR — V_5. Compression by the H.263+ coded achieved similar results to MPEG-4, except for errors correlated in the 5×5 window, being relatively large for the H.263+ codec. For the H.264 codec it is possible to compress with the $CR \approx 100$ keeping all $\overrightarrow{HVM}$ coefficients below their thresholds. Unfortunately, for the WMV9 codec, compression even for $CR \approx 96$ makes V_3 and V_4 coefficients exceed their thresholds.

The comparison under the Hosaka metric gave surprising results that, for some codecs compression with even the very high CR could allow for visually lossless compression. On the other hand, the subjective, ex-post-visual analysis of these video sequences performed by the authors contradicts the results achieved under the Hosaka metric, which probably excludes it from video evaluation.

4 General Conclusions

It has been proved that it is possible to effectively compress long-lasting and slow motion video sequences of surgery procedures while preserving their diagnostic features. The presented research concern bronchoscopic procedures. However, the proposed approach to combine subjective and objective evaluations could be applied for most other long-lasting surgery video recordings featuring a low level of motion. Considering several similarities between bronchoscopic recordings and other endoscopic recordings it is likely that the promising results are to be extended into these areas as well.

Acknowledgments. The work was supported by the Ministry of Science and Higher Education (Grants: 3T11D01727, N517438833). The authors thank the panel of physicians from Jagiellonian University for the evaluation efforts.

References

1. Dafonte, C., Gomez, A., Castro, A., Arcay, B.: Scientific papers: a proposal for improving icu assistance through intelligent monitoring and supervision. Technol. Health Care (2002)
2. Przelaskowski, A.: Vector quality measure of lossy compressed medical images. Computers in Biology and Medicine (2004)
3. Hosaka, K.: A new picture quality evaluation method. In: Proc. International Picture Coding Symposium (1986)

Techniques for Image Classification, Object Detection and Object Segmentation*

Ville Viitaniemi and Jorma Laaksonen

Department of Information and Computer Science, Helsinki University of Technology, P.O. Box 5400, FI-02015 TKK, Finland
{ville.viitaniemi,jorma.laaksonen}@tkk.fi

Abstract. In this paper we outline the techniques which we used to participate in the PASCAL NoE VOC Challenge 2007 image analysis performance evaluation campaign. We took part in three of the image analysis competitions: image classification, object detection and object segmentation. In the classification task of the evaluation our method produced comparatively good performance, the 4th best of 19 submissions. In contrast, our detection results were quite modest. Our method's segmentation accuracy was the best of all submissions. Our approach for the classification task is based on fused classifications by numerous global image features, including histograms of local features. The object detection combines similar classification of automatically extracted image segments and the previously obtained scene type classifications. The object segmentations are obtained in a straightforward fashion from the detection results.

1 Introduction

In this paper, we discuss techniques for recognising and locating objects of specified semantic classes in images when example images of that semantic class are given. In particular, we summarise the techniques (more details in [4]) that we employed for participating the VOC Challenge 2007 image analysis performance evaluation, briefly outlined in Section 2. Section 3 describes our generic software framework for visual similarity assessment. In Section 4 we concretise how the framework is applied to the tasks of the VOC challenge. The results of the VOC evaluation are summarised and discussed in Section 5.

2 Image Content Analysis Tasks of VOC Challenge 2007

In PASCAL NoE Visual Object Classes Challenge 2007 [1], machine learning systems are compared by their ability to recognise objects from a pre-specified set of 20 visual object classes in realistic scenes. The problem is formulated as a supervised learning problem in which a training set of labelled images is

* Supported by the Academy of Finland in the *Finnish Centre of Excellence in Adaptive Informatics Research* project.

M. Sebillo, G. Vitiello, and G. Schaefer (Eds.): VISUAL 2008, LNCS 5188, pp. 231–234, 2008.

Fig. 1. Examples of VOC Challenge 2007 images and their annotations

provided. We participated in three of the competitions of the 2007 challenge: two main competitions in image classification and object detection, and a smaller-scale taster competition in pixel-level object segmentation.

The image collection for the 2007 challenge consists of 9963 photographic images of natural scenes. Each of the images contains at least one occurrence of the 20 object classes. The images of the training set have been manually annotated with the bounding boxes of all the occurrences of the 20 object classes. Additionally, pixel-wise segmentation masks of approximately 8% of the training images were provided as training material for the segmentation taster competition.

3 Framework for Visual Object Classification

In the following description of our software framework we use the term target object to denote training and test set images in the case of the classification task and automatically extracted image segments in the case of the object detection task. Our method for tackling the VOC Challenge tasks is based on assessing the similarity of the visual properties of the test set target objects to the properties of the training set target objects. Figure 2 schematically outlines the general system architecture used for supervised similarity assessment. As the input, the framework takes a binary partitioning of the target objects of the training set into two classes. As the output a real number is produced for each of the target

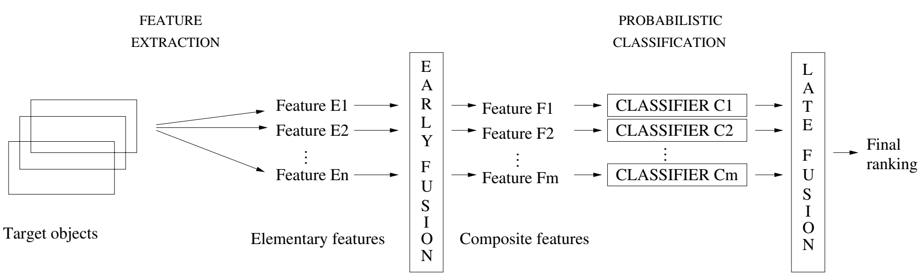


Fig. 2. The general architecture of the supervised similarity ranking framework

objects in the test set reflecting the estimated likelihood of the target object to belong to the first one of the classes of the binary partitioning.

In the first the visual properties of each target object are described simultaneously with several elementary feature vectors, each one reflecting a different aspect of the appearance, such as colour or texture. The next stage of the system is early fusion of the elementary feature vectors into composite features. The resulting sets of composite feature vectors are fed into supervised probabilistic SVM classifiers. Finally, in the last stage the outputs of the classifiers are fused together by an additional SVM layer.

4 Applying the Framework to the Challenge Tasks

In the classification task the test images are ranked according to their likelihood to contain objects from each one of the 20 classes. Here the framework is applied using whole images as target objects. The partitionings of the training images are produced in one-against-all fashion, thus resulting in one ranking per each object class. As elementary visual features, we use ten global or tiled image descriptors—such as colour and edge histograms—along with histograms of interest point SIFT features [2,5]. We employ several sets of SIFT features, varying parameters of the histogram generation such as codebook size and spatial subdivision of the images. The elementary features give rise to 141 composite features, and thus 141 separate classifiers. Before the final SVM fusion layer, we perform greedy forward/backward feature selection on the classifier outputs based on cross-validated training performance.

In the detection task the goal is to find and rank bounding boxes of objects of the 20 classes. The task is addressed by first segmenting the images with a rather rudimentary hierarchical region merging technique and then assessing the likelihood of each segment to correspond to the targeted object class. This problem is factored into product of two parts: 1) the likelihood of the image containing the segment to contain the object somewhere, and 2) the conditional probability of just the considered segment to correspond to the targeted object, given that the object is known to appear in the image. The former likelihood is readily available from the solution to the classification task. The latter one is obtained by applying the framework outlined in Section 3 with image segments as target objects. We use a feature set somewhat reduced but otherwise similar to that of the whole images to describe the image segments, augmented with two types of segment shape features. To produce the required rankings of bounding boxes—in contrast to irregularly shaped image segments—we simply list the bounding boxes of the segments in the rankings.

The submitted results for the segmentation taster task were synthesised in a straightforward manner from the bounding boxes detected in the detection task. For each pixel we checked all the detected bounding boxes containing the pixel. We set the label of the pixel to be that of the bounding box with the largest estimated probability if the probability exceeded a threshold. Otherwise the pixel was labelled as background.

5 VOC Challenge 2007 Results and Discussion

In this section we briefly report the our results in the VOC Challenge 2007 competitions we participated. For more details see [1]. In the classification contest, we placed 4th out of 19 participants, measured by median average precision (AP). The median AP we achieved was 0.506, the best median AP being 0.575. Our detection performance was modest, 7th best of 9 participants, when measured by counting class-wise placements in top-three of detection accuracy. Our method was in the top-three for five object classes. The object segmentation accuracy was determined for each object class by calculating the percentage of actual pixels of the class that have received the correct label. Mean accuracy over all the 20 classes and background class was used to rank the participants. In this ranking, our method was the best among 7 participants, with mean accuracy of 30.4%. All the segmentation contest entries except one were automatically derived from the bounding box detections.

Judging by the results, the classification performance of our method was relatively competitive. After a look in the other methods' details, it appears that the other top participants have used interest point histograms more effectively. In contrast, the feature fusion aspects of our architecture are more extensive than in the competing systems.

In the light of the modest detection performance of our system, the best overall segmentation accuracy among the challenge participants is rather surprising. It is somewhat mysterious why the segmentations derived automatically from our less accurate detections were better than similarly derived segmentations from more accurate detections. One explaining factor might be that our detection results consist of sets of overlapping bounding boxes that fully cover every image. After the challenge, it has been demonstrated [3] that the segmentation accuracy could still be markedly improved by re-segmenting the bounding boxes of our detections.

References

1. Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A.: The PASCAL Visual Object Classes Challenge (VOC2007) Results (2007), `http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html`
2. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
3. Shotton, J., Johnson, M., Cipolla, R.: Semantic texton forests for image categorization and segmentation. In: Proc. of IEEE CVPR (accepted, 2008)
4. Viitaniemi, V., Laaksonen, J.: Techniques for image classification, object detection and object segmentation applied to VOC Challenge 2007. Technical Report 2, Department of Information and Computer Science, Helsinki University of Technology (TKK) (2008)
5. Zhang, J., Marszałek, M., Lazebnik, S., Schmid, C.: Local features and kernels for classification of texture and object categories: a comprehensive study. International Journal of Computer Vision 73(2), 213–238 (2007)

A Visual Framework for the Definition and Execution of Reverse Engineering Processes*

Andrea De Lucia[1], Michele Risi[1], Luigi Rizzi[2], and Giuseppe Scanniello[2]

[1] Dipartimento di Matematica e Informatica, Università di Salerno, Via Ponte Don Melillo, 84084, Fisciano (SA), Italy
{adelucia,mrisi}@unisa.it
[2] Dipartimento di Matematica e Informatica, Università della Basilicata, Viale Dell'Ateneo, Macchia Romana, 85100, Potenza, Italy
gigi.rizzi@gmail.com, giuseppe.scanniello@unibas.it

Abstract. In this paper we present a visual framework developed as an Eclipse plug-in to define and execute reverse engineering processes aimed at comprehending traditional and web based information systems. Processes are defined in terms of UML activity diagrams, where predefined or newly developed software components can be associated to each activity. Components implemented using either traditional programming languages or software environments for data analysis (i.e., MATLAB or R) can be reused. Once the process has been fully defined the software engineer executes it to reverse engineering and comprehend software systems. The proposed visual framework has been evaluated on two case studies.

Keywords: Program comprehension, reverse engineering, legacy systems.

1 Introduction

In the software life cycle, the maintenance is the largest and the most expensive activity [32]. Starting after the delivery of the first release of the software, maintenance lasts much longer than the initial development phase [19]. During the maintenance phase a software system will be continuously changed and enhanced for several reasons. For example, changes are needed to meet new user requirements, to adapt software to interact with external entities, including people, organizations, and artificial systems, to correct faults, to improve performances, etc. [4]. However, a successful software system is condemned to change over time [21] as Lehman states in the first law "A program that is used in a real world environment necessarily must change or become progressively less useful in that environment" [20].

Technical and managerial problems contribute to the costs of software maintenance. The effort devoted to maintain a software system increases in case of an adequate documentation is lacking. However, supporting tools to reverse engineering existing or legacy systems can greatly reduce the maintenance effort. Reverse engineering generally

* This work has been supported by the project METAMORPHOS under grant PRIN-2006-2006098097.

M. Sebillo, G. Vitiello, and G. Schaefer (Eds.): VISUAL 2008, LNCS 5188, pp. 235–246, 2008.

supports the comprehension of existing software systems, abstracting higher level models from lower level representations of a software system. [7]. Typically, the code is the major source of information about the structure and the behavior, and is generally used to extract a set of potentially useful views [31].

There are different software tools available that can be used to reverse engineering and comprehend existing software systems, some freeware and others costing thousands of dollars or euros [27][34]. Generally, all-in-one reverse engineering and comprehension tools often lack. Hence, in case of a specific tool is needed a high effort is required to design and develop it. In the literature several software environments tools have been proposed to ease the production of reverse engineering tools [5][6][14]. These environments are generally based on some domain specific or scripting language that allows to easily prototype and/or generate reverse engineering tools. However, even using these environments, the development of reverse engineering tools might require a sensible effort and experience. It is worth noting that many reverse engineering components are available that might simply be combined within reverse engineering processes. Examples of such components are source code analyzers, clustering tools, graph visualizers, and so on. Most of the effort in this case, consists of integrating these components into the required reverse engineering process.

In such a scenario, workflow management technologies [35] could represent a viable solution. In fact, they have been conceived to support the modeling, the integration, and the orchestration of processes composed of different and heterogeneous software components. However, Workflow Management Systems (WfMSs) have not been used to integrate different tools and components during the definition and execution of reverse engineering and comprehension processes.

This paper presents a visual framework, to define and execute processes aimed at comprehending both traditional and web based information systems. This framework has been developed as an Eclipse plug-in and presents a visual environment where the software engineer defines processes expressed in terms of UML activity diagrams. Predefined or newly developed components can be associated to each activity of the process. Software components implemented either using traditional programming languages or software environments for data analysis (i.e., MATLAB or R) can be also reused. The framework also provides a set of predefined components ready to be used in the definition of the process. Once the process has been fully defined the software engineer executes it within the framework. In case the early activities of a process require a lot of time and they have been previously executed the software engineer can decide to skip them and run however the process. We assessed the framework on two case studies. In particular, two processes have been defined and executed to remodularize legacy information systems and to comprehend existing web applications [9], respectively.

The remainder of the paper is organized as follows. In Section 2 we present related work, while the framework is discussed in Section 3. The considered case studies are presented in Section 4, while final remarks and future work conclude the paper.

2 Related Work

In the following subsections we first discuss the relation of our work with workflow management and then the differences of our approach with respect to work on environments for generating reverse engineering tools. Finally, we discuss related work on clustering based reverse engineering tools that is related to the case studies presented in our paper.

2.1 Workflow Management Systems

Workflow Management Systems (WfMSs) [35] have been developed by researchers to provide support to the modeling, improvement, and automation of business and industrial engineering processes [8][15]. In case they are used to support software processes [18][22] WfMSs are also named as PSEE (Process-centered Software Engineering Environment). Most of the WfMSs are client-server systems, with centralized enactment facilities, although they do not exploit the web as basic infrastructure to ease the accessibility by remote users. Recent research on workflow management is focusing on the use of web technologies and/or specialized middleware to support distributed processes across organizations [2][8][15][22].

A number of Process Definition Languages have been proposed in the literature, based on several formalisms such as event-condition-action mechanisms [2], graph rewriting mechanism [18], Petri Nets [3], etc.. Several authors propose to adopt UML for representing business processes [12][25], as it has been conceived for the communication among people and it a notation that can be easily understood and used. Although the usefulness of WfMSs based on UML has been widely recognized in the software engineering field [12][18][25], they have not been used to integrate different tools during the definition and execution of reverse engineering and comprehension processes.

2.2 Tools for Developing Reverse Engineering Environments

Canfora *et al.* [5] propose a system for developing code analyzers. Indeed, the system is intended to produce analyzers to derive structural information at different abstraction levels, ranging from fine grained (e.g., control and data flow analysis and dependence) up to architectural models (e.g., structure charts, abstract data types, and object classes). Analyzers are automatically generated from a high-level specification expressed in a domain-oriented language and can be also embedded into C programs using foreign interfaces. The system and the underlying approach have been also assessed on a case study for software remodularization. The achieved results are presented and discussed in [6].

Ducasse *et al.* [14] describe the Moose reengineering environment. It is focused on a language independent meta-model, and offers services like grouping, querying, navigation, and advanced tool integration mechanism. Differently from our framework it does not provide a visual environment where processes to reverse engineering or comprehend legacy information systems can be defined and executed. Wong in [34] describes the Rigi system. This system provides several features and supports the comprehension and the re-documentation of existing software systems implemented in C, C++, and COBOL. A visual environment is also provided to help software engineers. Let us note that the Rigi system is adaptable to different languages only to analyze software systems from the structural point of view. The Software Refinery toolkit [27] uses an object-oriented database, called REFINE, to store a fine grained program model consisting of an attributed abstract syntax tree. This toolkit offers languages based on the procedural, declarative, and object-oriented paradigms to implement analyzers and to produce parsers and user interfaces, respectively.

Our approach is different from the previous approaches, as we concentrate on the definition and execution of reverse engineering processes and the integration and orchestration of the different components and analyzers required during the comprehension process.

2.3 Tools and Approaches Based on Clustering

Reverse engineering approaches are often focused on structural information. Nevertheless, the domain knowledge of the developers is generally embedded in the code comments. Kuhn *et al.* in [23] describe an approach to group software artifacts using Latent Semantic Indexing [10], a widely known and adopted information retrieval technique. The approach is language independent and tries to group source code containing similar terms in the comments. The authors consider different levels of abstraction to understand the semantics of the code (i.e., methods and classes). The tool implementing the approach is named Hapax and is built on top of the Moose reengineering environment [14]. Case studies are used to assess the approach and the tool support.

To group software artifacts clustering based approaches have been largely adopted in the past [1][9][11][28][30][33]. For example, Wiggerts [33] introduces clustering algorithms widely employed to group entities into software modules. This work has been extended by Anquetil and Lethbridge in [1]. In particular, the authors present a comparative study of different hierarchical clustering algorithms and analyze their properties with regard to software remodularization. In [29] the Arch tool is discussed. It is a graphical and textual structure chart editor for maintaining large software systems that can be used to group related procedures into modules using a clustering algorithm. Clustering algorithms have been also used to comprehend legacy web applications. For example, Di Lucca *et al.* [11] propose a clustering method to decompose a web application into groups of functionally related components. Ricca and Tonella in [28] propose a semi automatic approach to identify clusters of duplicated or similar pages to be generalized into a dynamic page. Similarly, in [30] the same authors propose a semiautomatic approach to identify and align static HTML pages whose structure is the same and whose content is in different languages. An approach based on a competitive clustering algorithm is proposed in [9] to identify similar pages at the structural level. Page structures are encoded into strings and then the Levenshtein algorithm is used to compute the distances between pairs of pages.

3 The Framework

The proposed framework enables to visually define and execute processes in terms of UML activity diagrams to reverse engineering or comprehend existing traditional and web information systems. The framework has been implemented as an Eclipse plug-in. A snapshot of the plug-in is shown in Fig. 1. In particular, this figure shows one of the processes used to assess the framework (further details can be found in Section 3).

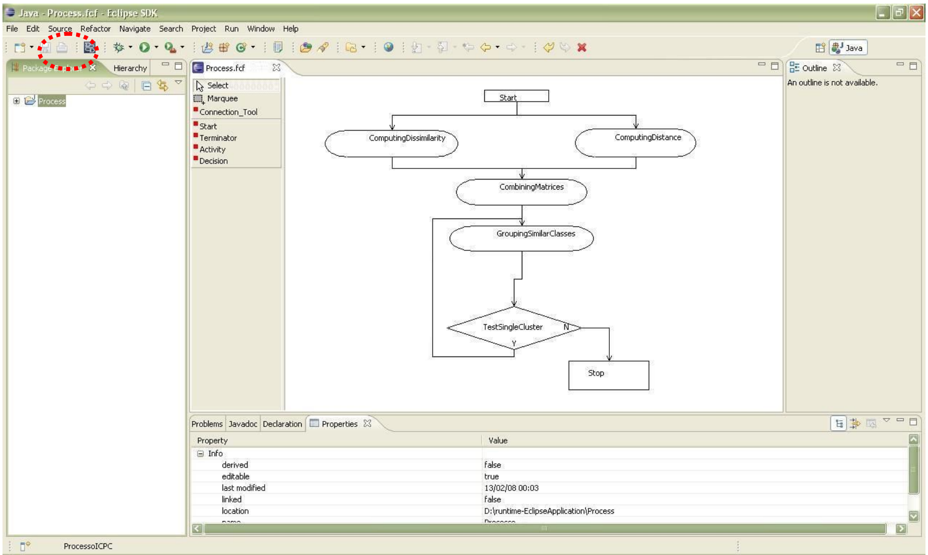


Fig. 1. The proposed plug-in

A software engineer first arranges the activities of a process and then associates to each activity a software component. These components implemented can be implemented using both traditional programming languages and software environments for data analysis. Indeed, we provided great emphasis on reusing software component implemented using the R[1] and MATLAB[2] software environments for data analysis. To integrate software components implemented in R the framework provides an automatic support using Rserve[2]. In particular, a Java class is used to remotely invoke procedures written in R, which in turn are encapsulated by employing an Adapter pattern [16]. MATLAB routines are integrated using the JMatLink engine, which is available under GPL license from http://www.held-mueller.de/JMatLink/. Similarly to the R procedures an Adapter pattern is used to wrap MATLAB routine.

In order to associate existing software components to the activities of the defined process the plug-in provides an easy to use mechanism. The framework provides predefined components implemented in R, MTLAB, and Java, while are ready to be reused. For example, Fig. 2 shows how an existing component implemented in R is associated to an activity (i.e., *CombiningMatrices*) of the process shown in Fig. 1. Information on the process in terms of activities and their relationships are stored in a XML file. This file also contains information concerning the software components associated to the activities of the process. An excerpt of the XML file of the process shown in Fig. 1 is depicted in Fig. 3. In particular, Fig. 3 shows how the activity CombiningMatrices and the *lsa_procedure.r* software component have been associated. The procedure of *lsa_procedure.r* that has to be invoked (i.e., *mergeMatrixFromFile*) when the process is executed is shown as well.

[1] http://www.r-project.org/
[2] http://www.mathworks.com/

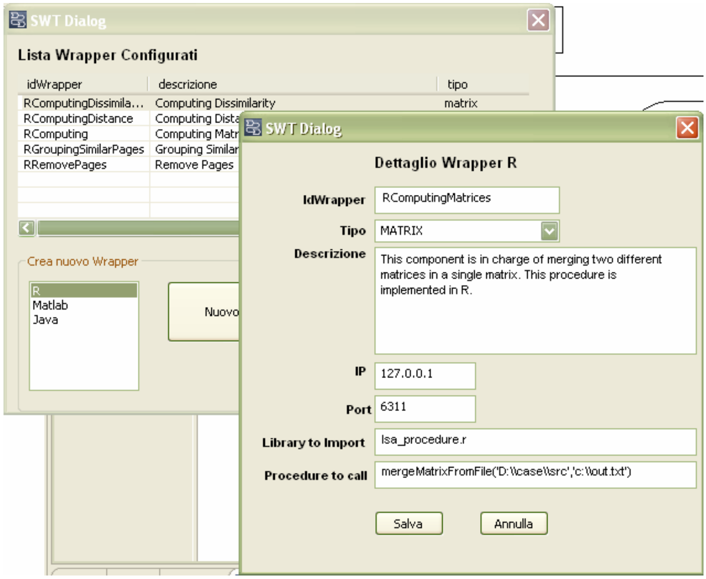


Fig. 2. Associating a existing software component to an activity

```
<matrix  id="RComputingMatrices" class-name="cddTools.MatrixGeneratorR" class-path="">
        <parametri-generali>
        <p-g name="IP" value="127.0.0.1" />
        <p-g name="PORT" value="6311" />
        <p-g name="PROC" value="mergeMatrixFromFile('D:\\case\\src','c:\\out.txt')" />
        <p-g name="LIB" value="source('lsa_procedure.r')" />
        </parametri-generali>
        <parametri-posizionali></parametri-posizionali>
</matrix>
```

Fig. 3. An excerpt of the XML file of the process

The execution of an R existing software component within the defined process is enabled using a wrapper that has been implemented once for all. An excerpt of the wrapper code is shown in Fig. 4. The wrapper uses the class *RManager* to enable the communication with existing components. The information specified during the process definition (see Fig. 1 and Fig. 2) is employed as well. For example, the values of the parameters allow the framework to get the location of the existing software components, to specify the library to use, and to invoke the procedures to be employed in the process execution. Software components implemented in MATLAB are managed in similar way through a middleware component implemented once for all. Consequently, the method to associate existing MATLAB or R components to the activities of a given process is nearly the same. The framework also enables the reuse of existing component implemented in Java. In such a case wrappers are not required.

The framework can be easily extended to manage existing software components implemented using different technologies. This is due to the flexible architecture we have designed. However, in case software components implemented using different technologies have to be managed a suitable software wrapper has to be implemented.

```
public Object esegui() throws Exception {
        String result="0";
        try {
          rUtility.RManager rMan = new RManager(this.getParametroGenerale("IP"),
                  Integer.parseInt(this.getParametroGenerale("PORT")),
                  this.getParametroGenerale("LIB"));
          rMan.init();
          rMan.start();
          result=rMan.procedure(this.getParametroGenerale("PROC").replace("'", "\'"));
          rMan.stop(false);
        } catch (RException e) {
                  logger.error("esegui()", e);
        }
        return "1";
}
```

Fig. 4. The wrapper code

In case a software engineer has to define a new activity, the framework proposes an editor where Java source code can be written. To specify conditions (i.e., diamond within a UML activity diagram) the same Java editor is used, which also provides features to highlight and indent the written code. The code of both the activities and the conditions is compiled within the editor and then executed when the software engineer runs the process.

The plug-in also provides a higher level of reusability. Indeed, a software engineer can reuse predefined processes and then associate to each activity a newly or an existing software component. This level of reusability was introduced as we observed that some reverse engineering approaches proposed in the literature [9][11][28][30][33] share nearly the same process.

Once the process has been fully defined the software engineering executes it within the framework (see dashed ellipse in Fig. 1). In case the early activities of a process require a lot of computation time and they have been previously executed the software engineer can decide to skip them. This is possible due to the fact that closer activities communicate by storing the produced artifacts on the prototype file system.

4 Case Studies

The framework has been assessed on different processes to reverse engineering and comprehend existing traditional and web based systems, respectively. In this section we presented and described two processes. The first process aims at remodularizing two existing objects oriented software systems, while the second has been conceived to identify similar web pages.

4.1 Software Remodularization

To remodularize software systems implemented in Java we defined the process depicted in Fig. 1, which is composed of the following phases: *ComputingDissimilarity*, *ComputingDistance*, *CombiningMatrices*, and *GroupingSimilarPages.*

To identify dissimilarity/distance between classes the comments have been considered. In particular, the ComputingDissimilarity activity uses Latent Semantic Indexing (LSI)

[10][17], a well known information retrieval technique, to compare pairs of classes from the semantic point of view. This phase takes as input the comment that has been previously extracted from the classes of a legacy system. The content is first normalized and then scanned to build the term-by-comment matrix. To obtain the latent structure of the system we apply a Singular Value Decomposition (SVD) to the term-by-comment matrix. For purposes of intuition, SVD can be geometrically interpreted. Generally, terms and comments could be represented as vectors in the k space (i.e., the singular values of the dimensionality reduction of the latent structure of the contents) of the underlying concepts. Similarly, each term and comment could be represented by a vector in the k space of the concepts.

Once the k space is computed the similarity between pairs of classes can be computed. To this end, we use the cosine between the vectors of the pairs of classes. This measure ranges from -1 (when they have a different latent structure) to 1 (when the latent structure is the same). To group classes we defined a measure obtained normalizing the cosine similarity from 0 (when the latent structure is the same) to 1 (when they have a different latent structure).

The Levenshtein string edit distance algorithm [24] is used in ComputingDistance to compare class comment. The Levenshtein algorithm is one of the most important dynamic algorithms for string matching. This algorithm is based on the notion of edit operation and consists of a set of rules that transform a given string into a target one. In particular, the Levenshtein edit distance is defined as the minimum number of insert, delete, and replace operations required to transform a source string x into a target string y. We assume that the insert and delete operations have cost 1, while the replacement has cost 2. The Levenshtein edit distance of the comments of the classes c_1 and c_2 ranges from 0, in case the content is the same, to 2 * maxlength(c_1, c_2), in case the content is completely different. We normalize this measure from 0 (when the comment is the same) to 1 (when the comment is different).

The matrices produced in the ComputingDissimilarity and ComputingDistance phases, are then used to produce a single matrix in the phase CombiningMatrices. In particular, for each pair of classes we computed the arithmetic mean of the measures based on the Levenshtein distance and the dissimilarity measure based on LSI. The resulting matrix is then provided as input to the activity GroupingSimilarClasses, which groups classes containing similar comments using a clustering algorithm [13]. Indeed, the k-means clustering algorithm [26] has been considered. This clustering algorithm classifies classes through a priori fixed number of disjoint clusters (the tuning value of our approach). The main idea is to define a centroid for each cluster of classes to identify. The centroids should be placed in a cunning way because of different location causes different results of the algorithm. In particular, they are placed as much as possible far away from each other. Generally, the k-means algorithm iteratively refines the initial centroids, minimizing the average distance/ dissimilarity, i.e. maximizing the similarity, of classes to their closest centroids. At each iteration, the clusters are built by assigning each cluster to the closest centroids. Here we adopted a variant of the k-means that minimizes the sum of the distances/dissimilarities instead of the sum of squared Euclidean distances. The considered variant is composed of two phases: *building* and *swap*. In the building step whether the centroids are specified or not, the algorithm looks for a good initial set of centroids. On the other hand, in the swap phase the algorithm finds a local minimum for the objective function, that is, a solution such that there is no single switch of an with a centroid that minimize the sum of the distances/dissimilarities.

The considered process presents a cycle (see Fig. 1), which enables us to execute the GroupingSimilarClasses activity according to different tuning values. In particular, given a legacy system composed of *n* classes, we used as tuning value numbers ranging from n/2 down to 1. The iteration is concluded when no single clusters (i.e., clusters containing only one class) are identified.

We have executed the defined process on two Java systems, i.e., jEdit 4.3 and JHotDraw 5.1. The former is available under GPL 2.0 license at http://www.jedit.org, while the latter is available at http://www.jhotdraw.org/ under LGPL license. The jEdit system is a widely employed programmer's text editor with hundreds of person-years of development behind it. On the other hand, JHotDraw is a two-dimensional graphics framework for structured drawing editors.

4.2 Grouping Similar Pages

The general process we have defined to group similar static and dynamic web pages according with a given concern is shown in Fig. 5. Suitable representations of the pages to consider in the identification of similar pages are produced in the Page Transformation phase. Successively, the Computing Distance Matrix phase uses the page representations to identify distances between the pairs of pages of a given web application. The distances between pairs of pages are then used to build the Distance Matrix, which is provided as input to the phase Clustering Web Pages. This phase uses clustering algorithms to group similar pages with respect to the adopted distance measure. We show here an instance of the process to identify groups of similar pages that are similar at the structural level. The structures of the web pages are encoded into strings in the Page Transformation phase. Distances between pairs of pages are computed using the Levenshtein algorithm [24] (see section 4.1 for details) on the strings encoding the web pages. Page distances are used in the Computing Distance Matrix phase to get the distance matrix of a given web application. In the Clustering Web Pages phase we have used one of the widespread partitional clustering algorithms to group observations from the statistical recognition perspective, i.e., WTA (Winner Takes All) [13].

WTA is a simplest artificial neural network clustering algorithm. It aims at grouping and representing similar patterns by a singular unit also called neuron. The grouping is done automatically based on correlations of the data points. The correlation between two data points is represented by the distance of the corresponding pages. Clusters, whose number is previously chosen, are presented at the input and are associated with the output nodes. The weights between the input nodes and the output nodes are iteratively changed until a termination criterion is satisfied. The network is said to be stable if no pattern in the training data changes its category after a finite number of learning iterations. Further detail can be found in [9].

The approach has been evaluated on two web applications developed using JSP technology. In particular, we considered the web application of the 14th International Conference on Software Engineering and Knowledge Engineering (SEKE 02) and the web application named SRA (Student Route Analysis). SEKE 02 was used to support the organizers and the academic community for paper submission, refereeing, and

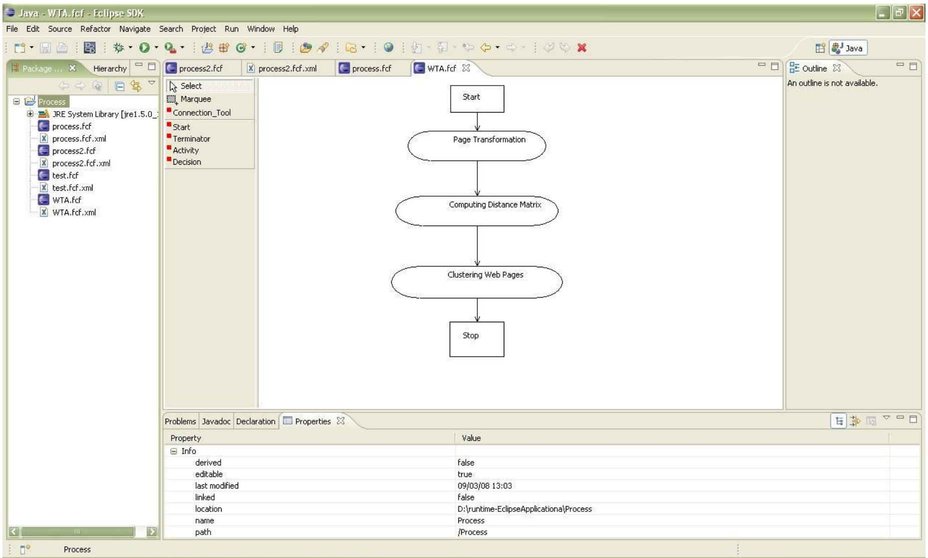


Fig. 5. A process to group similar pages

conference registration. On the other hand, the SRA web application was devised to provide the students in Politics Science at University of Salerno with statistics and information on their academic carrier. We have executed the defined process on these applications. To identify possible differences in using the proposed framework, we have compared the achieved results with the outcomes presented in [9]. As expected, we did not identify any difference in results.

5 Conclusions and Future Work

In this paper we have presented a framework developed as an Eclipse plug-in to define and execute processes to reverse engineering and comprehend existing web based and traditional software systems. The plug-in integrates a visual environment where processes expressed in terms of UML activity diagrams are specified. Software engineers complete the definition of the process associating newly developed or predefined software components (i.e., implemented in R, MATLAB, and Java). The association between existing components and activities is specifiable in easy way using a form based user interface. However, due to the designed architecture, the framework can be easily extended to manage existing software components implemented using different technologies. The framework also supports the reuse and the customization of previously defined processes. The framework has been assessed on processes to comprehend existing web applications [9] and remodularize legacy information systems, respectively.

We are currently working to further assess the proposed environment. Future work will be also devoted to further experiment the environment on different processes. We also plan to empirically validate the usefulness of the plug-in. To this end, students, academic researchers, and professional programmers will be selected as subjects. In particular, we plan to perform controlled experiments and actual industrial case studies with professional programmers. Usability studies will be also performed.

References

[1] Anquetil, N., Lethbridge, T.C.: Experiments with Clustering as a Software Remodularization Method. In: Proc. of 6th Working Conference on Reverse Engineering, pp. 235–255. IEEE CS Press, Los Alamitos (1999)

[2] Aversano, L., De Lucia, A., Gaeta, M., Ritrovato, P., Stefanucci, S., Villani, M.L.: Managing Coordination and Cooperation in Distributed Software Processes: the GENESIS Environment. Software Process Improvement and Practice 9(4), 239–263 (2004)

[3] Bandinelli, S., Di Nitto, E., Fuggetta, A.: Supporting cooperation in the SPADE-1 environment. IEEE Trans. on Software Engineering 22(12), 841–865 (1996)

[4] Canfora, G., Cimitile, A.: A Reference Life-cycle for Legacy Systems. In: Proc. of the ICSE Workshop on Migration Strategies for Legacy Systems, Boston, MA (1997)

[5] Canfora, G., Cimitile, A., De Carlini, U., De Lucia, A.: An extensible system for source code analysis. IEEE Trans. on Software Engineering 24(9), 721–740 (1998)

[6] Canfora, G., De Lucia, A., Di Lucca, G.A.: A System for Generating Reverse Engineering Tools: A Case Study of Software Modularisation. Automated Software Engineering 6(3), 233–263 (1999)

[7] Chikofsky, E.J., Cross, J.H.: Reverse Engineering and Design Recovery. A Taxonomy 7(1), 13–17 (1990)

[8] Cugola, G., Di Nitto, E., Fuggetta, A.: The JEDI event-based infrastructure and its application to the development of the OPSS WfMS. IEEE Transactions on Software Engineering 27(9), 827–850 (2001)

[9] De Lucia, A., Scanniello, G., Tortora, G.: Identifying Similar Pages in Web Applications using a Competitive Clustering Algorithm. Journal on Software Maintenance and Evolution 19(5), 281–296 (2007)

[10] Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., Harshman, R.: Indexing by Latent Semantic Analysis. Journal of the American Society for Information Science (41), 391–407 (1990)

[11] Di Lucca, G.A., Fasolino, A.R., De Carlini, U., Pace, F., Tramontana, P.: Comprehending web applications by a clustering based approach. In: Proc. of the 10th International Workshop on Program Comprehension, pp. 261–270. IEEE CS Press, Los Alamitos (2002)

[12] Di Nitto, E., Lavazza, L., Schiavoni, M., Tracanella, E., Trombetta, M.: Deriving executable process description from UML. In: Proceedings of 24th International Conference on Software Engineering, Orlando, Florida, pp. 155–165. ACM Press, New York (2002)

[13] Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification, pp. 576–581. Wiley-Interscience Publication, John Wiley & Sons, Inc., New York

[14] Ducasse, S., Girba, T., Lanza, M., Demeyer, S.: Moose: a Collaborative and Extensible Reengineering Environment. Tools for Software Maintenance and Reengineering, RCOST / Software Technology Series, Franco Angeli, pp. 55–71 (2005)

[15] Eder, J., Panagos, E.: Towards Distributed Workflow Process Management. AT&T Research Labs (1999)

[16] Gamma, E., Helm, R., Johnson, R., Vlissides, J.: Design Patterns: Elements of Reusable Object-Oriented Software. Addison-Wesley, Menlo Park (1995)
[17] Harman, D.: Ranking Algorithms. In: Information Retrieval: Data Structures and Algorithms, pp. 363–392. Prentice-Hall, Englewood Cliffs (1992)
[18] Heimann, P., Joeris, G., Krapp, C.A., Westfechtel, B.: DYNAMITE: dynamic task nets for software process management. In: Proceedings of the 18th International Conference on Software Engineering, Berlin, Germany, pp. 331–341 (1996)
[19] IEEE Standard for Software Maintenance. IEEE Std 1219-1998. The Institute of Electrical and Electronics Engineers, Inc. (1998)
[20] Lehman, M.M.: Lifecycles and the Laws of Software Evolution. Proc. of IEEE, Special Issue on Software Engineering 19, 1060–1076 (1980)
[21] Lehman, M.M.: Program Evolution. Journal of Information Processing Management 19(1), 19–36 (1984)
[22] Kaiser, G.E., Dossick, S.E., Jiang, W., Yang, J.J., Xi Ye, S.: WWW-based Collaboration Environments with Distributed Tool Services. World Wide Web Journal 1(1), 3–25 (1998)
[23] Kuhn, A., Ducasse, S., Girba, T.: Enriching reverse engineering with semantic clustering. In: Proc. of 12th Working Conference on Reverse Engineering, pp. 10–20. IEEE CS Press, Los Alamitos (2005)
[24] Levenshtein, V.L.: Binary codes capable of correcting deletions, insertions, and reversals. Cybernetics and Control Theory 10, 707–710 (1966)
[25] Marshall, C.: Enterprise Modellingwith UML: Designing Successful Software trough Business Analysis. Addison-Wesley, Reading (2000)
[26] Mcqueen, J.: Some methods for classification and analysis of multivariate observations. In: Proc. of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, pp. 281–297 (1967)
[27] Reasoning Systems, REFINE User's Guide, Reasoning Systems, Palo Alto, CA (1989)
[28] Ricca, F., Tonella, P.: Using Clustering to Support the Migration from Static to Dynamic Web Pages. In: Proc. of International Workshop on Program Comprehension, Portland, Oregon, USA, pp. 207–216 (2003)
[29] Schwanke, R.W.: An Intelligent Tool for Re-engineering Software Modularity. In: Proc. of the 13th International Conference on Software Engineering, pp. 83–92 (1991)
[30] Tonella, P., Ricca, F., Pianta, E., Girardi, C.: Restructuring Multilingual Web Sites. In: Proc. of International Conference on Software Maintenance, Montreal, Canada, pp. 290–299. IEEE CS Press, Los Alamitos (2002)
[31] Tonella, P., Potrich, A.: Reverse Engineering of Object Oriented Code. Springer Science, Heidelberg (2002)
[32] Zelkowitz, M., Shaw, A., Gannon, J.: Principles of Software Engineering and Design. Prentice-Hall, Englewood Cliffs (1979)
[33] Wiggerts, T.A.: Using Clustering Algorithms in Legacy Systems Remodularization. In: Proceedings of 4th Working Conference on Reverse Engineering, pp. 33–43. IEEE CS Press, Los Alamitos (1997)
[34] Wong, K.: Rigi blurb (1996), http://rigi.cs.uvic.ca/downloads/papers/pdf/blurb-rigi.pdf
[35] Workflow Management Coalition: Workflow Management Coalition Interface 1: Process Definition Interchange Process Model. Doc. no. WFMC-TC-1016-P (1999)

An Evolutionary General Purpose WebGIS to Disclose EGFR Mutations in Lung Cancer

Domenico Aliotta[1], Pietro Buffa[2], and Gennaro Iaccarino[3]

[1] ProdAl S.c.ar.l.
Università di Salerno, Via Ponte don Melillo, 84084 Fisciano (SA), Italy
nicoaliotta@prodalricerche.it
[2] Dipartimento di Scienze Biomediche, sez. Patologia Generale
Università di Catania, Via Androne 83, 95124 Catania, Italy
pietrobuffa@hotmail.com
[3] ISISLab - Dipartimento di Informatica ed Applicazioni "R.M. Capocelli"
Università di Salerno, Via Ponte don Melillo, 84084 Fisciano (SA), Italy
iaccarino@dia.unisa.it

Abstract. The epidermal growth factor receptor (EGFR) is a membrane spanning receptor tyrosine kinase that regulates cell proliferation, survival, and migration. Over expression and mutations in the kinase domain of EGFR are implicated in the development and progression of a variety of cancers. We present an evolutionary general purpose Geographic Information System, in which geographic information is replaced by biological data representing the main mutation regions within the kinase domain of EGFR. Our system can be used for generating hypotheses about relations between EGFR mutations and cancer; defining statistical, molecular, and pharmacological relationships; disclosing biological identification techniques, etc. Moreover this is an important example of how GIS capabilities can be used not only for geographic items, but for any kind of data, as a general purpose system.

1 Introduction

The latest remarkable advances in science came from the structural biology area [17], a branch of biochemistry concerning the study of the architecture and shape of biological macromolecules (proteins in particular) and what causes them to have the structures that they have. Molecular data visualization is considered a key aspect in the study of main biological processes; moreover, it has been enhanced by using "high performance computing" and by developing computer graphic methods, which have allowed a faster acquisition of structural and functional information on different proteins of biomedical interest (such as EGFR). Unfortunately, many data visualization tools offer a specific graphic vision for every kind of information (metabolic pathways data, structural data, genomic data, experimental data, etc) that often can be hardly integrated, so that possible correlations among them are not easily pinpointed. For instance, Protein Data Bank (PDB) [1] provides several tools and resources to study the structures

M. Sebillo, G. Vitiello, and G. Schaefer (Eds.): VISUAL 2008, LNCS 5188, pp. 247–258, 2008.

of biological macromolecules and their relationships with sequences, functions, and diseases; Pmut [7] is a software that allows to predict pathological mutations; CUPSAT [14] is a web tool useful to analyse and predict protein stability changes upon point mutations. The heterogenity of these tools forces biologists to use them in different steps of their work, without correlations with results, in an attempt to obtain the desired information. These difficulties necessitate the realization of a new tool (Web based), that combines all these features with a simple data management and shows results for a better scientific knowledge.

Geographic Information Systems (GISs) were born as systems for capturing, storing, analyzing, managing data and associated attributes which are spatially referenced to earth. Features are categorized separately and stored in different map layers, which share a common coordinate space. In particular, layers can be added for locating measured data, such as annual rainfall, geological events, territorial changes ando so on (evolutionary concept). This way of organizing data in a GIS, makes maps much more flexible to use since these layers can be combined in any useful manner. Similarly, biological structures (proteins, genes, etc.) define a biological space coordinates, with the same characteristics of any other, thus perfectly performable with a GIS. According with these bases, we plan on dividing a protein in four fundamental levels of increasing complexity: 1) Amino acid sequence level; 2) Secondary structure organization level; 3) Domains organization level; 4) 3-Dimensional organization level. Each of these levels provides specific information. Integrating these different informative levels and adding dynamically other possible data levels (i.e. the known protein isoformes, the found mutations, the direct/indirect interaction with other proteins, the correlated pathologies and therapies, etc), give us the possibility to obtain much more details, with a consequent increase of global knowledge on the studied molecules.

2 Related Works

WebGIS technology has extensive application prospects in the field of biological data. It is a technique that sets up geographical informations on the Web, linked to databases. Users can get information, mutually by WebGIS application, via Internet. With the rapid development of WebGIS technology, GIS based applications can be developed with low costs and little maintenance work. WebGIS based system has become very popular [11], [15]. The state of the art in this field is divided into two different branches: the first one is focused on designing conventional GISs, able to help medical researchers in the study of environmental medical rescues, genetics tendency to diseases in regions, and pharmacological statistics; the second one is focused on a new approach to define graphical information, in which biochemical features replace spatial data, improving interoperability, data search and management. Our system falls under this second category. In the following subsections 2.1 and 2.2 we show some examples of both kind of systems.

2.1 Conventional GIS for Biomedical Goals

One of the first approaches of this kind was shown in [8], where a WebGIS system was developed for the accurate mapping of tick-borne diseases, with specific attention to Lyme borreliosis, which may cause cardiac manifestations such as atrio-ventricular abnormalities, myocarditis, and dilated cardiomiopathy. Moreover, they obtained an experimental risk map from the model, by considering altitude, week of sampling and vegetation type, as predictor variables. As the authors also say, this is a representative example of how a GIS can be used for a quick geo-location of medical data, and the creation of a central archive for disease risk observation [8]. In 2004 Xiaolin Lu proposed, in [11], a framework for the WebGIS based urgent medical rescue CSCW for SARS disease prevention. He adopted GIS and multimedia tech-nologies to build an integrated collaborative environment. The system provided the geographic distribution of epidemic diseases using a WebGIS as graphical user interface; it was able to analyze spatial and temporal trends of epidemic disease, providing a lot of services (alarm the risk area, assess resource allocation, perform decision-making, etc.) against them [11].

2.2 Bio-Medical Oriented GIS

In 2006, Dolan et al. developed one of the first examples of spatial information systems, in which geographic informations represent genome data [5]. This system was named GenoSIS (Genome Spatial Information System), and it was an application that used the concepts and tools of geographic informations science for the interpretation and modeling of genome data [5]. This system reused existing spatial analysis, classification, querying and visualization tools for a workstation (not on the web) data analysis. It represented a novel approach to several problems of current interest to molecular geneticists.

3 Lung Cancer and Epidermal Growth Factor Receptor

In this section we show relationship between the most common mutations in EGFR kinase domain and lung cancer, that accounts for one third of all deaths from cancer worldwide. It represents an appreciable case of studie for our system. Like most cancers, lung cancer is a conglomeration of diseases of diverse aetiology, broadly divided into small-cell lung cancer (SCLC, comprising 20% of lung cancers), and non-small-cell lung cancer (NSCLC, comprising 80% of lung cancers). The epidermal growth factor receptor (EGFR) is a membrane spanning receptor tyrosine kinase (RTK) that regulates cell proliferation, survival, and migration. It has been widely demostrated that EGFR is implicated in NSCLC, glioblastoma, and breast cancer, where its oncogenic potential is stimulated by protein overexpression or by somatic “gain-of-function” mutations [10]. Sections 3.1 and 3.2 show the most common mutations found in EGFR kinase domain and interactions with other proteins. These items will be used for a primary investigations on our system.

3.1 The Most Common Mutations Found in EGFR-KD

EGFR has been intensely studied in its domain organization. A number of distinct mutations have been structurally identified around the active site cleft of the EGFR kinase domain (EGFR-KD) [16]. These include: point mutation within the nucleotide-binding loop in exon 18, small deletions in exon 19, insertions in exon 20 and point mutations in the activation loop in exon 21. Figure 1 shows the distribution of exons in the extracellular, transmembrane, and intracellular domains [16]. The most common mutation (40% of NSCLC cases) is an arginine for leucine substitution at position 858 (L858R). On the contrary, one of the less frequent (5% of NSCLC cases) is a glicine for serine substitution at position 719 (G719S). Interestingly, the clinical correlation between the presence of specific mutations and therapeutic response to inhibitors drugs such as gefitinib and erlotinib is mirrored in cell lines and EGFR-transfected cells. Cells bearing the mutant EGFR are in general more sensitive to drugs than cells expressing the wild-type EGFR [9]. Figure 2 shows a 3D ribbon representation of EGFR kinase domain. Our WebGIS prototype uses these as example data set.

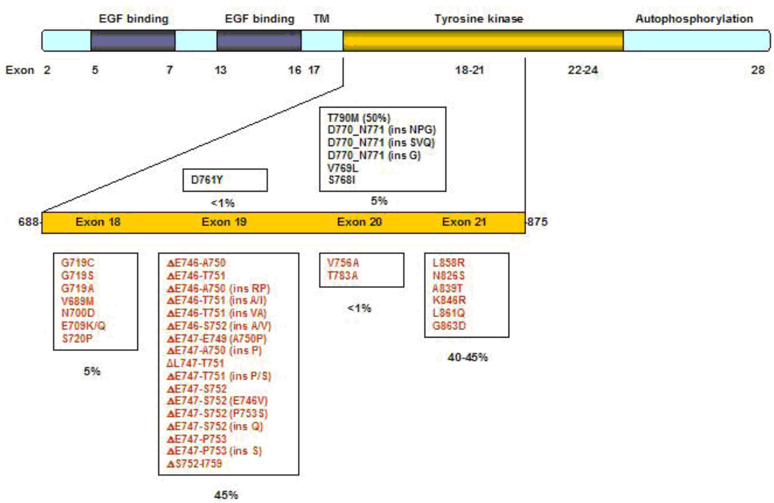


Fig. 1. A cartoon representation of the Epidermal Growth Factor Receptor (EGFR). It shows the distribution of exons in the extracellular (EGF binding), transmembrane and intracellular domains (comprising the tyrosine kinase and autophosphorylation regions). Exons 18/21, where relevant mutations are situated, are highlighted (yellow bar); and a detailed list of EGFR mutations is associated with sensitivity (red boxes) or resistance (black boxes) to the main drugs, "gefitinib" or "erlotinib". Moreover, frequencies of mutations are shown.

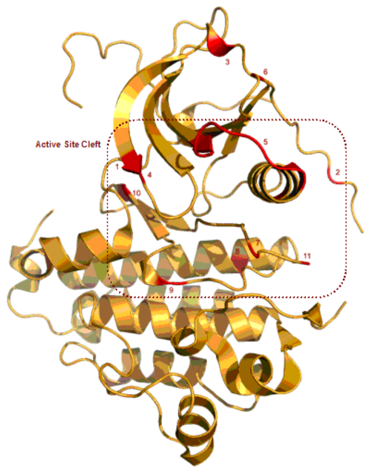


Fig. 2. A 3D ribbon representation of EGFR kinase domain (PDB:2itx). The protein regions more frequently subject to mutations are shown in red. 3-Dimensional structure representations were prepared by using the PyMol v0.98 [3].

Table 1. Predicted intracellular associations (*protein*^{2}*protein*)

Prot	ENSEMBL Code	Description
SHC1	ENSP00000336919	SHC transforming protein 1
GRB2	ENSP00000348125	Growth factor receptor-bound protein 2
CBL	ENSP00000264033	CBL E3 ubiquity protein ligase
SRC	ENSP00000353950	Proto-oncogene tyrosine-protein kinase Src
HBEGF	ENSP00000230990	Heparin-binding EGF-like growth factor precursor
RASA1	ENSP00000274376	Ras GTPase-activating protein 1
PLCG1	ENSP00000244007	1-phosphatidylinositol-4,5-bisphosphate phospho--diesterase 1
SOS1	ENSP00000263879	Son of seven-less protein homolog 1
TGFA	ENSP00000295400	Transforming growth factor alpha precursor
PTPN1	ENSP00000349544	Tyrosine-protein phosphatase, non receptortype 1
ERBB2	ENSP00000269571	Receptor tyrosine-protein kinase erbB-2 precursor
VAV2	ENSP00000317258	Vav-2 protein
CRK	ENSP00000300574	Proto-oncogene C-crk (P38)
PIK3R1	ENSP00000274335	Phosphatidylinositol 3-kinase regulatory alpha sub--unit
BTC	ENSP00000312050	Betacellulin precursor (BTC)
AREG	ENSP00000264487	Amphiregulin precursor (AR)
EPS8	ENSP00000281172	Epidermal growth factor receptor kinase substrate 8
PTPN11	ENSP00000340944	Tyrosine-protein phosphatase, non receptortype 11
ERBB3	ENSP00000267101	Receptor tyrosine-protein kinase erbB-3 precursor

3.2 The Most Common Interactions Found between EGFR and Other Proteins

Information on *protein*^{2}*protein* interaction is still mostly limited to a small number of model organisms, and originates from a wide variety of experimental and computational techniques. STRING (Search Tool for the Retrieval of Interacting Genes/Proteins), a database of known and predicted *protein*^{2}*protein* interactions led to 373 distinct organisms, and lists more than 1.5 million proteins for which associa-tions have been pre-computed [12]. The interactions include direct (physical) and indirect (functional) associations. This information is essential for a system-level understanding of cellular behaviour, and it is needed in order to place molecular functions of individual proteins into their cellular context. Table 1 lists predicted intracellular associations. In section 5.2 is shown how these data are available on the WebGIS.

4 System Overview

Our WebGIS infrastructure is based on an original modification and integration of several major open source software systems for data management (MySql), geographical data visualization through internet (MapServer), geographical data analysis (Quantum GIS), and Internet software management (Apache). A simple system architecture is shown in figure 3. Visual data and all the WebGIS functionalities are visualized in a browser window, and thus remain, as much as possible, independent from the hardware and the user's operative system. The service can be accessed from any personal computer, connected to the Web, using a common Internet browser or a desktop GIS application; providing also opportunities for the forthcoming mobile computing hardware. The infrastructure uses different modules of PMapper and MapServer, each one during a specific phase of the HTTP request life cycle. PMapper receives all client requests, from the Apache HTTP server, parses the MapFile and returns (to the HTTP server) the requested data through MapServer. MapServer is an open source development

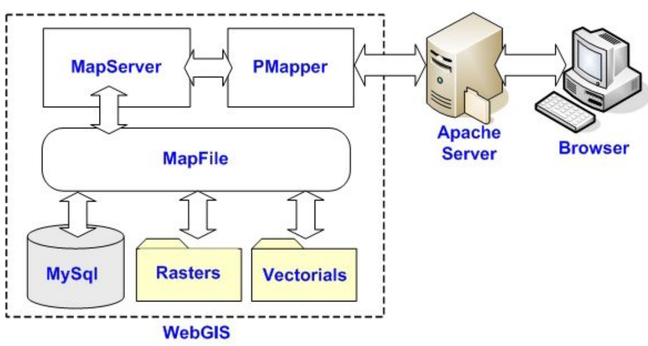


Fig. 3. A simple representation of the WebGIS architecture. The system can be accessed by simply using an Internet browser or a desktop GIS client. The Apache Web Server receives and processes HTTP requests, and invokes the MapServer (throw PMapper infrastructure). The MapFile includes information aboit data (rasters and vectorials) and is the system core. It is processed by MapServer and sent to the HTTP server (Apache) that will send it to the browser (client).

environment for building spatially-enabled internet applications; it provides to manage spatial data from the MapFile, producing the graphical output. MapFile defines the relationships between objects, polygons, points etc, where data are located, and how they have to be drawn. The MapFile has a hierarchical structure [13] in order to define the objects contained in it, and establishes a connection with data contained in MySql-DB. Moreover MapFile objects include information about vectorials and raster data. The application realized is multiplatform, and has been tested under Windows, Linux, and MacOSX. It supports a multitude of raster and vector data formats such as TIFF, GeoTIFF, and many others, via GDAL. The application can also use vector data from ESRI shapfiles, PostGIS, ESRI ArcSDE, Oracle Spatial, MySql and many others, via OGR. It is already able to access and provide data, using web services frameworks included in MapServer. The web services specifications for mapping come from the Open Geospatial Consortium, that developed mapping web services standards. The goal of using these WebServices is to improve interoperability between applications by creating some common interchange languages through common standards. The specifications empower technology developers to make complex spatial information and services accessible and useful with all kinds of applications.

5 Implementation

The polygons have been realized with Quantum GIS (QGIS), that is a user friendly open source Geographic Information System (GIS) that runs on Linux, Unix, Mac OSX, and Windows. QGIS supports vector, raster, and database formats, and allows browsing and creating map data on computers. It is licensed under the GNU General Public License. With QGIS we realized the shape files, then transformed in a MySql table, via shp2mysql tool. This tool takes in ESRI shape files and outputs SQL command files suitable for use in the MySql database (on version 4.1.0 and above). Then we used the files made with shp2mysql as source files for MySql command line client, and as "scripts" in the MySql query browser. After that, we wrote the MapFile. Map-File is the basic configuration mechanism for MapServer. The MapFile is the heart of MapServer. It defines the relationships between objects, points to MapServer where data is located and defines how things are to be drawn. The MapFile has a hierarchical structure, with the Map object being the "root". All other objects fall under this one. We established a connection between MapServer and our MySql-DB using OGR's MySql drivers to avoid having to set up an ODBC connection, and building a MySql select statement in our MapFile. We used PMapper framework because it has broad functionality and multiple configurations in order to facilitate the setup of our MapServer application based on PHP/MapScript.

5.1 Functionalities

GIS main window (fig. 4) is divided into three basic areas: 1) a main map generated by MapServer, surrounded by various control instruments, that reproduces

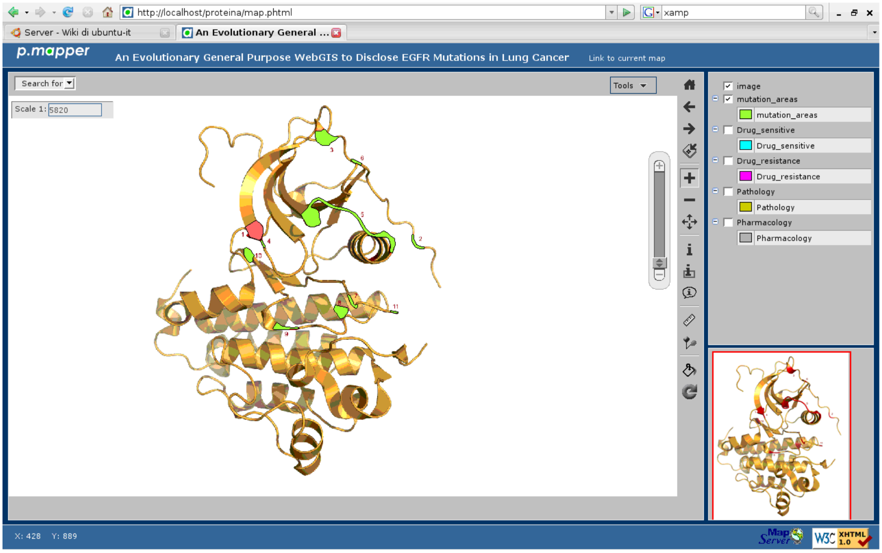


Fig. 4. WebGIS main window

the 3D ribbon representation of EGFR with all the activated layers; 2) an interactive legend through which users can activate and deactivate layers; 3) an interactive reference map, which offers also the navigation of the protein. The application offers several tools to make queries and research on the map; we provide a variety of queries ranging from the possibility of a punctual query-to-query based on an area, to the search for information based on text. All these functions can be activated using buttons in the pipe bar. Among the functionalities of our application there are the zoom and pan interface (without frames); by using them the user can browse the protein image easily. User can also zoom/pan through the map via keyboard keys, mouse wheel, reference map and slider bar. Users can query data in the database with three functions: identify, select and search; in fact there is a fully featured attribute search, including suggest, select boxes, etc., and query results can show joins from DB and hyperlinks (query results can be also exported). Query layout results (fig. 5) is really flexible because it is managed via JavaScript and can be shown with a pop-in identification by moving the mouse over the map, or with an i-frame into the map (auto-identify function). All data in the database can be printed out as HTML and PDF file. The application has a useful tool to measure distance and areas of the map. All data are figured also with HTML legends and various display styles of legend and layers/TOC, thanks to a very flexible configuration of functions, behavior and layout. The application has a multilingual user interface and predefined languages are: English, Italian, French and German. User can also add points of interest in the window with labels on the map.

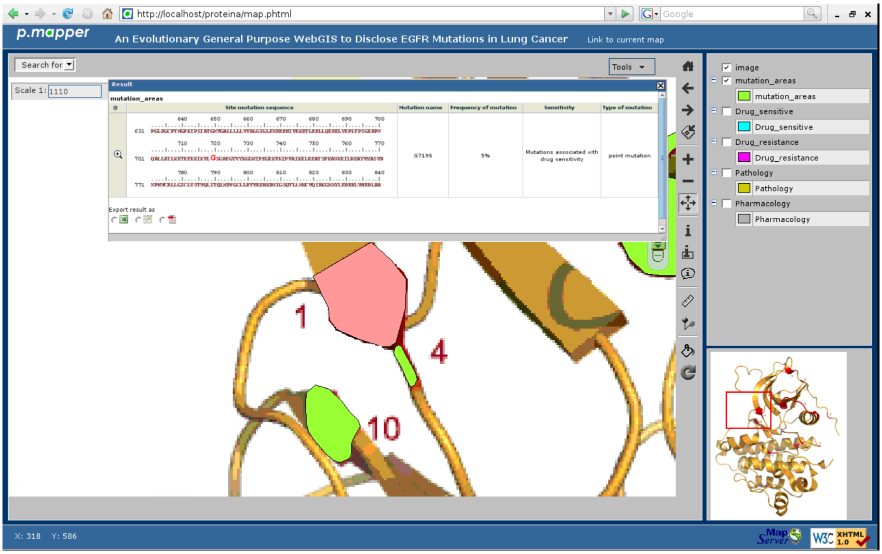


Fig. 5. An example of query layout. Mutation G719S: the pop-in shows the site mutation sequence, the name, the frequency, the sensitivity and the mutation type.

5.2 Layers

As shown in section 1, theoretical protein structure can be divided in four interactive levels of increasing complexity; some representeble as GIS layers ant other as database entry. in figures 4 and 5, the first prototype of the WebGIS provides five layers, each one containing information about EGFR mutations, as described in section 3. The first one contains information about mutations; in particular, in the database are collected data regarding the site mutation sequences, and the position where they happen, the mutations name, the frequencies with which they occur, and their typology (punctual or not). The second and third contain information about drug sensitivity/resistance; the fourth and fifth about pathology deriving from mutations, diseases, pharmacology experimentations and so on. Moreover, a lot of items regarding no-gis-data are provided. This category includes a dataset not depending on mutations, but regarding EGFR protein and cancer relationship. An example, introduced in 3.2, are the $protein^2protein$ interactions, that are viewable by clicking on the i icon of the toolbar; as shown in figure 6.

5.3 Accessibility

As widely shown in literature, a WebGIS, whatever its purpose, uses massive visual data, in order to represent different kinds of information on the Web. In the last ten years, scientists have provided several suggestions for web designers regarding how to make web pages accessible to everyone. Our system has been

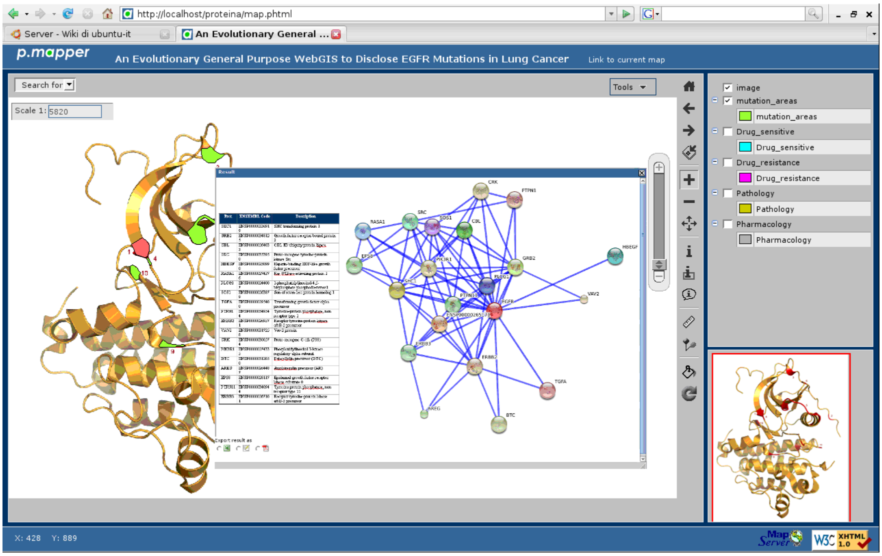


Fig. 6. Screenshot of no-gis-data. This example shows information regarding *protein*^{2}*protein* interactions (3.2) [12].

realized, considering guidelines and suggestions proposed in [4] and [6] as much as we could. In particular, our attention was focused on the most common color deficiencies vision (protanopy and deuteranopy) that affect 8% of the male population and 2% of the female, also known as colorblindness [4]. It represents a substantial subset of scientists and non-expert users for our system. In order to improve accessibility, our layers and spatial data were realized following (where needed) the *colorblind-filter* characteristics shown in [4]. Moreover, the used infrastructure (and version) guarantees *access-key* and *mouse-less* navigation, resulting W3C compliant.

6 Discussion and Future Works

Potential users of our application include biologists, doctors, researchers, as well as private collectors of biomedical data/objects, and non-expert people that will obtain information regarding diseases, statistics, therapies, and so on. Mutations in the EGFR kinase occur in approximately 16% of non-small-cell lung cancer and are one of its causes [18]; this is why there is still work to do in this field of research. Next step will be to implement a 3-Dimensional WebGIS, where 3D space will be used to store biomedical images linked from different databases, with different items, providing data-mining, graphical queries, and so on. This distributed 3D WebGIS could represent a huge improvement in biomedical science. Moreover, our goal will be to put in relation biological functions, associated diseases, related genes and gene-to-gene characteristics, allowing to manage, visualize and interact various level of information; in order to improve interoperability between different teams, working worldwide on medical research in cancer.

7 Conclusion

We have proposed a general purpose WebGis, based on the idea of spatial data replaced by biological data. It allows to manage various level of information on macromolecules (protein mutations, deriving diseases, pharmacology experimentations, and so on) with the ability to add new layers, or change items anytime, making it evolutionary. Our idea is to use GIS's strengths in order to improve 2D biological models and link them using multilayer, so it is possible to produce biological maps, that can be enriched with new layers, data, simulations, statistics and so on. Our goal is to show that general purpose GIS extends the concept of Geographical Information System, increasing usability for each kind of data, also providing a central scientific archive on the web. The strength of this system is to provide with researchers an easy-to-use service that is complete and multidisciplinary, and usable from different places. It is able to improve medical research in cancer mapping, analysing and exchanging data concerning proteins. It is possible thanks to the powerful tools of a typical WebGIS, such as WMS, WFS, and WMC. Using them it is possible to improve interoperability between applications, by creating some common interchange languages through common standards, so that applications can share and request vectors and rasters maps. They can be produced in plain image format, through Web Map Service (WMS); in vector map data and attributes (GML format), through Web Feature Service (WFS); and in a XML file format, saving WMS views through Web Map Context (WMC). In particular biomedical data, which often consist in different kinds of information, require simple instruments to be better arranged and displayed, so to ensure visibility, and to facilitate learning, research, therapy and so on.

Acknowledgement

The authors gratefully acknowledge the support of Antonio Aliotta for supervising preliminary version of this document, and would like to thank the anonimous reviewers for fruitful comments and suggestions which helped us to improve the final version of the paper.

References

1. Berman, H.M., et al.: The Protein Data Bank. Nucleic Acids Research 28, 235–242 (2000)
2. Chiang, J., Yu, H., Hsu, H.: GIS: a biomedical text-mining system for gene information discovery. Bioinformatics 20(1), 120–121 (2004)
3. De Lano, W.: PyMol: An Open-Source Molecular Graphics Tool. CCP4 Newsletter on Protein Crystallography, (40) (March 2002)

4. Del Percio, M., Iaccarino, G., Malandrino, D., Scarano, V.: Efficient edge-services for colorblind users. In: Proceedings of the 15th International conference on World Wide Web, pp. 919–920. ACM Press, New York (2006)
5. Dolan, M.E., Constance, C., Holden, M., Beard, K., Bult., C.J.: Genome as geography: using GIS technology to build interactive genome feature maps. BMC Bioinformatics 7, 416 (2006)
6. Erra, U., Iaccarino, G., Malandrino, D., Scarano, V.: Personalizable edge services for web accessibility. Universal Access in the Information Society 6(3) (2007)
7. Ferrer-Costa, C., Gelp, J.L., Zamakola, L., Parraga, I., de la Cruz, X., Orozco, M.: PMUT: a web-based tool for the annotation of pathological mutations on proteins. Bioinformatics 21(14), 3176–3178 (2005)
8. Furlanello, C., Marler, S., Menegon, S., Mancuso, S., Bertiato, G.: New WEBGIS technologies for geo-location of epidemiological data: an application for the surveillance of the risk of Lyme borreliosis disease. In: Proceedings of the 15th Int. Congress The new Frontiers of Arrhythmias, vol. 5(1), pp. 241–245 (March 2002)
9. Greulich, H., Chen, T.H., Feng, W., Jnne, P.A., Alvarez, J.V., Zappaterra, M., Bulmer, S.E., Frank, D.A., Hahn, W.C., Sellers, W.R., Meyerson, M.: Oncogenic transformation by inhibitor-sensitive and resistant EGFR mutants. PLoS Med. 2(11), 313 (2005)
10. Hynes, N.E., Lane, H.A.: ERBB receptors and cancer: the complexity of targeted inhibitors. Nat. Rev. Cancer 5(5), 341–354 (2005)
11. Lu, X.: A Web-GIS Based Urgent Medical Rescue CSCW System for SARS Disease Prevention. In: Li, M., Sun, X.-H., Deng, Q.-n., Ni, J. (eds.) GCC 2003. LNCS, vol. 3032, pp. 91–98. Springer, Heidelberg (2004)
12. von Mering, C., Jensen, L.J., Kuhn, M., Chaffron, S., Doerks, T., Krger, B., Snel, B., Bork, P.: STRING 7 - recent developments in the integration and prediction of protein interactions. Nucleic Acids Res. 35, D358–D362 (2007)
13. Mitchel, T.: Web mapping illustrated using open source GIS toolkits. O'Relly, pp. 120–127 (2005)
14. Parthiban, V., Gromiha, M.M., Schomburg, D.: CUPSAT: prediction of protein stability upon point mutations. Nucleic Acids Res. 34(Web Server issue), W239–242 (2006)
15. Shanzen, Y., Lizhu, Z., Chunxiao, X., Qilun, L., Yong, Z.: Semantic and Interoperable WebGis. In: International Conference on Web Information Systems Engineering (WISE 2001), Kyoto, Japan, December 3-6, vol. 2, pp. 42–48 (2001)
16. Sharma, S.V., Bell, D.W., Settleman, J., Haber, D.A.: Epidermal growth factor receptor mutations in lung cancer. Nature 7, 169–181 (2007)
17. Steitz, T.A.: Structural Biology: a mechanism for all polymerases. Nature 391, 231–232 (1998)
18. Yun, C., Titus, J., Boggon, L.Y., Woo, M., Greulich, H., Meyerson, M., Eck, M.J.: Structures of Lung Cancer-Derived EGFR Mutants and Inhibitor Complexes: Mechanism of Activation and Insights into Differential Inhibitor Sensitivity. Cancer Cell 11, 217–227 (2007)

Embedding Google Maps APIs into WebRatio for the Automatic Generation of Web GIS Applications

Giuseppe Avagliano[1], Sergio Di Martino[2], Filomena Ferrucci[1], Luca Paolino[1], Monica Sebillo[1], Genoveffa Tortora[1], and Giuliana Vitiello[1]

[1] Dipartimento di Matematica e Informatica (DMI), University of Salerno,
Via Ponte Don Melillo, I-84084 Fisciano (SA), Italy
{fferrucci,lpaolino,msebillo,tortora,gvitiello}@unisa.it
[2] University of Naples "Federico II", via Cinthia, I-80126 Napoli, Italy
dimartino@na.infn.it

Abstract. The success of WebML (Web Modeling Language) and of the supporting tool WebRatio for designing and generating data-intensive web applications suggested us to extend the approach to the Web GIS context. The proposal was based on Geo Server and Map Server, two standard, open solutions, to handle spatial data. In the present paper, we propose an alternative approach based on Google Maps, a freely available web mapping application provided by Google, which allows for the search and the visualization of geographic information. This solution is so diffuse that many Internet users are identifying it as a "standard" way for the presentation of geographical information. Moreover, Google Maps can be integrated into a Web application by exploiting Google Maps APIs. In the paper, we describe the proposed WebML-based visual language to design Web GIS applications and how it has been embedded into WebRatio to generate Web GIS applications by exploiting the potentialities of the Google Maps APIs. We also illustrate the architecture of the generated Web GIS applications together with the employed technologies and provide a sample example of design and generation of a Web GIS application.

1 Introduction

Recently, the increasing popularity of Rich Internet Applications has led to the development of Web applications, known as *Web GIS*, intended for dissemination and manipulation of spatial knowledge in specific domains. Compared with traditional Geographic Information Systems (GIS) [9, 10], Web GIS hold the potential to make geographic information available to a worldwide audience, allows Internet users to access GIS applications from their browsers without purchasing proprietary GIS software, and makes it possible to add GIS features to a wide range of network-based applications in business, government, education, etc. [12, 13].

However, software environments supporting the development of Web GIS applications are very hard to use and require specialized skill. Moreover, with the increased time-to-market pressure, it is no longer possible to deal with low-level issues, and create Web GIS applications from scratch. Thus, there is a growing need

M. Sebillo, G. Vitiello, and G. Schaefer (Eds.): VISUAL 2008, LNCS 5188, pp. 259–270, 2008.

for tools that allow developers to rapidly create this kind of Web applications and to rapidly modify them to meet the ever-changing business needs.

For traditional Web applications, many solutions are currently available. Among them, the *WebML* (*Web Modeling Language*) approach is based on a high-level, formal visual language specifically conceived to design "data-intensive" Web applications [1, 2, 13], which rely mainly on data management and movements. The approach is effectively supported by a CASE tool, *WebRatio* [1], which allows for the automatic generation of Web applications starting from a *WebML* visual specification. The approach has been successfully applied in many different contexts, from industry to academia, with positive results.

The observation that Web GIS can be considered as a particular class of data-intensive Web applications, being mainly devoted to handle (spatial) information, suggested us to extend the *WebML* methodology to deal with the Web GIS context. Indeed, the aim is to realize a visual environment that allows designers to model the required Web GIS application by spatially manipulating visual metaphors and then translates such visual schema into a self-contained application.

Preliminary results were introduced in [3-6]. The main contribution was a set of customized visual elements (*units*) specifically conceived to model relevant interaction and navigation operations, which usually characterize Web GIS (known as web mapping functionality). Moreover, the *WebRatio* CASE tool was enhanced to generate Web GIS applications by implementing the introduced units for the GIS-specific functionalities, and by exploiting the two standard open solutions Geo Server and Map Server for spatial data management.

In the present paper, we propose an alternative *WebML*-based approach to the generation of Web GIS applications. The new solution leverages on *Google Maps* [8] (*GM* in the following), a freely available Web mapping application, which allows the search and the visualization of geographic information, by seamlessly interacting with a smart user interface. Some of the most important advantages coming from the adoption of *GM* are the constant update of the geographical data (e.g.: roads), the appealing user interfaces, which exploits the most advanced Web technologies and counterbalances typical browser limitations, such as the use of a contextual right click, the use of the scroll wheel to zoom in/out, etc… Another advantage derives from the diffusion of *GM*. In fact, its widespread adoption by Internet users has made *GM* a *de facto* standard for presenting geographic information. This spreading is also due to the Google licensing policy which allows developers to freely integrate *GM* in their applications, through the *Google Maps APIs* [7]. These *API*s provide a number of utilities for manipulating maps and adding content to them through a variety of services.

These observations motivate the approach we propose in this paper. The *WebML*-based visual language has been enriched by introducing new customized units devoted to handle properties and components specific of *GM*, such as the *GM* viewer and the map type, which can be map, satellite or hybrid. Moreover, by exploiting potentialities of *GM APIs* the new customized units have been embedded into *WebRatio*, thus obtaining the automatic generation of the modeled Web GIS application.

The remainder of the paper is organized as follows. Section 2 recalls the main concepts of *WebML* and *GM*. In Section 3 we describe the proposed modeling concepts and notations which extend *WebML* for Web GIS applications. The architecture of the WebGIS applications generated by the *WebRatio* tool and the employed technologies are described in Section 4. In Section 5 we present an example of use of the proposed approach, while some final remarks conclude the paper.

2 Preliminaries

The solution we propose exploits both the expressiveness of *WebML* in describing a development process and the exhaustiveness of the *GM API*s in realizing a WebGIS application. By integrating such capability within a homogeneous framework, we have extended *WebRatio*, obtaining a CASE tool able to provide developers with a visual environment to design and implement Web GIS applications. In the following we recall some basic notions of *WebML* and *GM APIs* useful to describe the extension we propose.

2.1 *WebML*: A Modeling Language for Web Applications

WebML is a modeling language suited to support users in designing data-intensive Web applications [2], by providing them with a set of visual notations to model the content structure, data, and navigational aspects of a Web application.

The design of a Web application is based on two orthogonal perspectives: data and navigation. The former is usually described by means of the well-known ER data model, which characterizes all the relevant entities and relationships. The latter is specified by the *WebML Hypertext Model*, which describes how the previously defined contents should be arranged and provided within the Web application. In particular, the application structure and appearance are defined through four main levels of abstraction: *site views*, *areas*, *pages* and *units*. Moreover, semantics and syntax of notations allowed *WebML* creators to propose a specifically conceived commercial CASE tool, *WebRatio*, which automatically generates a J2EE-based Web application starting from the corresponding design models, thus effectively reducing time and costs of software development.

In [3-6] we proposed an extension of both *WebML* and *WebRatio* meant to support designers in generating Web-based GIS applications, which deal with complex, expensive and voluminous geographic data. In particular, recognizing Web GIS applications as a special case of data-intensive Web systems suggested us to employ the *WebML* approach to define a development process which took into account the complex nature of geographic data. Moreover, we introduced new *ad hoc* units, capable to model functionality, which characterizes those applications.

2.2 Google Maps and Its APIs

Google Maps [8] is a freely available web mapping application provided by Google that allows the search and visualization of geographic information. It provides both geo-referenced satellite images and vector maps, enhanced by a number of additional informative layers, related to points of interest such as restaurants, monuments and hotels. Moreover, *GM* can generate driving directions between any pair of locations.

The integration of many state-of-the-art web technologies, like Ajax, within the *GM* user interface has led to a smart and easy interaction, which involves the drag&drop mechanism for moving/selecting items, the scroll wheel to zoom in/out, the drag of a map for panning, and the double click to zoom and center the viewpoint. It is worth noting that all these metaphors are widely accepted for traditional desktop applications, but their diffusion in Web applications is very limited, due to well-known technical difficulties. Finally, in order to improve the reactivity of the application to the user panning, *GM* splits a map into a set of tiles that are smartly pre-cached.

From a development point of view, Google licensing policy allows web application developers to integrate *GM* in their sites, through the *GM APIs* [7]. These *API*s provide a set of utilities for manipulating maps and adding content to them through a variety of services.

3 Extending *WebML* for Web GIS Applications

Exploiting *WebML* for modeling Web GIS applications first implies the extension of the underlying visual language by defining new visual metaphors. In fact, besides the common interactions users may perform when navigating and executing a web application, it is necessary to take into account actions users may perform when interacting with a map, such as zooming, panning and spatial selection. Those actions are peculiar and characterize Web GIS applications.

The *WebML* extension we propose provides a new set of units, specifically conceived to model users' interaction with a map browser, where basic functionality for navigating and searching within a *GM* viewer is supplied.

In the following, we describe those units and illustrate their main properties.

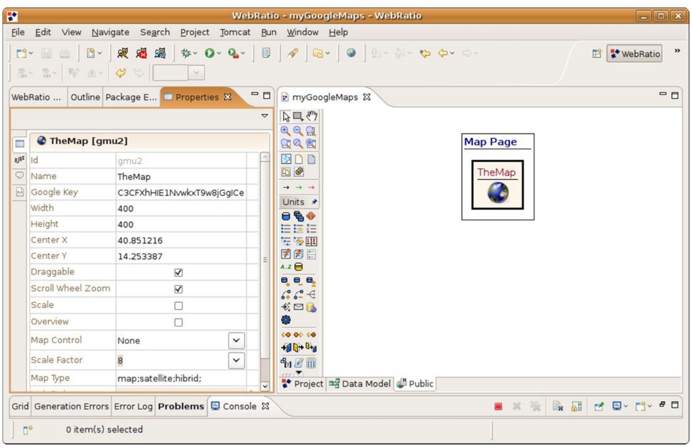


Fig. 1. The Icon we choose to represent a *MultiMap* Unit within WebRatio

The *MultiMap* Unit

Basically, the *MultiMap* Unit is a unit meant to represent the *GM* viewer, i.e. a graphic component within a Web browser able to render geographical data arranged

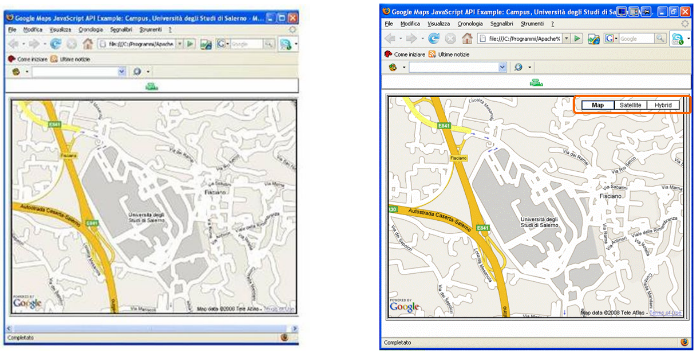


Fig. 2. How a MultiMap (Left), a MapControl (Right), appear within a Web browser

in layers. Fig. 1 illustrates the icon of a *MultiMap* Unit instance, namely *TheMap*, while its representation resulting in a Web browser is shown in Fig. 3 (Left).

This unit is characterized by the following properties:

- ***Name:*** The name chosen by the WebGIS designer for the *MultiMap* Unit instance;
  - **Type:** STRING;
  - **Mandatory:** YES;
- ***CenterX:*** The X coordinate of central point of the map;
  - **Type:** INT;
  - **Mandatory:** YES;
- ***CenterY:*** The Y coordinate of central point of the map;
  - **Type:** INT;
  - **Mandatory:** YES;
- ***Width:*** The width of the map control within the web page, in pixels;
  - **Type:** INT;
  - **Mandatory:** YES;
- ***Height:*** The height of the map control within the web page, in pixels;
  - **Type:** INT;
  - **Mandatory:** YES;
- ***ZoomFactor:*** The zoom level used the first time the map is rendered. It can be set in a range of values, from 0 (World View) to 16 (Street View).
  - **Type:** INT in [0..16];
  - **Mandatory:** NO;
- ***Scale:*** A parameter the developer can use to show a Scale Unit onto the map;
  - **Type:** Boolean;
  - **Mandatory:** NO; **Default** = OFF
- ***MapControl:*** A parameter the designer can use to add a control onto the map to zoom in, zoom out, move and re-center the initial map. Its representation in the web browser is shown in Fig. 3 (Right).

  - **Type:** Enumeration with the following values: *OFF* (Default), *Small*, *Large*;
  - **Mandatory:** NO;
- ***MapType:*** A parameter the designer can use to add a selection of the map type, i.e. it can be used to set the kind of map view, such as Satellite, Terrain, Streets, and so on. The default configuration results in a set of three buttons, as shown in Fig. 3 (Left).
  - **Type:** STRING. The various types must be concatenated by means of a semicolon (e.g.: Satellite; Streets;).
  - **Mandatory:** NO;
- ***Overview:*** A parameter the designer can use to add an overview map within the main *GM* viewer. It is meant to orient users by showing the location of the current view in the context of a larger area. Its representation in the web browser is shown in Fig. 3 (Right).
  - **Type:** Boolean;
  - **Mandatory:** NO; **Default** = OFF
- ***Scroll Wheel Zoom:*** A parameter the developer can use to enable or less the Zoom via mouse scroll wheel;
  - **Type:** Boolean;
  - **Mandatory:** NO; **Default** = OFF
- ***Google Key:*** The licensing key associated with the domain. It can be freely available, once activated a Google account.
  - **Type:** STRING;
  - **Mandatory:** YES;

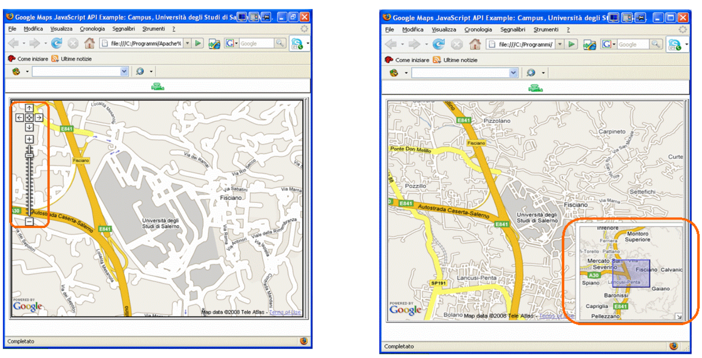


Fig. 3. How a MapType (Left) and an Overview (Right) appear within a Web browser

The *Marker* Unit

This Unit is meant to highlight onto the map one or more locations, starting from their real address. To this aim, it generates a web form, allowing the user to input the address. Once submitted, if such address exists, it is evidenced onto the map via a pinpoint. Thus, starting from a toponym, the *Marker* Unit allows converting it into a geographic reference and locates it on the map.

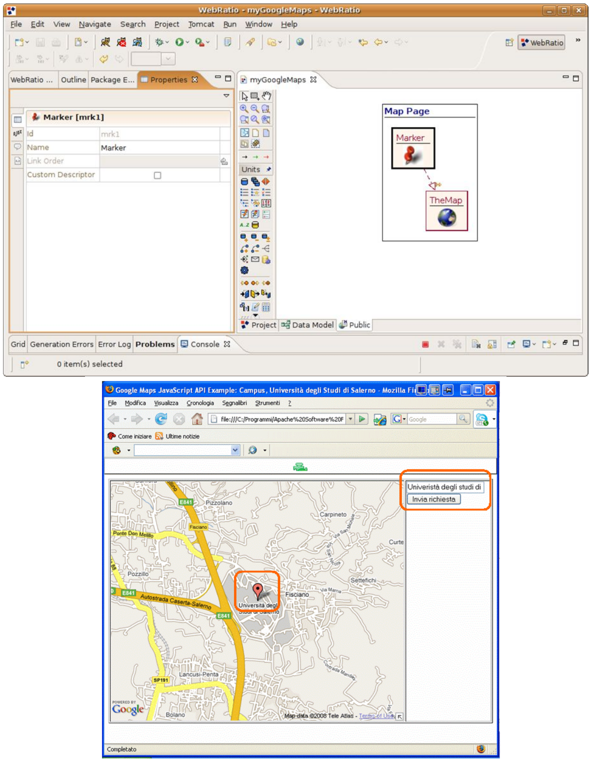


Fig. 4. An instance of a Marker Unit (Top) and its representation in a Web browser (Bottom)

In *WebML*, this unit requires a web form which allows the user to enter the address of the point of interest that may be possibly stored in the database. Fig. 4 (Top) shows how an instance of the *Marker* unit is specified within WebRatio, while the visual representation of the generated HTML code is depicted in Fig. 4 (bottom). Please note both the Form the user can exploit to specify the address (in the upper right corner), both the marker highlighting it onto the map.

The GetDirection Unit

The *GetDirection* Unit (see Fig. 5(Top)) allows us to calculate a route path between two addresses, and to highlight it onto the map. To this aim, it generates a web form to allow the user to input the two addresses required for the path. In order to improve its flexibility, we have contemplated also the possibility that the designer can set one of the two points. For instance, a developer of a Hotel web site could specify as destination point the lodge address, to generate a web page like "Insert your address to reach us".

The *GetDirection* Unit is characterized by the following parameters:

- *Name:* The name chosen by the WebGIS designer for the *GetDirection* Unit instance;
  - **Type:** STRING;
  - **Mandatory:** YES;

- ***Departure:*** The address of the starting point.
  - o **Type:** STRING;
  - o **Mandatory:** NO;
- ***Arrival:*** The address of the destination point.
  - o **Type:** STRING;
  - o **Mandatory:** NO;
- ***Unit:*** It specifies whether distances are shown in the imperial or in the metric scale unit. If not specified, distances will be indicated according to the system adopted by the starting country.
  - o **Type:** Enumeration with the following values: *KM*, *Miles*;
  - o **Mandatory:** NO;
- ***Avoid Highways:*** It specifies whether considering toll roads in the route calculation or not;
  - o **Type:** Boolean;
  - o **Mandatory:** NO;

The calculated path is overlapped to the map, as depicted in Fig. 5 (Bottom), while the set of turn-by-turn indications to follow is reported in a separated DIV of the web page.

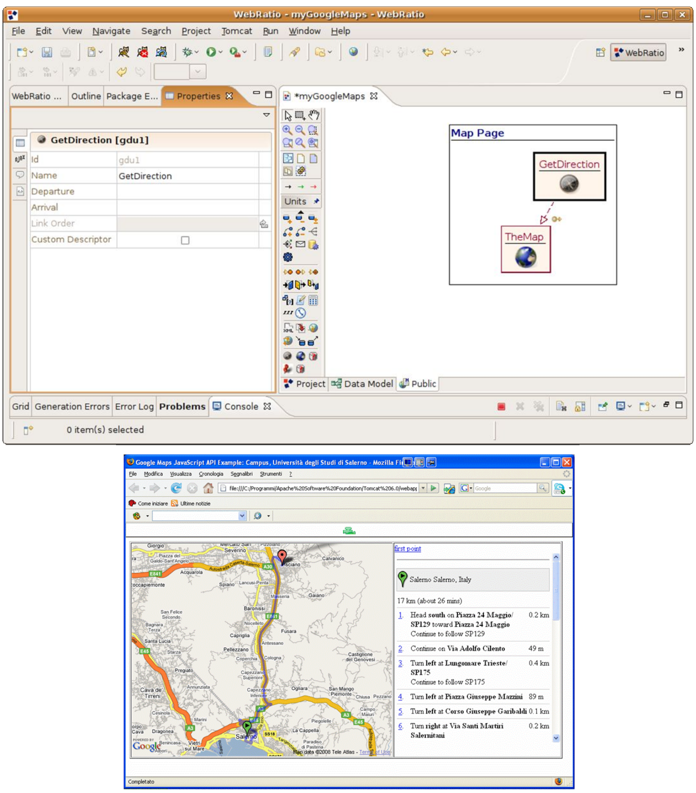


Fig. 5. The GetDirection Unit (Top) and its representation in a Web browser (Bottom)

4 The Generated Web GIS Applications: The Architecture and the Underlying Technologies

Once basic WebML units necessary for developing Web GIS applications have been designed, they have been implemented in WebRatio to support designers in the automatic code generation and delivering phases of Web applications. In particular, the tool generates the server side code, which handles and solves requests submitted via the Internet trough the Web browser.

Fig. 6 depicts the employed technologies and the architecture of a Web GIS application generated through *WebRatio*. The application is based on a three-tiered architecture and on a set of protocols and open technologies, such as Apache/Tomcat, Ajax and *GM APIs*, which guarantee the integration and interoperability between server modules and the client browser. Basically, the *WebML* units are translated into JSP pages and/or Servlet modules and successively mounted as a Tomcat/Apache web application. When users connect to a Web GIS, an automatic request is forwarded to the *GM* Server in order to receive maps to be displayed. This specific task will be automatically performed by the JavaScript code integrated within the downloaded web application.

As for navigation and searches, users' actions are translated into Ajax requests and then sent to the *GM* servers. They select data satisfying the request and sent back client HTML and JavaScript code, as well as new raster data to update the displayed map.

Finally, in order to manage markers, a database is included on the Web GIS server side which stores both coordinates and description associated with them.

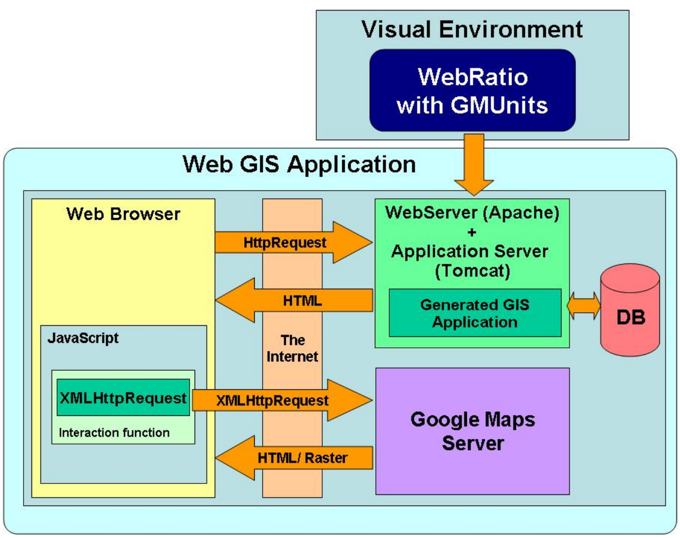


Fig. 6. WebRatio visual environment and the architecture of a generated Web GIS application

5 An Example

In order to prove the effectiveness of the proposed solution, in this Section we describe the design accomplished to realize a Map feature of our department Web site. In particular, we illustrate the *WebML*-based schema which contains both units provided by original *WebRatio* version and new units implemented through the extended version.

As for the *Hypertext & Web Mapping* design, the specific section of the web site is made up of two pages, as shown in Fig. 7:

- *A Home page* that contains the *External Link* instance of the *Index* unit meant to specify the external link to the other pages of the web site. Among them, one is pointing to the *Map Page*.
- *A* Map *Page* containing the *TheMap* instance of the *MultiMap* Unit, the *Marker* instance of the *Marker Unit* and, finally, *GetDirection1* instance of the *GetDirection Unit*.

Once generated, the Web GIS application provides the panning and zooming functionality, the map control buttons, the button for the visualization of the map type and the overview of the map. Moreover, users may insert and fix markers on the map through the *Insert Marker* link. Once a marker is attached, it is possible to click on it and visualize a dynamic form, as shown in Fig. 8.

Finally, through the *GetDirections* button, it is possible to calculate the route path connecting any arbitrary address (the city of Salerno in our example) and the campus of the University of the Salerno or vice versa. The result of this operation is visualized as both a polyline on the map and a list of turn-by-turn indications.

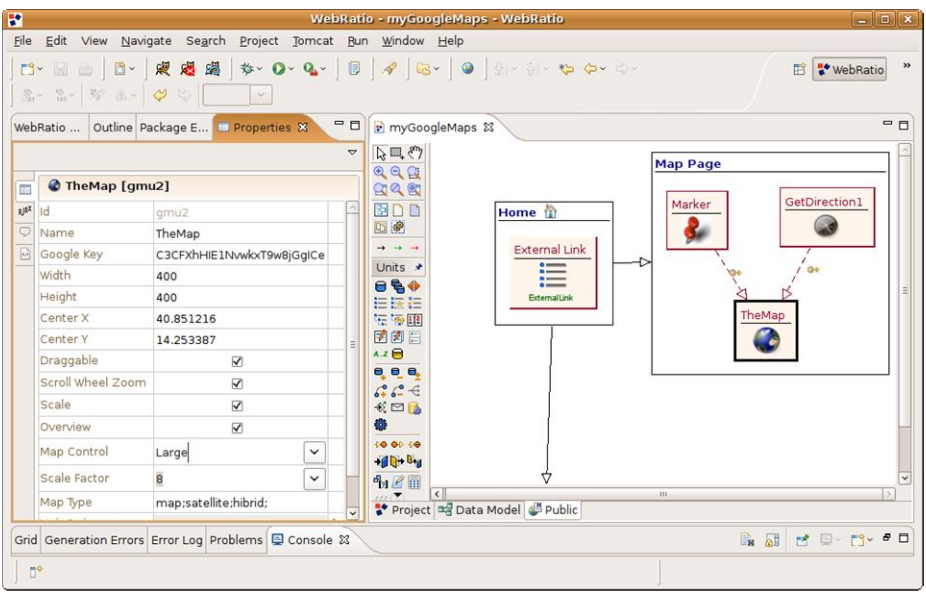


Fig. 7. The WebML-based schema for the proposed example

Fig. 8. The UI of the generated WebGIS application

6 Conclusions

In the present paper, we described an approach for designing and generating Web GIS applications that synergically combines the advantages of *WebML*/*WebRatio* and *GM*. In particular, we defined some visual notations to deal with web mapping functionalities and we embedded them into *WebRatio* by exploiting the potentialities of the *GM APIs*. The effectiveness of the proposed approach has been shown by exemplifying the modeling and construction of a Web GIS application.

As future work we plan to extend this solution by allowing developers both to select the required mapping service during the automatic generation task and to embed users own data to visualize. Indeed, it is worth to noting that currently other mapping services, such as Yahoo Maps [14], Microsoft Live Search Maps [11], are available over the Web, which share the same *GM* advantages. On the other side, it could be very useful for designer to show their own geographical data onto the map.

Acknowledgements

The authors wish to thank Dott. Giacomo Petillo for his support in the development of the solution proposed in this paper.

References

1. Ceri, S., Fraternali, P., Bongio, A., Brambilla, M., Comai, S., Matera, M.: Designing Data-Intensive Web Applications. Morgan-Kaufmann Publishers, San Francisco (2002)
2. Ceri, S., Fraternali, P., Bongio, A.: Web Modeling Language (WebML): a Modeling Language for Designing Web Sites. Computer Networks. The International Journal of Computer and Telecommunications Networking 33(1-6), 137–157 (2000)

3. Di Martino, S., Ferrucci, F., Paolino, L., Sebillo, M., Tortora, G., Vitiello, G., Avagliano, G.: Towards the Automatic Generation of Web GIS. In: Proc. of the 15th International Symposium on Advances in Geographic Information Systems (ACMGIS 2007), Seattle USA, pp. 7–9 (November 2007)
4. Di Martino, S., Ferrucci, F., Paolino, L., Sebillo, M., Vitiello, G., Avagliano, G.: A WebML-based visual language for the development of Web GIS Applications. In: IEEE Proceedings of the 2007 Symposium on Visual Languages and Human-Centric Computing (VL/HCC 2007) IEEE Press. Coeur d'Alene, Idaho, USA, pp. 23–27, 209–212 (September 2007)
5. Di Martino, S., Ferrucci, F., Ginige, A., Paolino, L., Sebillo, M., Vitiello, G., Avagliano, G.: A method to develop Web GIS. In: Proceedings of the International Workshop on Visual Languages and Computing (VLC 2007), San Francisco, California, USA, pp. 6–8, 229–234 (September 2007)
6. Di Martino, S., Ferrucci, F., Paolino, L., Sebillo, M., Vitiello, G., Avagliano, G.: A WebML-based Approach for the Development of Web GIS Applications. In: Benatallah, B., Casati, F., Georgakopoulos, D., Bartolini, C., Sadiq, W., Godart, C. (eds.) WISE 2007. LNCS, vol. 4831, pp. 385–397. Springer, Heidelberg (2007)
7. Google Maps APIs website, `http://code.google.com/apis/maps/`
8. Google Maps website, `http://maps.google.com`
9. Longley, P.A., Goodchild, M.F., Maguire, D.J., Rhind, D.W.: Geographic Information Systems and Science, p. 472. Wiley, Chichester (2001)
10. Longley, P.A., Goodchild, M.F., Maguire, D.J., Rhind, D.W.: Geographical Information System Principles, 2nd edn. Wiley, Chichester (2001)
11. Microsoft Maps Live, `http://maps.live.com`
12. Peng, Z., Tsou, M.: Internet GIS – Distributed GIServices for the Internet and wireless networks. Wiley, Chichester (2003)
13. WebML site, `http://www.Webml.org`
14. Yahoo Maps, `http://maps.yahoo.com/om`

DINDOW: Towards an Interaction Based on Spatio-temporal Memory

Jesús Ibáñez, Oscar Serrano, and David García

Department of Information and Communication Technologies,
Pompeu Fabra University, Barcelona, Spain
{jesus.ibanez,oscar.serrano,david.garcian}@upf.edu

Abstract. This paper describes the design and development of a novel system which allows non experienced users to easily interact with digital information and other users. In particular, the system allows the user to receive, handle and send media elements (pictures, text messages, videos). Instead of employing the file system to interact with information, the user interface promotes a kind of interaction which relies on spatial and temporal memory, which we believe to be more adequate.

1 Motivation

The work described in this paper has been carried out in the framework of the EC funded IST project ICING (Innovative CIties of the Next Generation) which, among other things, explores new ways of communication and interaction. In this context, needs for communication mechanisms and social awareness were found in several communities (elderly people, women association, families, etc.) through user studies. As a result of these studies, three main objectives were defined for the design of a system which fulfills the users' needs:

- The system should allow the user to visualise multimedia information received from the community members and corresponding to different media (text messages, pictures, and videos) in a flexible manner.
- The system should ease the local interaction with the multimedia elements.
- And finally, the system should promote interaction and social awareness among the community members.

An additional transversal requisite was that the system must be easily usable by non experienced users and even by technology reluctant users. These objectives were quite general. Thus, we had several open issues to research. We reviewed literature in related areas while designing the first prototype of the system. The design also evolved by taking into account the results from workshops conducted for gathering, from the users, feedback about the interfaces. As a result, the system presented in this paper, DINDOW (DIgital wiNDOW), was designed and developed.

In particular, the current version of DINDOW allows the user to receive, handle and send media elements (pictures, text messages, videos) in a very simple

M. Sebillo, G. Vitiello, and G. Schaefer (Eds.): VISUAL 2008, LNCS 5188, pp. 271–278, 2008.

way. Instead of employing the file system to interact with information, the user interface promotes a kind of interaction which relies on spatial and temporal memory, which we believe to be more adequate for our users.

The next subsections begins with a review of related work. Following this we describe the design and development of the system, including its architecture, user interface and interaction design. Finally we provide the conclusions and future work.

2 Related Work

Awareness systems, according to Markopoulos et al. [10], can be defined as systems whose purpose is to help connected individuals or groups to maintain a peripheral awareness of the activities and the situation of each other. The area of awareness system is a flourishing field of research, and interesting systems have been proposed, in the last years, for both workplaces [6][9] and social/family life [8][11][12][5]. In this sense, DINDOW is an awareness system which on the one hand can be used as a peripheral display (for instance it can be used as a notice board displaying pictures, or it can even subtly show details about the situation of the contacts, as the current weather in their locations) and on the other hand can be actively employed (for instance to manage the user's personal collection of multimedia elements or to send information to contacts).

It is well known that managing disparate data through traditional hierarchical storage and access interfaces is frustrating for users [3], specially for non experienced ones. As a consequence, different approaches and metaphors have been proposed to replace the desktop metaphor and its related hierarchical file system. A specially interesting alternative was proposed by Fertig et al. in Lifestreams [7], which uses a timeline as the major organizational metaphor for managing file spaces. It provides a single time-oriented stream of electronic information, and supports searching, filtering and summarization. Rekimoto extended this unidimensional idea to a bi-dimensional interface in TimeScape [14], which combines the spatial information management of the desktop metaphor and time travelling. Our system, DINDOW, extends these ideas by allowing time-travelling in any subregion of the user interface, and adding the ability to formulate natural and intuitive spatio-temporal queries.

3 Design and Development

In this section we describe the design and development of DINDOW. We first outline its overall architecture. Following this, we present the user interface and the interaction of the system.

3.1 Architecture

The overall architecture of the system is shown in figure 1. The communication manager is responsible for receiving and sending information. It consists of a

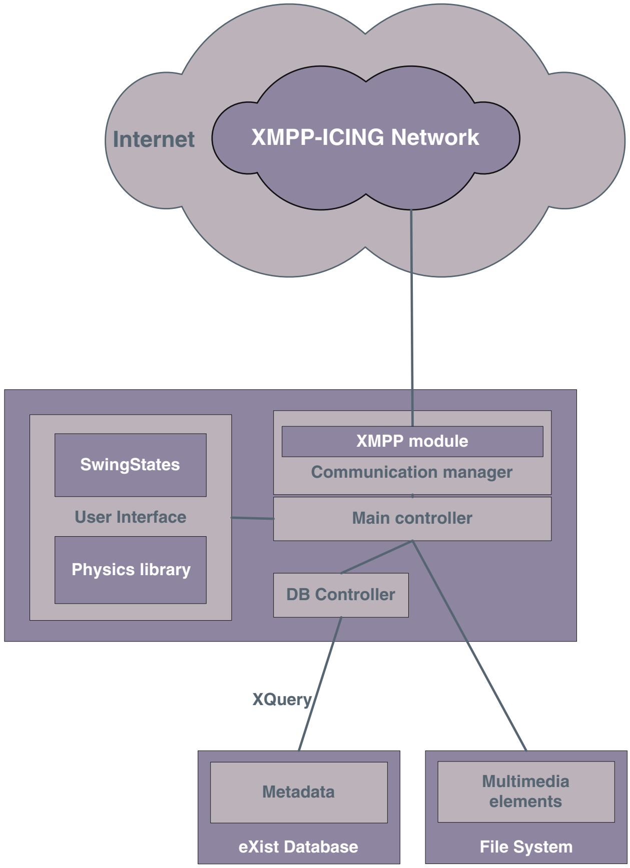


Fig. 1. Overall architecture

series of modules implementing different communication modalities. The current version contains a module implementing communication over the XMPP protocol.

The information stored by the system can be classified into two categories: multimedia elements and metadata. The multimedia elements are the images and videos received by the system from other users. These elements are automatically stored in the file system. The user does not need to know where these elements are physically located nor how they are named. The metadata are semantic annotations about the multimedia elements. These annotations include, for instance, the date when the element was received, its width and height, its location on the screen, its location on the file system, etc. The metadata are stored in a database. In particular, in the current version the metadata are stored in eXist [13], an open source native XML database. Thus, this metadata is accessed through queries programmed in the XQuery language [16].

The user interface incorporates novel interaction mechanisms, programmed with SwingStates. The user interface also employs a physics library to simulate the behaviour of group movement. The combination of both the novel interaction mechanisms implemented in the user interface and the metadata stored in the database, allows the user to naturally and intuitively interact with the multimedia elements.

3.2 User Interface and Interaction

As stated by Beaudouin-Lafon [2], the only way to significantly improve user interfaces is to shift the research focus from designing interfaces to designing interaction. This requires, among other things, powerful interaction models, beyond the usual ones. Moreover, as we were prototyping following an iterative process to design the system, we required advanced software libraries which ease the rapid development of new kinds of interaction. The typical libraries (like Java Swing) based on a set of well known graphical widgets are not appropriate, as they are oriented to create new user interfaces, not new interaction mechanisms. After surveying the state of the art, we finally decided to employ SwingStates [1]. SwingStates is a library that adds state machines to the Java Swing user interface toolkit. Unlike traditional approaches, which use callbacks or listeners to define interaction, state machines provide a powerful control structure and localize all of the interaction code in one place.

As shown in figure 2, the user interface is composed of three main regions (A, B and C). **Region A** is a scrollable space where the multimedia elements (pictures, videos, texts) are automatically added when they are received by the system. In particular, when a new element arrives, it appears on the right of this area, while the other elements in this space shifts to the left accordingly. Thus, all the elements in region A are temporally ordered. An element X was received earlier than an element Y if X is visually located to the left of Y. The user can scroll through the elements of this space. Thus, he can access any received element.

Fig. 2. User interface

When a new element is received, not only does the element appear on the right of region A, but also the user is warned by a particular sound. Moreover, the new element is enhanced with a colourful frame which identifies it as a new incoming element.

The user can group elements in folders but always keeping the temporal order. Thus, the folders are created in a simple way. The user only decides the extreme elements of the folder, that is, the oldest and newest element, and the folder is automatically created. Folders also appears temporally ordered on region A.

When several elements arrive together in the same message, they are automatically associated as a group. Then, these elements will behave as a group. For instance, the user can move the whole group by just moving one of the elements. The smooth movement of the group is simulated by employing the proper algorithms from a physics programming library.

Note that pictures, videos and text messages are clearly differentiated by the way they are represented. Pictures are displayed with the Polaroid aesthetic, while videos are presented with a film look, and text messages are shown in coloured PostIts such that the colour identifies the sender (that is, there is a particular colour per each sender).

Region B is an area which can be seen as a kind of advanced notice board. The user can move multimedia elements from region A to this area. When that happens, the elements which are moved to region B still remain visible in region A, although in ghost mode (with a degree of transparency). Elements moved from A to B, appear bigger in B than in A. However, relative size among elements is respected. That is, an element X will appear bigger than an element Y in region B if X as well appears bigger in region A.

The elements in region B will keep visible until the user decides to remove them. In this region, the multimedia elements can be freely moved, rotated and scaled. Moreover, video elements can be played and text annotation can be added to pictures.

Region B has a kind of spatio-temporal memory. For each element, its state-changes (location, size, rotation angle, time of change,...) are annotated in the database. Thus, by employing this memory it is possible to time-travel in this region. By travelling to the past, this region evolves by showing previous states of the region in an inverse time order. Elements appear and disappear, at a configurable rate, at the locations they were located at previous times, giving the user the impression that the region is travelling to past. Moreover, it is possible to time-travel in a subregion of region B. The user can mark a particular subregion by drawing its border on the screen. As a result, a hole appears in the just marked subregion. Then the user can time travel only in this subregion (see figure 3). Thus, he can access, for instance, a picture he remembers that was located at the top left corner by marking this corner and time-travelling to past. Time-travels to the future (from past times) are also allowed. Both, the direction of the time-travel (forward or backward) and its speed are shown through the animation of the needles of a clock which is shown while the time-travel takes place.

Fig. 3. Time-travel in a subregion

Natural spatial queries can be formulated in region B. To start the query, the user marks a subregion by drawing it on the screen. Then he can specify search criteria in an intuitive way. For instance, he can specify the sender of the element he is searching by dragging the picture of the sender from region C to the just marked subregion. Or he can can specify the approximate size of the element being searched, by just drawing it. Thus, queries are constructed by using simple gestures. In the mentioned example, the system will find and show elements which were located in the marked subregion and fulfill the specified search criteria (sender and approximate size).

Region C is a space that contains the pictures of the user contacts. On the one hand, these pictures can be used as clues to filter (for instance to look for elements sent by a particular contact). On the other hand these pictures are augmented with contextual information about the user. In particular they are augmented with a coloured frame showing the current weather at the user location. Finally, these pictures can also be employed to send multimedia elements to a contact. In particular, this is accomplished by dragging the element to be sent up to the picture of the contact.

4 Conclusions and Future Work

This paper has presented DINDOW, a novel system which allows non experienced users to easily interact with digital information and other users. Instead of employing the file system to interact with information, DINDOW promotes a kind of interaction which relies on spatial and temporal memory, which we believe to be more adequate for our users.

Future work includes the usability evaluation of the user interface. Moreover, the system will be evaluated as an awareness system. This evaluation will be based on both the ABC (Affective Benefits in Communication) questionnaire [15] and the IPO-SPQ (IPO Social Presence Questionnaire) [4].

Acknowledgments

This work has been partially funded by the European Union IST program through the project "ICING: Intelligent Cities for the Next Generation".

References

1. Appert, C., Beaudouin-Lafon, M.: Swingstates: adding state machines to the swing toolkit. In: UIST 2006: Proceedings of the 19th annual ACM symposium on User interface software and technology, pp. 319–322. ACM Press, New York (2006)
2. Beaudouin-Lafon, M.: Designing interaction, not interfaces. In: AVI 2004: Proceedings of the working conference on Advanced visual interfaces, pp. 15–22. ACM Press, New York (2004)
3. Cook, T.: It's 10 o'clock: do you know where your data are? Technology Review 98(1), 48–53 (1995)
4. de Greef, P., IJsselsteijn, W.A.: Social presence in a home tele-application. CyberPsychology and Behavior 4, 307–315 (2001)
5. Dey, A.K., de Guzman, E.: From awareness to connectedness: the design and deployment of presence displays. In: CHI 2006: Proceedings of the SIGCHI conference on Human Factors in computing systems, pp. 899–908. ACM Press, New York (2006)
6. Dourish, P., Bly, S.: Portholes: supporting awareness in a distributed work group. In: CHI 1992: Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 541–547. ACM Press, New York (1992)
7. Fertig, S., Freeman, E., Gelernter, D.: Lifestreams: an alternative to the desktop metaphor. In: CHI 1996: Conference companion on Human factors in computing systems, pp. 410–411. ACM Press, New York (1996)
8. Hindus, D., Mainwaring, S.D., Leduc, N., Hagstrom, A.E., Bayley, O.: Casablanca: designing social communication devices for the home. In: CHI 2001: Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 325–332. ACM Press, New York (2001)
9. Jancke, G., Grudin, J., Gupta, A.: Presenting to local and remote audiences: design and use of the telep system. In: CHI 2000: Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 384–391. ACM Press, New York (2000)
10. Markopoulos, P., de Ruyter, B., Mackay, W.E.: Awareness systems: known results, theory, concepts and future challenges. In: CHI 2005: CHI 2005 extended abstracts on Human factors in computing systems, pp. 2128–2129. ACM Press, New York (2005)
11. Markopoulos, P., IJsselsteijn, W., Huijnen, C., Romijn, O., Philopoulos, A.: Being There - Concepts, Effects and Measurements of User Presence in Synthetic Environments. In: Riva, G., Davide, F., IJsselsteijn, W.A. (eds.) Supporting social presence through asynchronous awareness systems, pp. 261–278. IOS Press, Amsterdam (2003)
12. Markopoulos, P., Romero, N., van Baren, J., IJsselsteijn, W., de Ruyter, B., Farshchian, B.: Keeping in touch with the family: home and away with the astra awareness system. In: CHI 2004: CHI 2004 extended abstracts on Human factors in computing systems, pp. 1351–1354. ACM Press, New York (2004)
13. Meier, W.: exist: An open source native xml database. In: Chaudhri, A.B., Jeckle, M., Rahm, E., Unland, R. (eds.) NODe-WS 2002. LNCS, vol. 2593. Springer, Heidelberg (2003)

14. Rekimoto, J.: Time-machine computing: A time-centric approach for the information environment. In: ACM Symposium on User Interface Software and Technology, pp. 45–54 (1999)
15. van Baren, J., IJsselsteijn, W.A., Romero, N., Markopoulos, P., de Ruyter, B.: Affective benefits in communication: The development and field-testing of a new questionnaire measure. In: PRESENCE 2003, 6th Annual International Workshop on Presence, Aalborg, Denmark (October 2003)
16. Walmsley, P.: XQuery. O'Reilly Media, Inc., Sebastopol (2007)

Perception Model for People with Visual Impairments

Pradipta Biswas, Tevfik Metin Sezgin, and Peter Robinson

Computer Laboratory, 15 JJ Thomson Avenue, Cambridge CB3 0FD,
University of Cambridge, United Kingdom
{pb400,mts33,pr}@cl.cam.ac.uk

Abstract. Scientists from many different disciplines (including physiology, psychology, and engineering) have worked on modelling visual perception. However this field has been less extensively studied in the context of computer science, as most existing perception models work only for very specific domains such as menu searching or icon searching tasks. We are developing a perception model that works for any application. It takes a list of mouse events, a sequence of bitmap images of an interface and locations of different objects in the interface as input, and produces a sequence of eye-movements as output. We have identified a set of features to differentiate among different screen objects and using those features, our model has reproduced the results of previous experiments on visual perception in the context of HCI. It can also simulate the effects of different visual impairments on interaction. In this paper we discuss the design, implementation and two pilot studies to demonstrate the model.

1 Introduction

Usability evaluation is an important step for successful design of any product. However user trials are often expensive and time consuming. Additionally for users with special needs, it is particularly difficult to get a representative population for a user trial. These difficulties with user trials led us to design a simulator that can model human computer interactions for people with a wide range of physical abilities and skills. In this paper we describe a particular component of this simulator - the visual perception model.

Computer Scientists have studied theories of perception extensively for graphics and, more recently, for Human-Computer Interaction (HCI). A good interface should contain unambiguous control objects (like buttons, menus, icons etc.) that are easily distinguishable from each other and reduce visual search time. In this work, we have identified a set of features to differentiate among different screen objects and we have used this set of features to reproduce the results of previous experiments on visual perception in the context of HCI. We have developed a prototype model of human visual perception for interaction with computer. It can also simulate the effects of different visual impairments on interaction. Unlike previous works, our model not only shows how a computer interface is perceived to a visually impaired person, but also it can simulate the dynamics of interactions with a computer.

M. Sebillo, G. Vitiello, and G. Schaefer (Eds.): VISUAL 2008, LNCS 5188, pp. 279–290, 2008.

2 Related Work

How do we see? This question has been addressed in many ways over the years. The Gestalt psychologists in early 19th century pioneered an interpretation of the processing mechanisms for sensory information [8]. Later the Gestalt principle gave birth to the top-down or constructivist theories of visual perception. According to this theory, the processing of sensory information is governed by our existing knowledge and expectations. On the other hand, bottom-up theorists suggest that perception occurs by automatic and direct processing of stimuli [8]. Considering both approaches, recent models of visual perception incorporate both top-down and bottom-up mechanisms [14]. This is also reflected in recent experimental results in neurophysiology [12, 17].

Knowledge about theories of perception has helped researchers to develop computational models of visual perception. Marr's model of perception is the pioneer in this field [14] and most of the other models follow its organization. However it was never been implemented in a practical system [18]. In recent years, a plethora of models have been developed (e.g. ACRONYM, PARVO, CAMERA etc. [18]), which have also been implemented in computer systems. The working principles of these models are based on the general framework proposed in the analysis-by-synthesis model of Neisser [14] and mainly consist of the following three steps:

1. **Feature extraction:** As the name suggests, in this step the image is analysed to extract different features such as colour, edge, shape, curvature etc. This step mimics neural processing at the V1 region of brain.
2. **Perceptual grouping:** The extracted features are grouped together mainly based on different heuristics or rules (e.g. the proximity and containment rule in the CAMERA system, rules of collinearity, parallelism and terminations in the ACRONYM system [18]). Similar type of perceptual grouping occurs in V2 and V3 regions of the brain.
3. **Object recognition:** The grouped features are compared to known objects and the closest match is chosen as the output.

In these three steps, the first step models the bottom-up theory of attention while the last two steps are guided by top-down theories. All of these models aim to recognize objects from a background picture and some of them have proved successful at recognizing simple objects (like mechanical instruments). However they have not demonstrated such good performance at recognizing arbitrary objects [18]. These early models do not operate at a detailed neurological level. Itti and Koch [10] present a review of some computational models, which try to explain vision at the neurological level. Itti's pure bottom-up model [10] even worked in some natural environments, but most of these models are used to explain the underlying phenomena of vision (mainly the bottom-up theories) rather than prediction.

In the field of Human Computer Interaction, the EPIC [11] and ACT-R [1] cognitive architectures have been used to develop perception models for menu searching and icon searching tasks. Both the EPIC and ACT-R models [4, 9] are used to explain the results of Nielsen's experiment on searching menu items [15] and found that users search through a menu list in both systematic and random ways. The ACT-R model has also been used to find out the characteristics of a good icon in the context of an icon-searching task [6, 7]. However the cognitive architectures emphasize modeling human cognition and so the perception and motor modules in these systems are not as well developed as the reminder of the system. The working principles of the perception models in EPIC and ACT-R/PM are simpler than the earlier general-purpose computational models of vision. These models do not use any image processing algorithms. The features of the target objects are manually fed into the system and they are manipulated by handcrafted rules in a rule-based system. As a result, these models do not scale well to general-purpose interaction tasks. Modelling of visual impairment is particularly difficult using these models. An object seems blurred in a continuous scale for different degrees of visual acuity loss and this continuous scale is hard to model using propositional clauses in ACT-R or EPIC. Shah et. al. [20] have proposed the use of image processing algorithms in a cognitive model, but they have not published any results about the predictive power of their model yet.

3 Design

We have developed a perception model as part of a simulator for HCI. The simulator takes a task definition and locations of different objects in an interface as input and then predicts the cursor trace, probable eye movements across the screen and task completion time, for different input device configurations (e.g. mouse or single switch scanning systems) and undertaken by persons with different levels of skill and physical disabilities. The architecture of the simulator is shown in Figure 1. It consists of the following three components:

The Application model represents the task currently undertaken by the user by breaking it up into a set of simple atomic tasks using the KLM model [5].
The Interface Model decides the type of input and output devices to be used by a particular user and sets parameters for an interface.
The User Model simulates the interaction patterns of users for undertaking a task analysed by the task model under the configuration set by the interface model. It uses the sequence of phases defined by the Model Human Processor [5]. The perception model simulates the visual perception of interface objects. The cognitive model determines an action to accomplish the current task. The motor-behaviour model predicts the completion time and possible interaction patterns for performing an

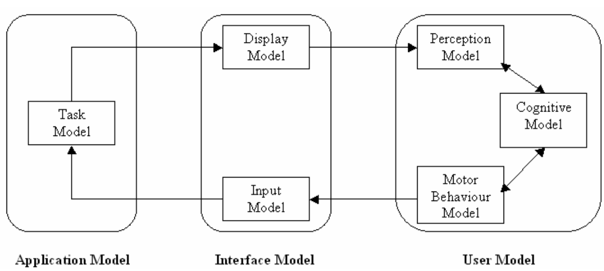


Fig. 1. Architecture of the Simulator

action. The details of the simulator and the cognitive and motor-behaviour models can be found in two separate papers [2, 3]. In the following sections we present the perception model in detail.

Modelling perception

Our perception model takes a list of mouse events, a sequence of bitmap images of an interface and locations of different objects in the interface as input, and produces a sequence of eye-movements as output. The model is controlled by four free parameters: distance of the user from the screen, foveal angle, parafoveal angle and periphery angle (Figure 2). The default values of these parameters are set according to the EPIC architecture [11]. The model can also be used to simulate the effect of different visual impairments.

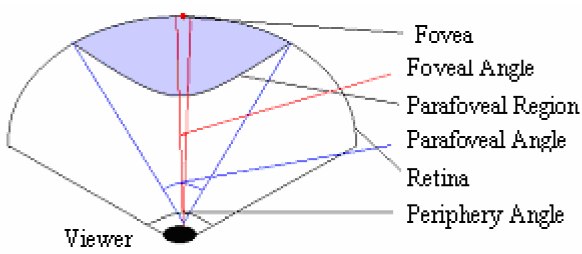


Fig. 2. Foveal, parafoveal and peripheral vision

We perceive something on a computer screen by focusing attention at a portion of the screen and then searching for the desired object within that area. If the target object is not found we look at other portions of the screen until the object is found or the whole screen is scanned. Our model simulates this process in three steps (Figure 3).

- Scanning the screen and decomposing it into primitive features
- Finding the probable points of attention fixation
- Deducing a trajectory of eye movement

The perception model represents a user's area of attention by defining a focus rectangle within a certain portion of the screen. The area of the focus rectangle is calculated from the distance of the user from the screen and the periphery angle (Figure 2). However it has already been found that we can see objects even which are out of attention (obviously with less accuracy [10]) and so the size of the focus rectangle varies with the number of probable targets in its vicinity. If the focus rectangle contains more than one probable target (whose locations are input to the system) then it shrinks in size to investigate each individual item. Similarly in a sparse area of the screen, the focus rectangle increases in size to reduce the number of attention shifts.

The model scans the whole screen by dividing it into several focus rectangles, one of which should contain the actual target. The probable points of attention fixation are calculated by evaluating the similarity of other focus rectangles to the one containing the target. We know which focus rectangle contains the target from the list of mouse events that was input to the system. The similarity is measured by decomposing each focus rectangle into a set of features (colour, edge, shape etc.) and then comparing the values of these features. The focus rectangles are aligned with respect to the objects within them.

Finally, the model shifts attention by combining three different strategies,

Nearest strategy [6,7]**:** At each instant, the model shifts attention to the nearest probable point of attention fixation from the current position.
Random Strategy: Attention randomly shifts to any probable point of fixation.
Cluster Strategy: The probable points of attention fixation are clustered according to their position and attention shifts to the cluster centre of one of these clusters.

We choose any one of these strategies probabilistically.

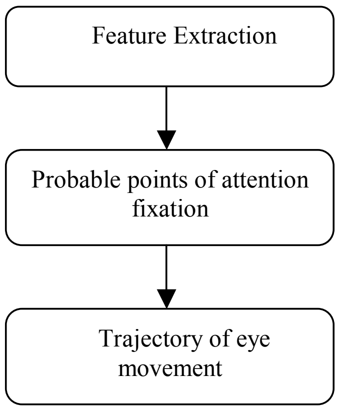
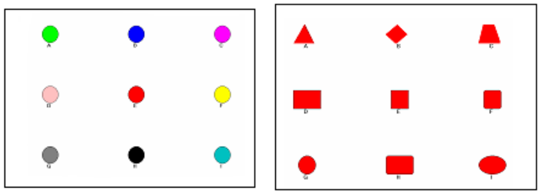


Fig. 3. Simulating visual perception

Pilot Studies

Study 1- Comparing performances for colour and shape recognition
In a computer screen, any target can be characterised by two properties – its colour and shape. In this study, we have investigated which of the features is easier to detect for impaired vision. We compared the reaction times people take to recognize a target from distractors of same colour and different shape and vice versa (Figure 4). Prior to each session, the participants were told about the target (e.g. a red circle) and then instructed to point to the target as soon as they could find it. We measured the reaction time between target display and recognition. We used nine types of targets of different colours and shapes. We recruited 10 participants (6 male, 4 female and average age 25.4), who did not have any colour-blindness and had no visual impairment that could impede their vision after correction. We simulated visual impairment by using translucent filters from the Inclusive Design Toolkit [22] and considered four conditions (normal vision, mild acuity loss, severe acuity loss and central vision loss). The reaction times are shown in Figures 5. As can be seen from the Figures 5, shape recognition takes more time in general and especially for severe acuity loss and central vision loss. With the filters (simulating vision loss), participants took more time to



a. b.

Fig. 4. a. Screen to test colour recognition b. Screen to test shape recognition

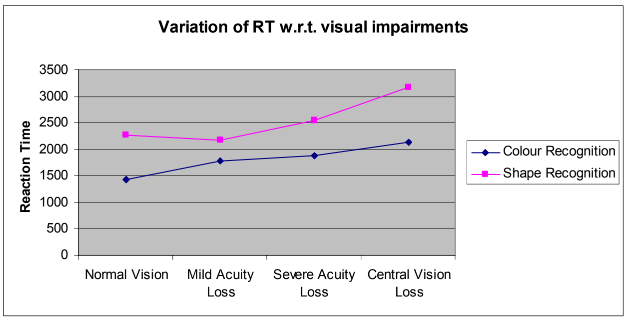


Fig. 5. Variations of reaction time (in msec) for different impairments

differentiate between target and distractors of same colour and different shapes than the other case and some of them even reported that they could not detect the corners of the shapes.

Guided by this study, we developed algorithms to simulate the process of colour and shape recognition. We used the colour histogram matching algorithm [16] to measure and compare the colours, the Sobel operator [16] for edge detection and the shape context algorithm [21] for shape measurement. We simulated severe acuity loss by a low pass Gaussian filter. We found that the colour histogram matching algorithm can work well even for a blurred screen; however the shape context matching algorithm does not. In particular, the edge detection algorithm, which is runs as a precursor to the shape context algorithm, fails to detect edges in a blurred screen. This is also consistent with the result we found in the study: with blurred vision people take more time to detect edges and thus to differentiate shapes from one another. However the colour information is not lost by blurring (as long as the colours contrast with background) and the colour-histogram matching algorithm finds it easier to recognize colour in the same way as the human participants. These results can be extended in future to predict reaction time from the colour histogram and shape context matching coefficients.

Study 2- Defining the best set of features to predict the probable points of fixation

The second study considered the best set of features to predict the probable points of fixation. For the pilot study, we assumed that, users' attention would fix on icons which were same as the target icons in a screen instead of other types of icons. For example, if the target was a PDF file then attention would mostly be fixed on most of

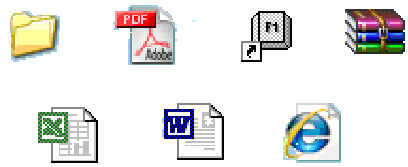


Fig. 6. Icons used in pilot study

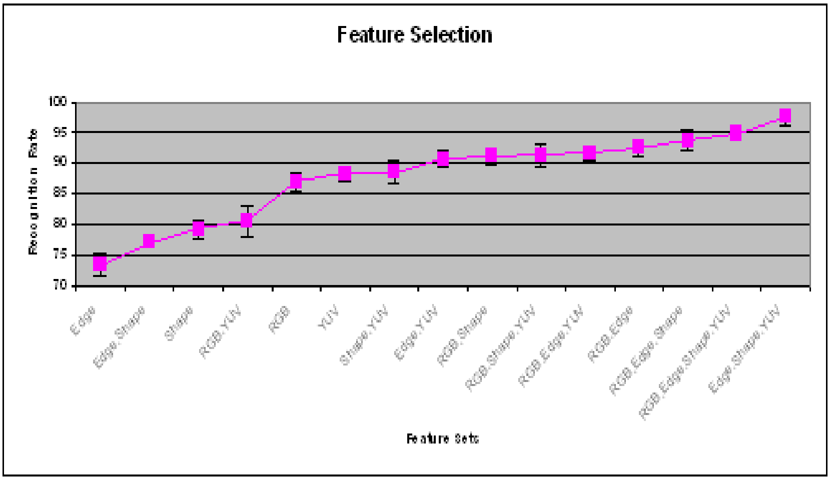


Fig. 7. Classifier performance for different feature sets

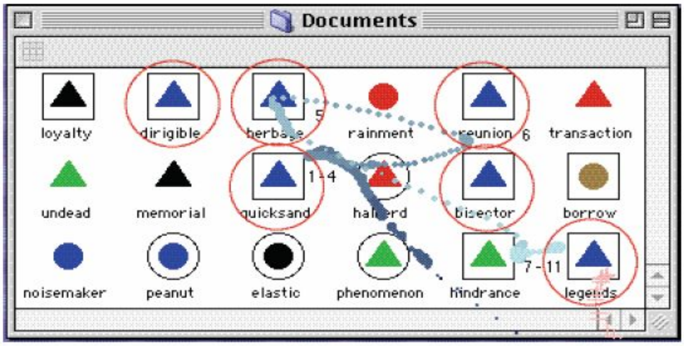
the PDF icons in the screen. We considered seven different types of icons (Figure 6) and looked for the best classification performance for different feature subsets. We used a backpropagation neural network as classifier. Figure 7 shows the classification performance for 15 different subsets of the Colour in RGB, Colour in YUV, shape and edge features. The error bars show the standard deviation for 30 runs for the best classifier parameters. As can be seen from Figure 7 the best results are obtained for the Colour (YUV), shape and edge features.

Validation

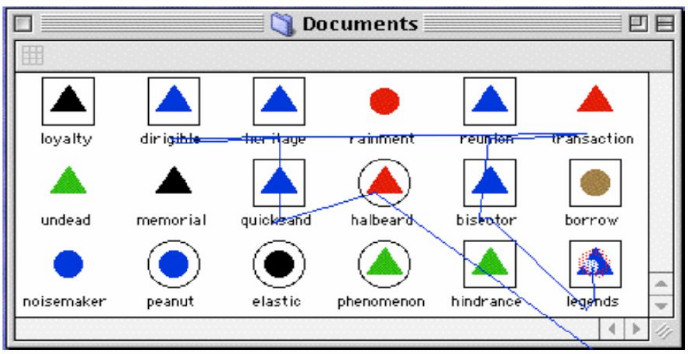
We do not yet have eye-tracking data of our own, so we compared the performance of our result to some previous eye-tracking data [6,7]. Figure 8 shows the actual eye-tracking data of a previous experiment (Figure 8a), prediction of the previous model (Figure 8b) and the prediction by our model (Figure 8c). It can be seen that our model successfully identified all the probable points of fixation.

Modelling visual impairment

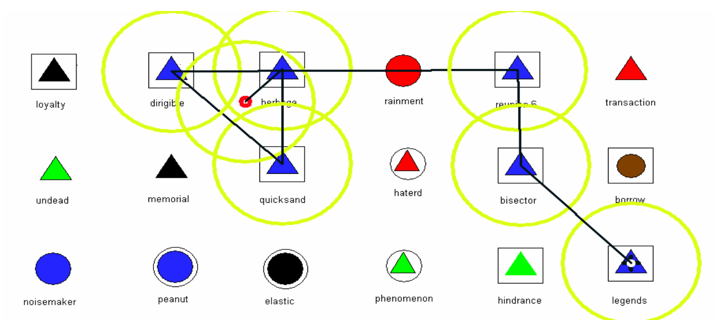
Our model can also simulate the effects of different visual impairments on interaction. To cover a wide range of visual impairments, we have modelled it in three different levels - in the first level the system simulates different diseases (currently Maccular Degeneration, Diabetic Retinopathy, Tunnel vision and Colour-Blindness). In the next



a. Eye tracking data [from 6, 7]



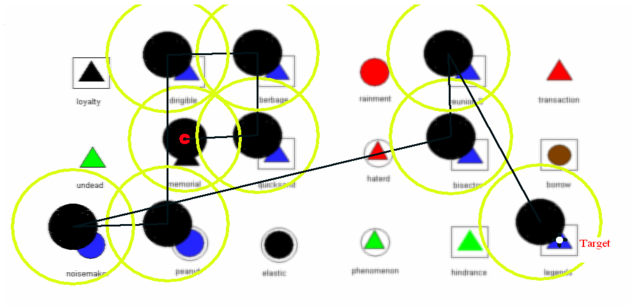
b. Eye movement prediction from previous model [6, 7]



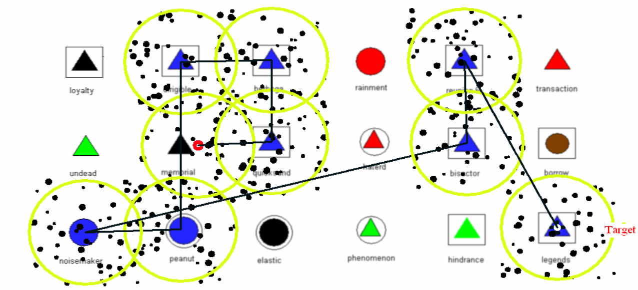
c. Eye movement prediction from our model

Fig. 8. Validating the model

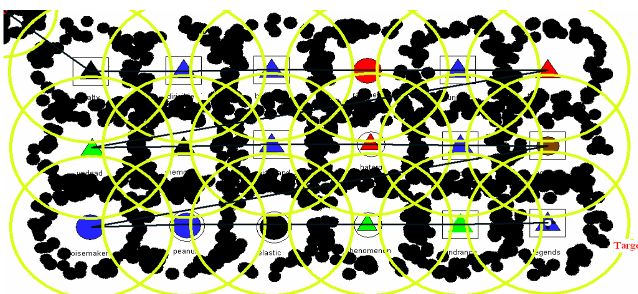
level it simulates the effect of change in different visual functions (e.g. Visual acuity, Contrast sensitivity, Visual field loss etc.). In the last level, it allows different image processing algorithms to be run (e.g. Filtering, Smoothing etc.) on input images to manually simulate the effect of a particular impairment. This approach also makes it easier to model the progress of impairment. The previous simulations on visual



a. Eye movement prediction for Maccular Degeneration



b. Eye movement prediction for Diabetic Retinopathy



c. Eye movement prediction for Tunnel Vision

Fig. 9. Eye movement prediction for different visual impairments

impairments model the progress of impairment by a single parameter [22, 23] or using a large number of parameters [24]. In our system, the progress of any impairment can be modelled either by a single parameter or by changing the values of different visual functions. For example, the extent of a particular case of Maccular Degeneration can be modeled either by a single scale or by using different scales for visual acuity and central visual field loss. Additionally, most previous work (like the Visual Simulator Project [23] or the Inclusive Toolkit [22]) simulates visual impairment on still images for a fixed position of eyes. Unlike those works, our model not only shows how a computer interface is perceived by a visually impaired person, but also it can simulate the dynamics of interactions with a computer. Figure 9 shows a few demonstrations of our simulator. In all these figures, the desired target is marked with the text 'Target'. The black line indicates the trajectory of eye movements through a series of intermediate points of attention fixation marked with rings.

Figure 9a shows a sequence of eye movements for Maccular Degeneration. As can been seen from the figure, the whole screen becomes blurred since the patient is using peripheral vision and black spots appear in the centre of point of fixation due to central field loss. In case of Diabetic Retinopathy (Figure 9b), some random black spots appear at the region of attention fixation due to damage of blood vessels inside the eyes. In both of these cases the number of points of fixation is greater than in normal vision (Figure 8) since patients need to investigate all blue targets due to blurring of the screen. For tunnel vision (Figure 9c), the patient cannot use any peripheral vision, so he can never see the screen as a whole and can only see a small portion of it. So all the targets need to be examined and eyes have to move systematically from left to right and top to bottom until it reaches the target.

4 Discussion

The first study proves (at least qualitatively) the credibility of colour histogram and shape context algorithms to model colour and shape recognition processes for both normal and impaired vision. The second study shows that they can also be used to identify icons besides primitive shapes (with more than 90% accuracy). Table 1 presents a comparative analysis of our model with the ACT-R/PM and EPIC models. Our model seems to be more accurate, scalable and easier to use than the existing models. However, in real life situations the model also produces some false positives because it fails to take account of the domain knowledge of users. This knowledge can be either application specific or application independent. There is no way to simulate application specific domain knowledge without knowing the application beforehand. However there are certain types of domain knowledge that are application independent that is they are true for almost all applications. For example, the appearance of a pop-up window immediately shifts attention in real life, however the model still looks for probable targets in the other parts of the screen. Similarly, when the target is a text box, users focus attention to the corresponding labels rather than other text boxes, which we do not yet model. There is also scope to model perceptual learning. Currently our neural network (used as a classifier) trains itself after each execution, but there is no way to remember a particular location, which would be used for the same purpose as before. For that purpose, we could consider some high level features like the caption of a widget, handle of the application etc. to remember the utility of a location for a certain application. These issues did not arise in previous works since they modelled very specific and simple domains [4, 6, 7, 9].

We are still undertaking further comparisons of our model with previous models. Currently we are working on an experiment to track users' gaze while they try to recognize a target from a real life application, rather than primitive shapes. We will simulate impairment using filters as our first study. Then we will try to predict the points of attention fixation and eye movements using our model. We are also working to predict the visual search time using the EMMA model [19], which will also help to evaluate the model.

Table 1. Comparative analysis of our model

	ACT-R/PM or EPIC models	Our Model	Advantages of our model
Storing Stimuli	Propositional Clauses	Spatial Array	Easy to use and Scalable
Extracting Features	Manually	Automatically using Image Processing algorithms	
Matching Features	Rules with binary outcome	Image processing algorithms that give the minimum squared error	More accurate
Modelling top down knowledge	Not relevant as applied to very specific domain.	Considers the type of target (e.g. button, icon, combo box etc.).	More detailed and practical
Shifting Attention	Systematic/ Random and Nearest strategy	Clustering/ Nearest /Random strategy	Not worse than previous, probably more accurate

5 Conclusions

In this paper we have presented a perception model that can be used to evaluate and compare the visual feedback provided by different computer interfaces. The model is part of a larger system that is used to evaluate interfaces with respect to a wide range of skills and physical abilities [2, 3]. Our perception model takes a list of mouse events, a sequence of bitmap images of an interface and locations of different objects in the interface as input, and produces a sequence of eye-movements as output. The model supports existing theories on visual perception and it can also explain the results of most of the experiments done on visual perception in the field of Human-Computer Interaction. The model can also simulate the effect of different visual impairments on interactions. Unlike previous work, our model not only shows how a computer interface is perceived to a visually impaired person, but it can also simulate the dynamics of interactions with a computer. Currently we are in the process of calibrating the model using an eye-tracker.

Acknowledgements

We would like to thank the Gates Cambridge Trust for funding this work. We also like to thank the students of University of Cambridge for taking part in our experiments

References

[1] Anderson, J.R., Lebiere, C.: The Atomic Components of Thought. Erlbaum, Hillsdale (1998)

[2] Biswas, P., Robinson, P.: Automatic Evaluation of Assistive Interfaces. In: Proc. of the ACM Intl. Conf. on Intelligent User Interfaces (IUI), pp. 247–256 (2008)

[3] Biswas, P., Robinson, P.: Simulation to Predict Performance of Assistive Interfaces. In: Proc. of the 9th Intl. ACM SIGACCESS Conf. on Computers & Accessibility (ASSETS 2007), pp. 827–828 (2007)

[4] Byrne, M.D.: ACT-R/PM & Menu Selection: Applying A Cognitive Architecture To HCI. International Journal of Human Computer Studies 55 (2001)
[5] Card, S., Moran, T., Newell, A.: The Psychology of Human-Computer Interaction. Lawrence Erlbaum Associates, Hillsdale (1983)
[6] Fleetwood, M.F., Byrne, M.D.: Modeling icon search in ACT-R/PM. Cognitive Systems Research 3(1), 25–33 (2002)
[7] Fleetwood, M.F., Byrne, M.D.: Modeling the Visual Search of Displays: A Revised ACT-R Model of Icon Search Based on Eye-Tracking Data. Human-Computer Interaction 21(2), 153–197 (2006)
[8] Hampson, P., Moris, P.: Understanding Cognition. Blackwell Publishers Ltd, Oxford (1996)
[9] Hornof, A. J. & Kieras, D. E., Cognitive Modeling Reveals Menu Search Is Both Random & Systematic. In: Proc. of the ACM/SIGCHI Conference on Human Factors in Computing Systems, pp. 107–115 (1997)
[10] Itti, L., Koch, C.: Computational Modelling of Visual Attention. Nature Reviews, Neuroscience 2, 1–10 (2001)
[11] Kieras, D., Meyer, D.E.: An Overview of The EPIC Architecture For Cognition & Performance With Application To Human-Computer Interaction. Human-Computer Interaction 14, 391–438 (1990)
[12] Luck, S.J., et al.: Neural Mechanisms of Spatial Selective Attention In Areas V1, V2, & V4 of Macaque Visual Cortex. Journal of Neurophysiology 77, 24–42 (1997)
[13] Marr, D.C.: Visual Information Processing: the structure & creation of visual representations. Philoso-phical Transactions of the Royal Society of London B 290, 199–218
[14] Neisser, U.: Cognition & Reality. Freeman, San Francisco (1976)
[15] Nilsen, E.L.: Perceptual-motor Control in Human-Computer Interaction (Technical Report No. 37), Ann Arbor, MI: The Cognitive Science & Machine Intelligence Laboratory, the Univ. of Michigan
[16] Nixon, M., Aguado, A.: Feature Extraction & Image Processing, 1st edn. Elsevier, Oxford (2002)
[17] Reynolds, J.H., Desimone, R.: The Role of Neural Mechanisms of Attention In Solving The Binding Problem. Neuron 24, 19–29, 111–145 (1999)
[18] Rosandich, R.G.: Intelligent Visual Inspection using artificial neural networks, 1st edn. Chapman & Hall, London (1997)
[19] Salvucci, D.D.: An integrated model of eye movements & visual encoding, Cognitive Systems Research (January 2001)
[20] Shah, K., et al.: Connecting a Cognitive Model to Dynamic Gaming Environments: Architectural & Image Processing Issues. In: Proc. of the 5th Intl. Conf. on Cognitive Modeling, pp. 189–194 (2003)
[21] Belongie, S., Malik, J., Puzicha, J.: Shape Matching & Object Recognition Using Shape Contexts. IEEE Transactions on Pattern Analysis & Machine Intelligence 24(24), 509–521 (2002)
[22] Inclusive Design Toolkit (accessed in March 27, 2008), `http://www-edc.eng.cam.ac.uk/betterdesign/downloads/visualsim.html`
[23] Vision Simulator (Accessed in March 27, 2008), `http://www.my-vision-simulator.com/`
[24] Visual Impairment Simulator (Accessed in February 27, 2008), `http://cita.rehab.uiuc.edu/software/vis/index.php`

An Analysis of Improvements for Voter Interfaces in Polling Station and Remote Electronic Voting Systems

Maider Huarte, Maria Madarieta, Iñaki Goirizelaia, and Juan José Unzilla

University of the Basque Country, School of Engineering
Alda. Urquijo s/n
48013 Bilbao
maider.huarte@ehu.es, maria.madarieta@ehu.es,
inaki.goirizelaia@ehu.es, juanjo.unzilla@ehu.es

Abstract. Due to the rapid growth of computer networks and advances in cryptography, Electronic Voting Systems are becoming a real option for voters. These systems must fulfil requirements such as accuracy, invulnerability, privacy, verifiability, convenience, flexibility and mobility, protected by the required security schemes. Those requirements guarantee that elections run on such systems are universal, equal, secret and free. Voter interfaces for those systems should be designed so that no eligible voter feels discriminated. Thus, they should be adapted even for disabled and illiterate voters. In this paper, we analyse different interfaces used in already developed and tested electronic voting systems, organizing them into polling station and remote system types. Then, we describe the interface of electronic voting systems and propose improvements to adapt them to all eligible voters. Afterwards, we perform the same analysis and corresponding proposals for remote electronic voting systems.

Keywords: Electronic Voting Systems, electronic interfaces, hardware for handicapped.

1 Introduction

Communication interfaces between humans and computers have been and continue to be in constant evolution. The generalized use of these machines, while intended for non-expert computer users, favours the development of ever easier to use, but ever more complex designs. As in any other electronics field, interface evolution can be classed into two categories; hardware and software.

The evolution of the hardware led to graphic colour screens, printers, speakers, earphones, card readers, etc. for data output; keyboards, special help keyboards, mouse, scanners, microphones, etc. for data input. Software evolution examples are user friendly windows interface Operating Systems and programmes, as well as multimedia integration (audio, video, etc.).

Most of the aforementioned hardware and software interfaces have been sufficient to significantly improve human-machine communication to the point of being a key factor in the success of the spread of New Information and Communication Technologies.

M. Sebillo, G. Vitiello, and G. Schaefer (Eds.): VISUAL 2008, LNCS 5188, pp. 291–301, 2008.

However, not all of them are designed with all users in mind, as the disabled, the elderly, and the illiterate require especially adapted interfaces for their individual needs.

2 Electronic Voting Systems

In the case of electronic voting system interfaces, each system uses its own interface composed of some of the hardware and software elements previously mentioned. It is accepted that democratic elections must be universal, equal, secret and free ([1]); as so, electronic voting systems must inherently fulfil requirements such as accuracy or integrity, invulnerability or eligibility, privacy, verifiability, convenience or accessibility, flexibility on ballot formats and mobility of voters ([2][3]).

So that the whole system can maintain these properties, the interfaces of these electronic voting systems must be the appropriate ones. Thus, depending on the architecture of each system, the interface requirements will differ. A distinction shall be made between two different types of electronic voting in order to analyse their requirements and propose new solutions. This division is due to the fact that voter-system interaction takes place in a controlled environment (polling station) or uncontrolled (remote [3]), which mostly affects the properties of invulnerability, privacy and mobility:

- Polling Station E-voting: the voter must physically go to a designated place where the system interface is located, and such a place would be controlled to guarantee the properties of invulnerability and privacy. On the other hand, mobility requirement is negatively affected. The use of technology can mean that more locations can be made available than in conventional systems, resulting in increased mobility of the Polling Station E-voting systems.
- Remote E-voting: these systems presently in use, favour maximum mobility (they can be used any time anywhere) but do not guarantee voter privacy, making it susceptible to coercion and vote buying and selling. Additionally, many do not offer secure voter validation, thus affecting the requirement of invulnerability. Despite the limitations of such systems, they are often used as a substitute or complement to traditional remote voting (i.e. postal voting).

  Although there have been pilot experiences with Remote E-voting by telephone and digital TV ([4]), the majority of the solutions lean toward the use of the Internet, owing to its widespread use (both geographically and socially).

In our opinion, all voting systems do not fulfil completely the universality, equality, secrecy and free properties. In fact, in recent years it has been indicated that one of the major problems with the protocols of e-voting and i-voting (Remote E-voting via internet), was the so-called digital gap ([5] [6]), where not all citizens have equal access to the new technologies, and the change to an electronic interface would result in an impediment to the right to vote for certain sectors of society (the elderly, the illiterate, or those on limited income). Today, however, the digital gap is closing thanks to the proliferation of all types of electronic equipment, and especially in the case of e-voting, may even be bridged with the appropriate social measures.

Nevertheless, there still exists a group of citizens for whom the necessary measures have not been taken, and for whom complete access to voting systems is not available: the disabled. Relatively profound changes to standard interfaces are required as the disabled presently must depend on assistance in order to vote, thereby negating their privacy.

Therefore, e-voting system interfaces, be it Polling Station E-voting Systems or Remote E-Voting Systems, should integrate all the properties equally for all voters.

3 Polling Station Electronic Voting System Interfaces

The interface itself can be composed of two different parts perfectly defined according to its purpose:

- Voter validation interface: e-voting systems could electronically guarantee the property of invulnerability. To that end, specific electronic tools need to be built into the validation interface to authenticate and recognize the right to vote of the possible voters. Data output could be a screen, where the software guides the user through each step toward validation. Data input would depend on how and on what basis the system validates the voter. Thus, if the validation is based on data voter knows, like a PIN, a keyboard and/or a mouse could be used; if it is based on something voter has, namely a Smart Card, CD, USB stick, etc. or in voter biometric parameters (fingerprint, manual signature, etc.) specific readers or ports could be used.
- Voting interface: this interface permits a validated voter to cast their vote, and they are found inside the voting booth in the polling station to guarantee the property of privacy. Already in use voting systems are based on paper, mechanical lever machines, punch cards, Optical Mark Recognition (OMR) and Direct Recording Electronic (DRE). It is stated that for a voting system to be considered e-voting, it must involve the use of electronic means in at least the casting of the vote ([1]) and as such, there ought to be an electronic voting interface. Thus, OMR and DRE systems are the ones to be considered electronic. On DRE based systems, the electronic voting interface can be a set of buttons or a touch screen. On OMR systems, the interface is a scanner and specific software for mark recognition.

With the aim of guaranteeing universal, equal, secret and free properties in these voting systems, we propose the following adaptations shown in table 1:

Table 1. Proposals for the adaptation of Polling Station Electronic Voting Systems

	Blind	Deaf	Physically Disabled
Electronic interface	Data output: audio Data input: audio mouse, Braille keyboard	Data output: screen Data input: virtual keyboard on touch screen, mouse	Data output: screen Data input: quadriplegic adapted mouse
Voting Booth	Audio via earphones	Adequate medium for written instructions	Sufficient room for wheelchair

Our proposal for adaptation is based on the use of functioning senses to overcome barriers. Thus, blind voters could receive necessary instructions in audio form via earphones to guarantee privacy, and then express their opinion by typing on a Braille keyboard and/or using an audio mouse. These same audio instructions need not be exclusively for the blind, but could also be used to assist illiterate voters, or even as an added support to those in the digital gap. Deaf voters require the least amount of adaptation as long as their vision is not impaired. However, care should be exercised that all audio instructions are also provided in writing, so as to avoid any possible discrimination. In the case of voters with physical disabilities, not all would require specially adapted equipment (those who have manual mobility). For those who may be quadriplegic, we propose the use of adapted mouse which operate by mouth, eyes or facial gestures ([7] [8]).

In summary, the adaptation of an electronic voting interface for all types of voters would entail the use of a touch screen and audio via earphones as data output; Virtual keyboard on touch screen, audio and quadriplegic adapted mouse as data input; all enclosed in a voting booth with ample space.

3.1 Example: Demotek

Demotek ([9]) is a Polling Station E-voting system, designed for closed list elections. Each ballot cast can be valid (a blank ballot or one of the available options, such as a particular political party or yes/no in referendums) or a spoiled ballot. As a result, the tallying method is quite simple, since each ballot cast may only have one value.

Demotek was designed to meet requirements as stipulated in the electoral law of the Basque Government, which establishes how the voters cast the vote. In summary, it is as follows:

- The voters are to identify themselves at the appropriate polling station.
- All the polling station officials are to verify the voter validity by checking the eligible voter list.
- The voter is to hand their ballot in an envelope to the president of the polling station.
- The president of the polling station is to announce that the voter is casting their vote, and then is to enter the vote into a transparent ballot box.
- The electoral authorities and representatives of political parties are to make note that the voter's envelope was introduced in the ballot box.

In this voting protocol, the voter needs to have their vote ready before identity validation. A prepared vote consists of entering an official ballot paper into an envelope, which along with all the official ballots is provided by the voting authority in order to cast valid votes. When a cast vote does not correspond to any officially recognized ballot it is considered to be spoilt. The voting authority sets up the voting booths with all the officially recognized ballots and envelopes so that the ballot may be prepared in absolute privacy.

The Demotek system, which follows the same protocol, incorporates a number of improvements to the privacy of vote preparation and tallying.

3.1.1 Privacy of Vote Preparation: Obligatory Special Booth

The voter who wishes to use the Demotek system, must first go into the voting booth before approaching the ballot box, invalidating family-voting (a form of coercion where the coercer may exert control over the voter from ballot preparation to ballot casting).The Demotek voting booth is equipped with a detection system, which monitors the booth to ensure that there is only one person at a time inside it. As with conventional voting booths, the Demotek ones must contain all the official ballots. In this case, these ballots are adapted for the Demotek vote tallying mechanism, as shown in Fig. 1.

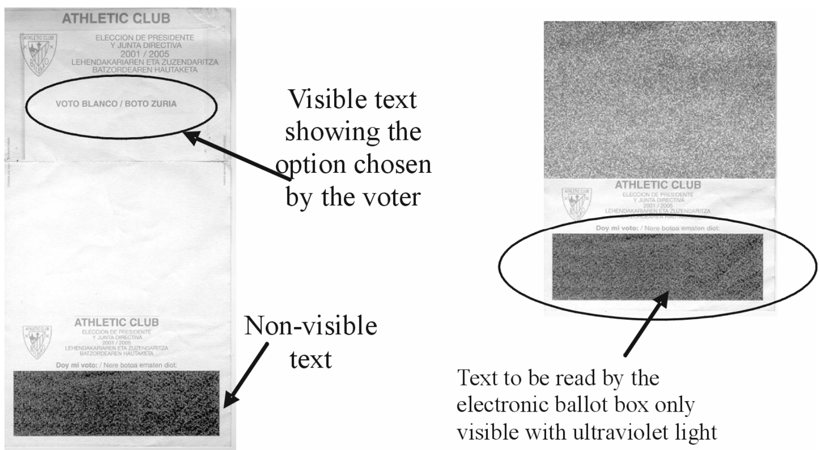


Fig. 1. Demotek ballots format, unfolded (left) and folded (right)

Demotek ballots are very similar to conventional ballots, except for two major differences. The first one is a non visible text that has codified the value of the vote in such a way that can only be read via a special ballot reader with ultraviolet light, which is used for the electronic tallying. The other difference is that the Demotek ballots do not need to be entered into envelopes as they can be folded along a fold-line and as a result the value of the ballot is hidden from view (right image at Fig. 1). Once folded, the codified section will be accessible to the Demotek ballot reader facilitating tallying.

In order to prevent sabotage to, or errors with the ballots the voting booth must contain a ballot-checking device, so that by inserting the codified section of the ballot, the voter could make sure that the vote to be counted is of the expected value (the one on visible text section). The voter is permitted to check as many ballots as they wish and additionally, the device should not have memory to record all the checked ballots.

Outlined below are necessary additional proposals given the Demotek vote preparation procedure the adaptation for all types of voters.

Voters without barriers. All the required steps to follow for vote preparation should be clearly indicated on a screen, making the system more convenient, and the change from paper ballots less traumatic. The vote preparation steps are as follows:

- Choose a paper ballot
- Insert the codified section into the slot of the ballot-checking device
- Check the readout on the screen
- Fold the ballot before exiting the voting booth

Illiterate voters. The steps to be taken for vote preparation should be given audibly, which is not incompatible with the steps for those without barriers. The screen could also be additional support to the audio, with illustrative pictograms, candidate photos, etc.

Blind voters. The audio system proposed for illiterate voters would be equally valid for blind voters, with a few additional instructions for total adaptation. Given that the voting booth would be an unfamiliar space to the blind voter, the audio instructions ought to include a description of the exact location of the various elements such as paper ballots, the ballot-checking device, as well as any other assistance necessary.

The visible text of the paper ballots could include a Braille version which should not be noticeable when folded. However, this is not sufficient guarantee for a blind voter that their ballot would be tallied correctly, as the Demotek ballots may be counted manually by checking the visually printed value and not that printed in Braille, since the tallying is performed by sighted people. Perhaps it would be easier and more economical to label the ballot trays in Braille and that the ballot-checking device to include a scanner and OCR software. The checking device would immediately corroborate that both the written and codified information are the same and then audibly confirm that to the voter.

Deaf voters. The adaptation for voters without barriers could be sufficient for deaf voters, as long as the audio instructions are supported by clear and graphic instructions on the screen.

Physically disabled voters. In the Demotek case the proposed adaptation for general Polling Station E-voting systems is insufficient for quadriplegic, since the Demotek paper ballots must be handled manually. One possible solution could be the incorporation of a mechanical arm ([10]) in the voting booth which would be controlled by the quadriplegic adapted mouse. Such an arm could physically pick up the paper ballot and insert it in the ballot-checking device. Then the device could also fold the ballot.

Finally, the disabled voter would only require assistance in taking their folded ballot from the voting booth to the president of the polling station with their right to privacy intact, as no-one else participated in the preparation of the vote.

3.1.2 Election Results: Electronic Ballot Box Tallying

The improvements of Demotek in vote tallying begin at the moment of vote casting, when the vote is entered into the ballot box. At the end of the election period, the ballot count is available immediately, and the data can be sent directly to a central counting location.

The voter exits the voting booth with their vote folded so that its value cannot be seen. The voter then approaches the president of the polling station to be validated as a legitimate voter. The vote is handed to the president of the polling station, who then enters it into a Demotek ballot box, which requires two steps.

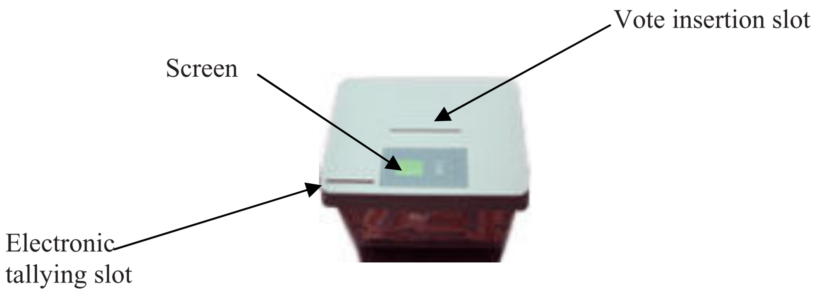


Fig. 2. Demotek ballot box

The Demotek ballot box is electronic, which is classified as an OMR ballot box. As can be seen in Fig. 2, there are two slots, one is an electronic reader, and the other is to physically enter the vote into the ballot box.

The president of the polling station enters the electronically readable part, into the electronic tallying slot. The system reads the value of the vote, while also checking that it is a Demotek ballot, and once confirmed the second vote insertion slot opens. The president of the polling station shall enter the ballot into this second slot. The software of the ballot box does two steps once it detects the paper ballot:

- The corresponding vote value register increases by one. That value, held in temporary memory when it was recorded by the electronic tallying slot, is erased, thus guaranteeing privacy.
- The number of cast votes increases by one, which is shown on the screen (Fig. 2), so that all polling station officials shall have confirmation that the vote has been tallied.

The sensors in the vote insertion slot can detect whether more than one ballot is entered or if a ballot is not entered fully, in which case the value read from slot one does not increase, nor does the number of cast votes.

The Demotek ballot box contains a simple interface, just a screen telling the president of the polling station and officials the number of cast ballots and instructions to follow. Given that this interface is only intended for election officials, it is considered that although it ought to be simple to operate to bridge the digital gap with most individuals, it is unnecessary to adapt it to the needs of all disabled groups.

4 Remote Electronic Voting Systems

As mentioned in Section 2, the Remote E-voting systems differ from Polling Station E-voting systems in that the voter-system interaction occurs in an uncontrolled environment any time anywhere, favouring mobility requirement. As it is impossible to ensure whether the voter is alone or not at the time of casting, invulnerability and privacy properties are the most seriously affected ones.

In the case of i-voting, the voter is connected to the system via Internet, requiring a PC adapted to their needs. In Estonia in 2007 ([11]), the first binding elections were held where i-voting was employed as another voting option for voters, and not only for citizens residing abroad or with mobility barriers. The problems with i-voting, derived from the lack of privacy, such as coercion and vote buying and selling, were solved by multi-casting ([12] [13]). Multi-casting means that a voter can cast more than one vote in a given election, but the only last one will be tallied. The idea is that if the voter is coerced into voting a given option, they could cast a true vote when they are free of the coercer. In the case of vote buying and selling, the transaction becomes meaningless, as the buyer is not absolutely sure whether the voter will cast another vote or not.

In our opinion, multi casting does not solve all the problems it is intended to solve. Perhaps it does solve the problem of vote buying and selling, but this is not so for all cases of coercion, in particular the so called family-voting. In family-voting, the coercer could exert absolute control over the voter, keeping the voter under 24 hour watch until the last day in order to oblige them to comply with their will.

This is the reason why we believe that the use of remote electronic voting should not depend on multi casting. Instead adapted interface design should be researched in order to resolve the problems that arise from having to vote in an uncontrolled environment.

For such interfaces, we propose (Table 2) the next requirements:

Table 2. I-voting system interface requirements approach

Interface requirements	System requirement fulfilled
Portable	Mobility
Personalizable	Invulnerability
Tamper resistant	Invulnerability Privacy
Cryptographic capability	Privacy
WAN (Internet) connection	Mobility
Autonomous*	Mobility Privacy

*Autonomous requirement is divided into three aspects: appropriate peripherals and software (for private interaction with all type of users), adequate power supply (battery and mains connection) and elements and/or mechanisms which permanently guarantee vote privacy (non-traceability of the vote-voter relationship, secure storage/elimination of critical data when needed).

In a classic scheme of remote electronic voting via anonymous channel, the voter must complete a validation phase and another of vote casting. In the validation phase, the system identifies the voter as eligible and accreditation is issued. This accreditation is generated so that nobody, not even the part of the system that issued it, will be able to link it to the voter. This is possible thanks to Chaum's blind signature cryptographic technique. In the ballot casting phase the system receives a vote with valid authorization, via an anonymous channel (unable to identify the voter).

Our proposal for a Remote E-voting interface takes into account this procedure in two steps, one for the voter validation and the other for vote casting. Taking that in mind and the requirements stated in table 2, our proposed interface is made up by three elements: a personalization card for the validation of the voter, a voting card for vote casting and a simple device (without processing capability). This device should be composed of input and output portable peripherals adapted to all types of voters, and the necessary ports which accept the aforementioned cards, and an Internet connection.

In the following paragraphs we explain how an interface of these characteristics could comply with the previously mentioned system requirements.

Portability. This requirement is satisfied by the fact that all three elements mentioned are portable, making the interface portable as well.

Personalizable. This requirement is met thanks to both cards. All eligible voters would have a Personalization card with all the necessary data for the system to ensure identification (i.e. electronic identity card, [10]). Therefore, the interface could be personalizable, and only its owner could operate with it. In addition, it should be able to cryptographically sign and perform encryption and decryption calculations.

All eligible voters would also have a voting card. This card, however, must not be personal, since it will guarantee the necessary anonymous channel for the vote cast. The system must be unable to identify which voting card was issued to each voter, obliging appropriate measures to be taken. The card would be programmed to execute the various stages of interaction with the system, as long as the device is operating with the appropriate personalization card. Thus, in order to guarantee privacy, the voter's first operation would be always to personalize the interface, by recording onto the voting card which personalization card to operate with. When communication with the system is over a public channel, the voter card would prepare and transmit the appropriate message to the personalization card to sign it, or it will request that a received message be decrypted. When communication with the system is over an anonymous channel, it will carry out the encryption and decryption by itself.

Tamper resistant. To comply with this requirement, both cards must also be tamper resistant. The personalization card must be tamper resistant so that neither the identification data nor the cryptographic keys can be altered in any way. Additionally, the voting card must too be tamper resistant, so that neither its programming nor the secret data (such as the cryptographic voting key for example), can be altered nor discovered. As an additional security measure, this card ought to be renewed every election.

Cryptographic capability. The capability to carry out cryptographic calculations is guaranteed with the personalization and voting cards described.

WAN connection capacity. This is achieved by the appropriate ports on the device, and the necessary software on the voting card, as the device itself does not have any processing capability.

Autonomy. The requirement of autonomy (private interaction peripherals, battery supply and security mechanisms) is met with an appropriate design of the device and suitable functionality of the voting card.

The device requires an output peripheral that shows data only to the voter (avoiding any other present individuals). The input peripheral would accept the data entered only assisted with the output peripheral (mouse). Thus, our design proposal for the device would be a screen incorporated into eyeglasses for sighted-voters, or an earphone for blind voters. Such device would positively identify the personalization card user-owner, and activate a detection system so that the whole process would abort should the voter attempt to remove it (perhaps to give it to a possible coercer or vote-buyer). Finally, the peripheral input could be a type of mouse or virtual keyboard, so that only the voter can see what is entering through it.

5 Conclusions

This article introduces an analysis of the situation of voter interfaces in electronic voting systems, as well as several proposals to improve them. Our analysis has been divided into two major blocks, according to the different conditions in which the interfaces are found: controlled and uncontrolled environments.

The first analysis was of electronic voting systems where the interaction between voter and system occurs in a controlled environment, that is to say, in a polling-booth. The result of our analysis was to propose a series of general approaches to adapt the interfaces to all types of voters, particularly keeping in mind voters with different disabilities. This would entail the use of a touch screen and audio via earphones as data output; Virtual keyboard on touch screen, audio and quadriplegic adapted mouse as data input; all enclosed in a voting booth with ample space. As a specific example of Polling Station E-voting systems, we have examined Demotek, used in the Basque Country. This led to new specific approaches, which mechanically handle and fold Demotek paper ballots.

The second analysis was of Remote E-voting systems, where there is no polling-booth, and the interaction between voter and system occurs in an uncontrolled environment any time anywhere. Due to the complexity of maintaining the invulnerability and privacy of the electoral process, these systems have been tested once in binding elections, using multi casting technique. Our proposal for Remote E-voting system interfaces avoids multi casting because of its weaknesses. It would be a screen incorporated into eyeglasses for sighted-voters, or an earphone for blind voters, with adapted mouse and virtual keyboard. All this, using leading technology with which Remote E-voting systems could reach its full potential making it possible to hold legally binding really universal elections for all type of voters.

References

1. Council of Europe: Legal, operational and technical standards for e-voting. Recommendation rec(2004)11 adopted by the committee of ministers of the council of Europe and explanatory memorandum. Council of Europe, Strasburg (2004)
2. Cranor, L.F., Cytron, R.K.: Sensus: A Security-Conscious Electronic Polling system for the Internet. In: Proceedings of the Hawaii International Conference on system Sciences, pp. 561–570. IEEE Computer Society, Los Alamitos (1997)

3. Puiggali, J., Morales-Rocha, V.: Remote Voting Schemes: A Comparative Analysis. In: Alkassar, A., Volkamer, M. (eds.) VOTE-ID 2007. LNCS, vol. 4896, pp. 1–15. Springer, Heidelberg (2007)
4. Electoral Comisión: The shape of elections to come: A strategic evaluation of the, electoral pilot schemes (2003), http://www.electoralcommission.org.uk/about-us/03pilotscheme.cfm
5. Madise, Ü., Vinkel, P., Maaten, E.: Internet Voting at the Elections of Local Government Councils on (October 2005), http://www.vvk.ee/english/report2006.pdf
6. Mäenpää, O.: E-Governance: Effects on civil society, transparency and democracy, http://www.iiasiisa.be/iias/aicor/rap_pdfs/Olli_Maenpaa.pdf
7. Surdilovic, T.: Fuzzy Mouse Cursor Control System For Computer Users with Spinal Cord Injuries. In: Thesis presented in Partial Fulfillment of Requirements for the Degree of Master of Science in the College of Arts and Sciences Georgia State University (2005)
8. Meng, L.F.: Applications of Computer Access Approach to Persons with Quadriplegics". In: Miesenberger, K., Klaus, J., Zagler, W., Burger, D. (eds.) ICCHP 2004. LNCS, vol. 3118, pp. 857–864. Springer, Heidelberg (2004)
9. Goirizelaia, I., Espinosa, K., Martín, J.L., Lázaro, J., Arias, J., Igarza, J.J.: An Electronic Secure Voting System Based on Automatic Paper Ballot Reading. In: Sanfeliu, A., Martínez Trinidad, J.F., Carrasco Ochoa, J.A. (eds.) CIARP 2004. LNCS, vol. 3287, pp. 470–477. Springer, Heidelberg (2004)
10. Leroux, C. et al.: Robot grasping of unknown objects, description and validation of the function with quadriplegic people. In: IEEE 10th International Conference on Rehabilitation Robotics, pp. 35–42 (2007)
11. Madise, Ü., Martens, T.: E-voting in Estonia 2005. The first practice of country wide binding Internet voting in the world. In: Krimmer, R. (Ed.) ELECTRONIC VOTING 2006, Proceedings of the 2nd International Workshop Coorganized by Council of Europe, ESF TED, IFIP WG 8.5 and E-Voting.CC, pp. 13–26 (2006)
12. Volkamer, M., Grimm, R.: Multiple Casts in Online Voting: Analyzing Chances. In: Krimmer, R. (Ed.) ELECTRONIC VOTING 2006, Proceedings of the 2nd International Workshop Co-organized by Council of Europe, ESF TED, IFIP WG 8.5 and E-Voting.CC, pp. 97–106 (2006)
13. Skagestein, G.: Are Vegard Haug, Nødtvedt E., Rossebø J.: How to create trust in electronic voting over an untrusted platform. In: Krimmer, R. (ed.) ELECTRONIC VOTING 2006, Proceedings of the 2nd International Workshop Co-organized by Council of Europe, ESF TED, IFIP WG 8.5 and E-Voting.CC, pp. 107–116 (2006)

Reasoning about Hand-Drawn Sketches: An Approach Based on Intelligent Software Agents

Giovanni Casella[1], Vincenzo Deufemia[2], Viviana Mascardi[1], Maurizio Martelli[1], and Genoveffa Tortora[2]

[1] Dipartimento di Informatica e Scienze dell'Informazione – Università di Genova
Via Dodecaneso 35, 16146, Genova, Italy
{casella,mascardi,martelli}@disi.unige.it

[2] Dipartimento di Matematica e Informatica – Università di Salerno
Via Ponte don Melillo, 84084 Fisciano (SA), Italy
{deufemia,tortora}@unisa.it

Abstract. Sketching is a powerful means to represent objects and reason on them. In this paper we describe an integrated environment, conceived as a multi agent system, that brings together sketch recognition functionalities and decision support facilities. In this environment, intelligent agents are exploited both for managing the process of recognition of the sketched objects, and for supporting users in solving decisional problems. We explain our approach and its potential by means of a running example taken from the domain of building's safety.

1 Introduction

Sketches play multiple roles [1]: they serve as an external memory to augment the limitation of human cognitive abilities, act as the medium that users use to communicate, and serve as the triggers that enable reasoning [2]. Humans see in a sketch more than a static arrangement of arbitrary symbols: they always consider the meaning underlying the sketch and its potential transformations. A sketch often represents the solution to a problem that the user has in mind. Of course, not every sketch is an admissible solution to that problem. In fact, the placement of sketched objects must respect a set of constraints that depend on the sketch's domain and semantics.

Recent years' experience suggests that a great improvement to hand-drawn sketch recognizers can be introduced by enhancing their level of "intelligence", i.e., enabling them to show an intelligent behavior to the user and to help him/her to find solutions to a problem [3]. A computer system might help the user by providing both recognition functionalities of hand-drawn symbols and domain specific knowledge for extracting the underlying sketch meaning and for supporting the user to reason about it. Users should interact with the drawing components in a natural and user-friendly way for obtaining real-time intelligent feedback from the system, and the system should propose innovative solutions to solve particular problems. An intelligent sketch system should embed a problem solving process involving specific domain knowledge, and should present feedback in

M. Sebillo, G. Vitiello, and G. Schaefer (Eds.): VISUAL 2008, LNCS 5188, pp. 302–314, 2008.

appropriate form, without distracting the user from the tasks in which he/she is engaged [4].

In this paper we present an approach based on agent technology for integrating domain-specific reasoning facilities with sketch recognition functionalities. The system architecture consists of the “Sketch Recognition” component, devoted to recognizing elements drawn by the user and to resolving interpretation ambiguities that may arise; the “User Reasoning Support” component, that checks that symbols drawn by the user respect a set of constraints, and, upon request, suggests a re-organization of them; and the “Interface” component, that provides the graphical interface between the user and the system.

To show how the system might be used in practice, we discuss its adoption in the building’s safety domain. Given a building/room plant and a customizable set of criteria to meet as input, our system might allow the user to draw objects that are relevant for safety aspects (fire extinguishers, tables, chairs, closets, windows, doors, lights), to check if the given safety criteria are met by the chosen placement of the objects, and, if not, to reason about how to meet them by re-positioning the objects, and finally to propose suggestions to the user.

The paper is organized as follows. Section 2 motivates this research and describes the case study that will be considered throughout the paper. Section 3 describes why intelligent software agents represent a suitable approach for sketch recognition and reasoning. Sections 4 and 5 describe the proposed agent-based system and its application to the case study. The related work is discussed in Section 6. Finally, conclusions and further research are discussed in Section 7.

2 Motivation and Case Study

In many situations, being able to sketch a diagram using an input device, and having the diagram components recognized in an automatic way by a software application, is a great advantage. In fact, this approach allows to save time and paper, to share the diagram with colleagues which are spread all over the world, to teach how a diagram should be correctly drawn, to archive and retrieve it in an electronic form. These advantages are well-known: the literature discusses many examples of software systems able to recognize hand-drawn sketches in very different domains [3,4,5]. However, there are situations where the user’s needs go far beyond the use of a software system just for recognizing symbols in the correct way. For example, consider an employee that needs to check that all the safety requirements imposed by his/her country’s law, are met by the building where he/she works.

The employee would be surely happy to use a software application that allows him/her to load the building plant in some format, and draw, upon the plant, tables, closets, fire extinguishers, and all the objects that may change location over time, as well as mark some doors as emergency exits. But he/she would be even happier, if the application would allow him/her

– To understand that a table too close to an emergency exit violates the safety constraints, or that the fire extinguishers located within a room, are either not enough, or not located in the most convenient way;
– To move the symbols representing furniture, lights, fire extinguishers, and so on, in order to find a setting that meets the safety constraints;

and would be able to propose, in a pro-active way, a re-arrangement of the objects that meet the safety constraints, in case of their violation.

A software system like this must integrate capabilities coming from three research domains:

– From the domain of automatic hand-drawn sketch recognition, the system must borrow the ability to recognize hand-drawn symbols, and to detect and resolve conflicts among their interpretation;
– From the domain of geometric modeling, it must borrow the ability to model physical objects and to reason about spatial relationships among them;
– Finally, from the domain of artificial intelligence, it must borrow the ability to act as a "pro-active" and "situation aware" expert system, supporting the user in finding violations of constraints, explaining why and where the constraints are violated, and proposing alternative solutions.

Provided that the rules for recognizing free-hand drawn symbols, for verifying the allowed spatial relations among them, and for reasoning on their semantics (determined by the constraints that they must respect), are customizable, such a system would prove useful in many disparate domains. For example, it might help a chemical engineer in reasoning on chemical reactions: the engineer might sketch a chemical as a molecular graph with atoms for nodes and bonds for edges, and the system would be able to recognize the characters that identify atoms and the lines that represent bonds (sketch recognition ability), check that spatial relations among symbols are met (geometric modeling capability), reason over its semantics (artificial intelligent capability).

The domain that we will consider for showing the potential of the proposed system is that of building's safety already introduced in the beginning of this section. Ensuring that a room or a building respects all the safety criteria concerning accessibility of emergency exits, availability of fire extinguishers, position of emergency lights, and so on, is extremely important for saving human lives in case of fires and other calamities. The safety criteria to meet are stated by law, and change from country to country.

As an example, the map in Fig. 1 shows the plant of a public library. The library is composed by three rooms. The one depicted on the upper part of the figure is the place where books are stored and contains four bookcases and a desk. The one below is the reading room containing two tables and a closet. Finally the small one on the right is the bathroom. The system will recognize all the hand-drawn symbols representing the objects relevant for the safety domain, as detailed in Section 5, and will check that they are correctly placed in space according to physical rules (a table cannot intersect a closet), and will reason about these objects following rules determined by the safety domain (a

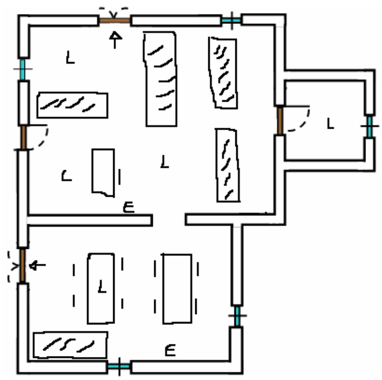


Fig. 1. A public library with furniture

closet placed in the middle of a room does not violate any physical or geometric constraint, but could violate a safety constraint if it makes an emergency path longer than a given threshold).

3 Intelligent Agents for Recognizing and Reasoning on Hand-Drawn Sketches

The architecture of the system that we propose, consists of the three modules described in detail in Section 4: *User Interface*, *User Reasoning Support*, and *Sketch Recognition*. Recognizing sketched symbols is demanded to the Sketch Recognition module, that integrates algorithms and solutions from the research domain of *automatic hand-drawn sketch recognition*. Verifying that the spatial relationships among symbols are satisfied is demanded to the User Reasoning Support, that must be able to reason about geometric concepts and thus integrates *geometric modeling* capabilities. Finally, reasoning over the drawn symbols in order to support the user in finding their right placement is once again demanded to the User Reasoning Support, that also adds some *artificial intelligence* to the system. Despite to their different abilities, the components that build our system share a common factor: they exploit *intelligent agents* in their implementation.

Following [6,7], an agent can be viewed as a software entity characterized by:

– *Autonomy:* An agent is not passively subject to a global, external flow of control; instead, it has its own internal execution activity, and is pro-actively oriented to the achievement of a specific task.
– *Situatedness:* An agent performs its actions while situated in a particular environment, and it is able to sense and affect such an environment.
– *Sociality:* Agents work in open operational environments hosting the execution of a multiplicity of agents. In these multi-agent systems (MASs), the global behavior derives from the interactions among the constituent agents.

Also *reactivity* is an important feature of agents [8], since they must be able to respond in a timely fashion to changes that take place in the environment. Finally, for many authors an explicit representation of *human-like mental attitudes* such as beliefs, obligations, permissions, is also required for characterizing agents.

According to the definition above, our system is a MAS. In our previous work, we have discussed why most of the entities that compose the Sketch Recognition Module are intelligent agents [9]. The User Reasoning Support module exploits agents as well: we have designed it in such a way that its functionalities are provided by the Decision Support Agent, a "deliberative" agent equipped with both "geometric modeling rules" and "application domain rules" represented by means of Deontic Logic [10] extended with nonmonotonicity. This kind of logic allows the agent to easily model and reason about what the user is permitted to draw, what he/she is forbidden to, what he/she is obliged to, and supports a "default" reasoning thanks to its nonmonotonic component (*human-like attitudes*). The agent will use these rules to check that what the user draws, satisfies both geometric and domain-dependent constraints, and will pursue the goal of avoiding their violation (*pro-activeness*). In case of violation, the agent, without any intervention from the user, will look for an alternative arrangement of the drawn objects (*autonomy*). If an "easy" solution cannot be found in a reasonable amount of time (*reactivity*), the agent will start to interact i with the user in order to collaborate for finding a solution (*sociality*). The virtual sheet where the user draws constitutes the environment of the system, and each agent in the system must perceive and must react to changes that occur inside it (*situatedness*).

In the sequel of the paper, we will show our use of agent technology for providing an intelligent and user-friendly support to users' decisions.

4 The Architecture of Our System for Reasoning about Hand-Drawn Sketches

The architecture of the proposed system is shown in Fig. 2. In Section 4.1 we will briefly describe the User Interface and Sketch Recognition modules, whereas in Section 4.2 we describe the Decision Support module by illustrating its functionalities and its architecture. Section 4.3 discusses how all the modules of our system will interact in order to support the user in the most precise and efficient way.

4.1 User Interface and Sketch Recognition Modules

The two modules described in this section, already implemented and tested, have been presented in [9].

The User Interface module manages the interaction between the user and the system when the user draws new symbols and when he/she manipulates (deletes, moves, resizes) them. The module integrates a Graphical User Interface for editing sketches. Since different sketch editors may be used for different application domains, new GUIs must be "pluggable" inside the system. In that case, the

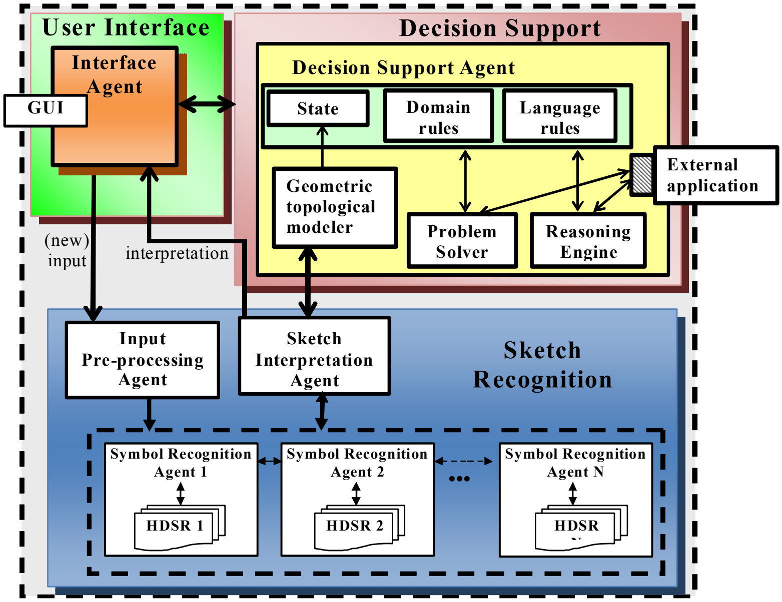


Fig. 2. The architecture of the system

Interface Agent is responsible for converting the information produced by the newly plugged editor into a format compliant with our system, and vice versa. It is also responsible for informing the agents belonging to the Sketch Recognition and to the Decision Support modules about both new strokes drawn by the user, and transformations of symbols previously drawn.

The Sketch recognition module provides the functionality of recognizing symbols belonging to a given visual language, hand-drawn by the user using a context based approach. The Input Pre-processing agent is responsible for segmenting and classifying the strokes arriving from the User Interface module. The recognition process performed by the intelligent agents devoted to symbol recognition (Symbol Recognition Agents, SRAs for short) and to the correct interpretation of the sketch (Sketch Interpretation Agent, SIA for short), is based on the knowledge about the language and about the symbols context, which is used for disambiguating the recognized symbols. SRAs exchange contextual information, which is sent to the SIA that solves possible conflicts and gives an interpretation of the sketch drawn. At the lowest level the symbols of the domain language are recognized by applying suitable Hand-Drawn Symbol Recognizers (HDSRs, for short) to the input strokes.

4.2 Decision Support Module

Functionalities: The Decision Support module, not completely implemented yet, provides the following functionalities.

Computing both simple and complex geometric and topological relations among symbols. Given two symbols $S1$ and $S2$, already recognized by the SIA, example of atomic geometric and topological relations between them are "$on\_the\_left\_of(S1, S2)$", "$included\_in(S1, S2)$", "$distance(S1, S2, D)$". The Decision Support Agent (DSA, for short) must be able to exploit its knowledge about geometry and topology in order to decide, given two symbols $S1$ and $S2$, whether $S1$ *intersects* $S2$, or if it is *in front of* it, which is the *distance between* them, and so on. The "geometric and topological modeler" component depicted in Fig. 2 is responsible of computing simple and general geometric and topological relations among the symbols drawn by the user starting from the information provided by the SIA. Since, according to the application domain, the DSA may need to compute relations that are more complex than those calculated by the geometric and topological modeler, and that cannot be decided a priori and once and for all, it must be able to access external components devoted to running optimized, ad hoc algorithms. If, for example, an application working on hand-drawn sketched graphs requires the ability to compute the shortest path, an external module will be accessed in order to run Dijkstra's algorithm on the input graph.

Modeling language dependent constraints on symbols. Spatial constraints change from visual language to visual language; for example, in the representation of an electronic circuit, a wire may graphically intersect another wire, while in a Use Case Diagram, no intersections between symbols are allowed.

In our reference domain, where symbols represent architectonic elements and furniture, no partial intersections of symbols are allowed while inclusion of some symbols might be permitted (the symbol representing a light may be entirely included on a table, since it might be placed above it). We may represent rules about relations of symbols using normative concepts such as permitted, forbidden, and obligatory:

–It is permitted that a symbol representing a light is included in a symbol representing a table.
– For any couple of symbols $S1$ and $S2$, it is forbidden that $S1$ intersects $S2$.
– It is obligatory that a symbol representing a door touches a symbol representing a wall.

Modeling domain dependent constraints on symbols. Besides constraints depending on the visual language, there are also constraints that depend on the specific application domain of the language. For example, the criteria for arranging furniture in a room in such a way that the comfort of people is ensured, are different from those for ensuring safety, although the set of available symbols and the language constraints they undergo, are the same. Also the domain dependent constraints may be easily expressed in term of "permitted", "forbidden", and "obligatory".

Verifying that all constraints are satisfied by the current sketch. The recognized symbols may or may not satisfy the constraints stated by the rules that the DSA possesses. The DSA is in charge of verifying if a violation takes place by running a "reasoning engine" that checks that all the rules are satisfied in the current state, consisting of the logical representation of the current placement of symbols.

Helping the user in finding alternative solutions. In case of violation of constraints, the DSA must inform the user that a violation has occurred. The DSA must also try to find, in a timely fashion, an alternative arrangement of the symbols drawn by the user, such that constraints are respected. A "problem solver" is in charge of this activity. If the alternative arrangement cannot be found in a reasonable amount of time (which may happen, since finding the right placement of objects considering a set of constraints is computationally expensive), the system must start a collaboration with the user for finding a "guided solution" to the constraint violation problem.

Architecture: The components of the Decision Support Agent are the following.

Geometric & topological modeler. It receives geometric information about symbols drawn so far by the SIA, creates a logical representation of these symbols (the agent's state) consistent with the logical representation used for constraints, and computes a set of "simple" topological and geometric relations, when needed by the engine or by the problem solver that have to verify which constraints are satisfied.

State and rules. The DSA is equipped with two sets of rules (also named "constraints" in the paper): those defining what must, can, and cannot be done with the language symbols (language rules), and those defining what must, can, and cannot be done according to the application domain (domain rules). The rules represent the agent program, and operate over the agent state that, as anticipated, consists of the logical representation of drawn symbols and of their placement.

Reasoning engine. A reasoning engine reasons about the rules in order to verify that they are respected by the current state. In Computer Science terms, the engine is an interpreter for the agent's program and state. Since the rules may include both "simple" and "complex" relations, the engine must be able to access both the geometric and topological modeler, for computing the former, and any external application that computes the latter. A wrapper, represented by the dashed box attached to the external component in Fig. 2, must implement the conversion of representations between the logical one used by the reasoning engine, and the one used by the external application. The engine must implement a nonmonotonic, forward reasoning. Nonmonotonicity is necessary to avoid stating in the rules everything that is permitted, and everything that is forbidden. Forward reasoning is necessary because the engine starts reasoning from the data it possesses (the current state) and applies recursively all the rules until all of them succeed, or one of them fails, raising a rule violation.

Problem solver. Finally, a problem solver looks for alternative solutions that meet the rules, re-arranging the drawn symbols (namely, finding a state different from the current one) in a way that does not raise conflicts with the agents' program. The problem solver must use the engine in order to verify the correctness of the solutions (new states) that it finds.

State and rules' representation: For representing the agent's state, we have chosen first-order logic atoms of type $represents(S, Sym)$, where S is an identifier and Sym is one among the symbols of the language; $has\_bounding\_box(S, BB)$, where BB is a couple of points defining the symbol's bounding box; $starts(S, StartingPoint)$; and so on. The relations between symbols may be represented as atoms as well: $on\_the\_left\_of(S1, S2)$, $contained\_in(S1, S2)$, and so on.

For representing the agent's rules, we have chosen *deontic logic*, which is the logic to reason about ideal and actual behavior. From the 1950s, von Wright [10] and others developed deontic logic as a modal logic with operators permission (**P**), obligation (**O**), and prohibition (**F**). Providing details about deontic logic is out of the scope of this paper; for more information about this topic, the reader may refer to [11].

By using deontic logic, we can easily express both language and domain rules such as "For any couple of symbols $S1$ and $S2$, it is forbidden that $S1$ intersects $S2$", which is represented by

$$\forall S1, S2\ \mathbf{F} intersects(S1, S2) \tag{1}$$

or, in an electronic engineering domain, "It is obligatory that a symbol representing a CPU is included in a symbol representing a motherboard", that becomes

$$\begin{aligned} \forall S1, S2((represents(S1, cpu) \wedge represents(S2, motherboard)) \\ \Rightarrow \mathbf{O}\ included\_in(S1, S2)) \end{aligned}$$

and so on.

The Decision Support Module can be implemented by exploiting the IMPACT framework [12] that allows the user to define rules that embed deontic operators and inside which calls to external code may appear. However, IMPACT is a commercial product and we are instead aiming at developing a free system. For that reason, we are considering the recently developed RBSLA framework (`http://ibis.in.tum.de/projects/rbsla/index.php`) that implements a rule-based system able to integrate deontic modalities and is freely available under GNU license.

4.3 System Behavior

The system behavior is described by the following algorithm:

1. Initially, the DSA state is empty
2. The user draws the symbol S
3. S is recognized by the Sketch Recognition module, and is passed to the geometric and topological modeler of the DSA that creates its representation as a logical atom and adds it to the current DSA state
4. The DSA runs the reasoning engine using the language and domain rules as program, on the current state
5. If the current state does not violate the rules, then the user can go on drawing (step 2), else

5.1 The user is informed of the violation, and the problem solver is run, in order to find an alternative solution by moving only the last drawn symbol
5.2 If the solution is found in an amount of time lower that a given threshold, then the user is informed of the found solution, else the user is asked to select a symbol that the problem solver will try to move, going back to step 5.1.

5 Checking Safety Criteria in Buildings with Our System

The best way to illustrate how our system might be exploited to cope with real problems is through an example.

Fig. 1 shows a building plant where some furniture symbols have been sketched. We suppose that the user needs only to sketch furniture since building plants (including walls, windows, and doors) are loaded from external files. In order to face the problem of building's safety, the user has to sketch

– Furniture having a height that could occlude the light (i.e., wardrobe, libraries, and so on), represented by rectangles containing dotted lines;
– Furniture that cannot obstruct the light (i.e., tables, desks, and so on), represented by empty rectangles;
– Chairs represented by lines placed near low furniture (see Fig. 1).

Moreover the user may sketch an arrow to specify an emergency exit, an "L" to represent a light, and an "E" to represent an extinguisher. In order to support the user in finding the right placement of lights, furniture, and fire extinguishers, the DSA uses a couple of external modules to compute the emergency paths and the poorly illuminated areas. Some of the security constraints are specified by the following deontic rules:

– It is forbidden to have a symbol representing a fire extinguisher behind a symbol representing a closet
  $\forall S1, S2\ ((represents(S1, fire\text{-}extinguisher) \wedge represents(S2, closet)) \Rightarrow \mathbf{F}\ behind(S1, S2)$
– It is forbidden to have any symbol intersecting an emergency path
  $\forall P, S\ (emergency\_path(P) \Rightarrow \mathbf{F}\ intersects(S, P))$[1]

Fig. 3(a) shows the plant of Fig. 1 with emergency paths represented by dashed lines and poorly illuminated areas represented by grey rectangles. When the user draws a new symbol (or moves an existing one) the DSA checks if any rule is violated. As an example, if the user draws the table on the left of the lower room of Fig. 3(b), the emergency paths are updated and the DSA detects a

[1] The value P of the logical variable that appears in "$emergency\_path(P)$" must be computed by an external resource, and must be represented in such a way that the geometric and topological modeler can verify the intersects relation between it and the representation of any drawn symbol.

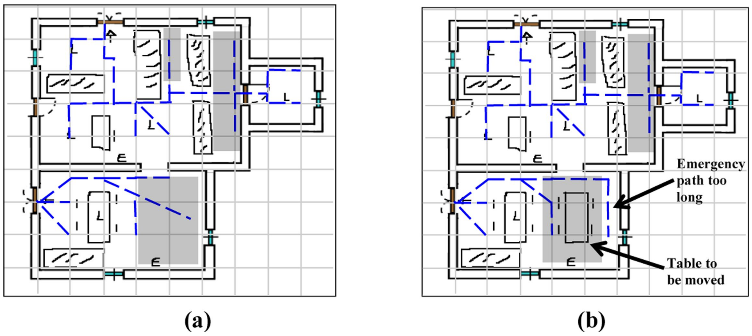


Fig. 3. Emergency paths and poorly illuminated areas of a sketched plant (a) and violation of security constraints (b)

constraint violation, i.e., an emergency path becomes too long. By applying the algorithm described in Section 4.3, the DSA suggests to the user to move the table towards the wall at the bottom of the plant, in order to shorten the length of the emergency path.

If the user does not want to move table, he/she can indicate to the system another object to move in order to find a solution. In our example the user could indicate to the system the table placed on the right of the bottom room, and the system would try to find a solution by moving only that table.

6 Related Work

In the following we discuss some sketch-based systems providing reasoning functionalities.

If we consider the reasoning functionalities provided by our proposed approach, to our knowledge, only the Design Evaluator system addresses similar issues. Indeed, it is an intelligent sketch system that offers critiquing annotations on drawings to facilitate design reflections [4]. The critiques are generated by applying design rules, coded as Lisp predicates, on the recognized graphical objects. However, "Design Evaluator" presents a monolithic architecture in spite of our modular and easy customizable agent oriented system. Moreover, Deontic Logic enables us to better formalize rules characterizing the considered language/problem.

sKEA (Sketching Knowledge Entry Associate) is designed for capturing knowledge from sketches [13]. sKEA can acquire several kinds of information from the sketches, such as semantic information, positions, what relations one glyph has with others, and which glyphs are conceptually and visually similar. The matching capability of sKEA can be used to suggest users what glyphs would be added and where they would be added. sKEA avoids the recognition problem by requiring the user to indicate when he/she begins and finishes drawing a new object as well as the interpretation of the object.

NuSketch is a sketch system that provides several reasoning services, including analogical reasoning and geographic reasoning [14]. Based on the nuSketch architecture, nusketch Battlespace (nSB), a system specialized for military reasoning, has been created [3]. The system provides a sketching interface for drawing military battle plans, which are understood using qualitative spatial reasoning. However, the system does not attempt to perform shape recognition of the sketches. Rather, it depends on voice input and specific selection procedures from the user to define object types and names.

7 Conclusions and Future Works

In this paper we have proposed an integrated environment, conceived as a multi agent system, to support user's reasoning through sketching. We have already demonstrated that the agent technology is suitable for recognizing hand-drawn symbols and for solving ambiguities that may arise in their interpretation [9]. The extension of our system, namely the Decision Support module, also exploits intelligent agents. We have motivated our choice of having an entirely agent-based system, and highlighted its advantages. This choice represents the major difference between our proposal and the related work.

Our future work is to complete the implementation of the system for the presented case study, and to design and develop innovative solutions to enhance user interaction with the system maximizing the advantages of the proposed approach.

References

1. Purcell, A.T., Gero, J.S.: Drawings and the design process. Design Studies 19, 389–430 (1998)
2. Kavakli, M., Gero, J.S.: Sketching as mental imagery processing. Design Studies 22, 347–364 (1999)
3. Forbus, K., Usher, J., Chapman, V.: Sketching for military courses of action diagrams. In: Proc. of IUI 2003, pp. 61–68. ACM Press, New York (2003)
4. Oh, Y., Do, E.Y.L., Gross, M.: Intelligent critiquing of design sketches. In: Proc. of AAAI Fall Symp. Making Pen-Based Interaction Intelligent and Natural (2004)
5. Landay, J.A., Myers, B.A.: Sketching interfaces: Toward more human interface design. IEEE Computer 34(3), 56–64 (2001)
6. Jennings, N.R.: An agent-based approach for building complex software systems. Communications of ACM 44(4), 35–41 (2001)
7. Lind, J.: Issues in agent-oriented software engineering. In: Ciancarini, P., Wooldridge, M.J. (eds.) AOSE 2000. LNCS, vol. 1957. Springer, Heidelberg (2001)
8. Jennings, N.R., Sycara, K., Wooldridge, M.: A roadmap of agent research and development. Journal of Autonom. Agents and Multi-Agent Syst. 1, 7–38 (1998)
9. Casella, G., Deufemia, V., Mascardi, V., Costagliola, G., Martelli, M.: An agent-based framework for sketched symbol interpretation. Journal of Visual Languages & Computing 29(2), 225–257 (2008)
10. Von Wright, G.: Deontic logic. Mind, 1–15 (1951)

11. Meyer, J.J.C., Wieringa, R.J.: Deontic Logic in Computer Science. John Wiley and Sons, Chichester (1993)
12. Rogers, T.J., Ross, R., Subrahmanian, V.S.: IMPACT: A system for building agent applications. J. Intell. Inf. Syst. 14(2-3), 95–113 (2000)
13. Forbus, K., Usher, J.: Sketching for knowledge capture: A progress report. In: Proc. of IUI 2002, pp. 71–77. ACM Press, New York (2002)
14. Forbus, K., Ferguson, R., Usher, J.: Towards a computational model of sketching. In: Proc. of IUI 2001, pp. 77–83. ACM Press, New York (2001)

Handling of Task Hierarchies on the Nepomuk Social Semantic Desktop

Marko Brunzel[1] and Roland M. Mueller[2]

[1] DFKI GmbH–German Research Center for Artificial Intelligence, Kaiserslautern
marko.brunzel@dfki.de
[2] University of Twente, Enschede, The Netherlands
r.m.mueller@utwente.nl

Abstract. The idea of the Semantic Web is becoming more and more reality. Recent developments in Semantic Desktop research bring semantically enriched data onto the knowledge workers desktop environment. A crucial application for today's knowledge worker is task management, which goes beyond simple ToDo lists.

In this paper we explain why task management applications in social semantic environments need an appropriate user interface to take advantage of the opportunities ahead. Then we propose the usage of an UI which is tailored towards the underlying data schema: the task management ontology.

One challenge for task management based on flexible semantic schemas in social environments is to scope on what should be shared and what not. Sharing and transfer of semantic, graph-represented information is highly desired. We propose WYSIWYS: What You See Is What You Share. This paradigm gives the human user an interface where he/she maintains control over what is shared and what not.

1 Introduction

Recent developments in the Semantic Web research field address the usage of semantics on the desktop. The Nepomuk [3] project aims to implement the Social Semantic Desktop. The core message of the social semantic desktop is that arbitrary information pieces, like existing desktop resources, can be associated semantically with each other.

Task management is especially important for today's knowledge workers. Nepomuk task management is a knowledge-management application that provides a generic infrastructure that handles tasks and integrates task-related information from different sources. Nepomuk task management is based on the Task Model Ontology (TMO) [1]. The TMO aims to provide a shared schema where the relevant entities of the task domain are captured. By using the Nepomuk Social Semantic Desktop, the user will obtain a graph of associated information artifacts. The knowledge worker's tasks are part of this graph.

Handling task hierarchies is encompassed by the issues of (1) visualization and (2) editing. For collaborative tasks, additionally (3) scoping of the parts

M. Sebillo, G. Vitiello, and G. Schaefer (Eds.): VISUAL 2008, LNCS 5188, pp. 315–318, 2008.

to be shared is a challenge. The exchange of granular information, which is represented by an RDF graph, bears the problem of how the information to be shared should be separated from what should be kept private. In this paper we propose WYSIWYS: What You See Is What You Share. WYSIWYS is about providing a UI where the information to be shared can be displayed at a glance.

Task Management in a social environment requires ad-hoc sharing of tasks and related information. For information which is based on RDF data this comes down to the challenge of determining an appropriate sub-graph from the overall RDF-graph within a repository. This can be done by statically defining what should belong to a compound instance. But for the ad-hoc sharing of tasks, a static, fixed closure is not sufficient. There are plenty exceptions expected, where more or less information should be exchanged. The proposed UI paradigm can help in such a situation by providing the user with a content oriented visualization. The knowledge worker can determine the closure that should be exchanged in a flexible manner. Predefined rules of what to include can still be used as the default but can be easily superseded by the user.

2 The DisTREEbute UI

In the case of semantically-represented data on the desktop, the problem of ad-hoc publication and sharing emerges. In the following, we shall present the user interface and visualization for ad-hoc ontology sharing. The aim of the prototype *Dis*TREE*bute*[1] is a compact visualization of entire task hierarchies, which are described based on a domain ontology.

Solutions for ontology visualization and ontology user interfaces can be classified in a continuum between two extremes: generic RDF visualizations and special domain user interfaces. Both alternatives have their drawbacks. Generic RDF visualizations normally do not confer to the specific usage patterns in a scenario and have low usability. Special tailored user interfaces are rigid, can't be customized to changing ontologies, and do not fit to the circumstances and requirements of the Social Semantic Desktop. The DisTREEbute prototype can be positioned in the middle of this continuum: on the one hand it is flexible and customizable for changing ontologies, on the other hand it has built-in domain logic which improves the visualization and usability. The visualization of an instance in the hierarchy can be adjusted according to a domain ontology.

In Figure 1 a screenshot of DisTREEbute is shown. By selecting a start node in the ontology graph, a hierarchy is obtained. The most important element of the task hierarchy is the task with its name and an automatically created invisible ID. The task names are displayed directly as nodes. For other properties that have a maximum cardinality of 1 (e.g. the *task state*) there are corresponding sub-nodes who display the name and the value of the property. For properties that have a cardinality of more than one value, *grouping nodes* bundle together all properties of a certain type. Often not all possible properties have values.

[1] Scottish for distribute.

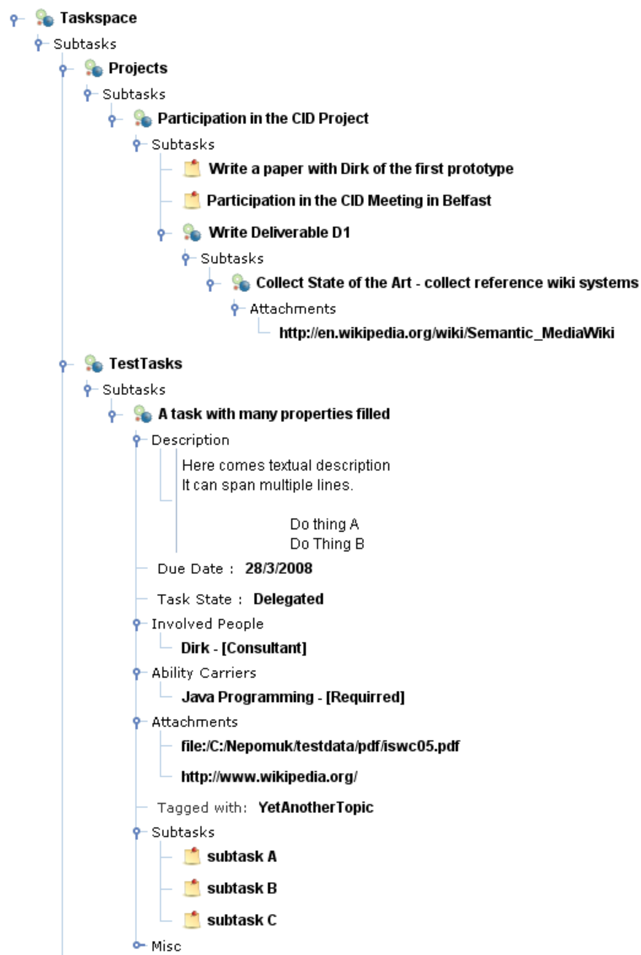


Fig. 1. Screenshot of DisTREEbute

DisTREEbute considers this sparseness of the instances and shows only used properties.

The main feature of the user interface is the view definition: this scoping follows the principle "What You See Is What You Share" or WYSIWYS. The user can expand or collapse parts of the task hierarchies and thereby determine the scope of the shared ontology instances. Therefore, the scope of the displayed and subsequently shared information can be determined by two principles:

1. By unfolding (expanding) and folding (collapsing) the task hierarchy, the scope can be broadened and narrowed by a all children of a node.
2. By checking (selecting) and / unchecking (unselecting) checkbox the scope can be determined on the properties and individual children.

After the selection, the selected items, resp. the RDF statements which are underlying the selected items, might be sent to another participant or shared within a group of people.

3 Related Work

Today's work is in majority collaborative work. For collaboration and remote desktop sharing, the principle What You See is What I See (WYSIWIS) is known [4]. In contrast to WYSIWIS, in our approach only fragments of the task hierarchy should be shared and there is only a partly overlap of ontologies. Both sides could be partially different ontologies.

Regarding the display of semantic information, the Tabulator Redux system developed by Berners-Lee et al. [2] is relevant. Tabulator Redux allows to display RDF data as hierarchies.

4 Conclusions and Future Work

In this paper we proposed the WYSIWYS (*What You See Is What You Share*) paradigm which can be applied to determine what parts of a hierarchy, respectively graph, should be shared with remote collaborators.

RDF-Graphs in social settings raise new and difficult issues concerning the privacy of the information. WYSIWYS can be used to determine a sub-graph which can be transferred to another person. By using a UI according to WYSIWYS, the user stays in charge of what is shared. WYSIWYS therefore tackles a major security and privacy issue for the emerging field of social-semantic information exchange. DisTreebute provides a prototypical implementation of the WYSIWYS paradigm and visualizes tasks which are based on the Nepomuk Task Management Ontology.

Acknowledgements. This work was supported by the European Union IST fund (Grant FP6-027705, project NEPOMUK).

References

1. Bernardi, A., Brunzel, M., Grebner, O., Ong, E., Riss, U., Roth-Berghofer, T.: Task Management Model, Technical report, Deliverable 3.1 of Project Nepomuk, http://nepomuk.semanticdesktop.org/xwiki/bin/download/Main1/D3-1s/D3.1_v10_NEPOMUK_Task_Management_Model.pdf
2. Berners-Lee, T., Hollenbach, J., Lu, K., Presbrey, J., Prud'ommeaux, E., schraefel, mc.: Tabulator Redux: Browsing and writing linked data. In: WWW 2008, Workshop: Linked Data on the Web (LDOW 2008) (2008)
3. Groza, T., Handschuh, S., Moeller, K., Grimnes, G., Sauermann, L., Minack, E., Mesnage, C., Jazayeri, M., Reif, G., Gudjonsdottir, R.: The nepomuk project - on the way to the social semantic desktop. In: Pellegrini, T., Schaffert, S. (eds.) Proceedings of I-Semantics 2007, pp. 201–211, JUCS (2007)
4. Stefik, M., Bobrow, D.G., Foster, G., Lanning, S., Tatar, D.: Wysiwis revised: early experiences with multiuser interfaces. ACM Trans. Inf. Syst. 5(2), 147–167 (1987)

Musimage: A System for Automatically Presenting Pictures Depending on the Music Being Played

Jesús Ibáñez, David García, and Oscar Serrano

Department of Information and Communication Technologies,
Pompeu Fabra University, Barcelona, Spain
{jesus.ibanez,david.garcian,oscar.serrano}@upf.edu

Abstract. This paper presents Musimage, a novel system which displays pictures according to the songs being played at the same time. By using the interface the user selects the songs to be played, but the pictures are automatically selected. For each song to be played, the system selects a set of pictures, according to various criteria corresponding to some features of the song. In this sense, the pictures to be shown are, metaphorically, triggered by the songs.

1 Motivation

Music triggers recollections. By listening to a particular song, we can remember events that happened and feelings we had in the epoch we used to listen to that song in the past. Some songs do not evoke concrete instants but epochs in our life. Songs we used to listen to in 1994 will probably trigger some recollections from that year or the next few years. This is the idea that led us to design and develop Musimage, the system presented in this paper. Our original idea was to design a user interface which on the one hand accompanies the user in this recollection process, and on the other hand is able to illustrate the song.

Musimage displays pictures according to the songs being played at the same time. By using the interface (see figure 1), the user selects the songs to be played, but the pictures are chosen automatically. For each song to be played, the system selects a set of pictures, according to various criteria corresponding to certain features of the song (namely lyrics and year). The topics of a song are automatically obtained from its lyrics through a categorization process. These semantic features (topics and year) are then used as cues for collecting pictures from both the user's personal picture collection and picture servers on the Internet.

2 Overall Working

Although in the system we present in this paper we deal with songs (as triggers) and pictures (as triggered medias), when designing the architecture we tried it to be flexible and extensible, so that it is able to work with other medias and

M. Sebillo, G. Vitiello, and G. Schaefer (Eds.): VISUAL 2008, LNCS 5188, pp. 319–322, 2008.

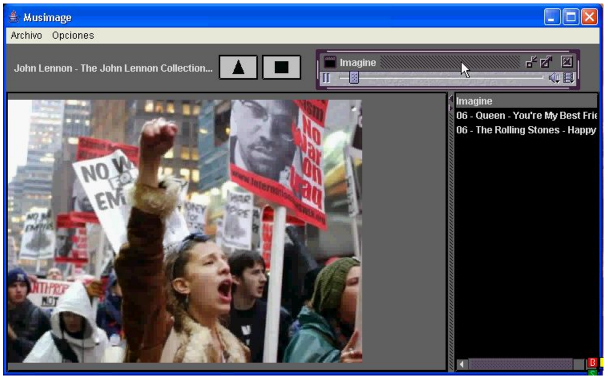


Fig. 1. User interface

interrelations among them. As a result of these considerations and requirements, the system architecture is a multi-agent system [1].

The system employs text categorization techniques to assign a set of weighted topics to a song lyrics. These weighted topics are then contrasted with the pictures topics. Therefore, the system requires two previous step for it to work properly:

– On the one hand, it needs the categorizer of song lyrics to be previously trained with a representative set of lyrics.
– On the other hand, it requires that the user's personal collection of pictures is annotated with keywords (topics) and date.

In order to illustrate the overall functioning of Musimage, we show next the ordered series of steps which are followed in a typical case of use:

1. The user selects a song or a list of songs to listen.
2. The system analyzes the ID3 description of each song. The ID3 codifies meta-data in the mp3 files. These meta-data include the song title, artist, album, year, genre, comments, and they can even include the lyrics. The system obtains the four first mentioned meta-data.
3. The system retrieves the lyrics corresponding to each song from a lyrics server on the Internet. To find the lyrics, it uses the meta-data obtained from the ID3 description.
4. The system automatically assigns a list of weighted topics to a song by applying text categorization techniques to the song lyrics. The list of weighted topics is a list of pairs topic/weight where the weight indicates the degree with which the song belong to this topic. To achieve the automatic categorization, the current system employs the Naive Bayes algorithm included in the Weka libraries [2].
5. The system selects a set of pictures to be shown while each song is played, taking into account the following song characteristics: list of topics, year and temporal length.
6. The songs in the list are sequentially played, and for each song being played the corresponding selected set of pictures is displayed.

The next section details the process indicated in the 5th step, where the system selects the set of pictures to be shown while each song is played, taking into account several song characteristics (list of topics, year and temporal length).

3 Selection of Pictures

Given a song s to be played, the system selects a subset of pictures to be shown while playing the song. In particular, n pictures are selected, where n is calculated as follows:

$$n = \frac{length(s)}{refreshTime}$$

where $length(s)$ is the duration of the song s and $refreshTime$ is the time that every picture is exposed. More specifically, the system selects the n pictures with the greatest value of similarity with respect to the song s. Similarity between a song s and a picture p is defined as:

$$similarity(s, p) = ((c_1(s, p) \times w_1) + (c_2(s, p) \times w_2))$$

where the similarity is calculated as the weighted addition of two different criteria, c_1 and c_2. c_1 is a criterion which measures the similarity by taking into account both the year when the song was recorded and the year when the picture was taken. On the other hand, c_2 is a criterion which measures the similarity by estimating the affinity between the topics of the song and the topics of the picture. Both criteria c_1 and c_2 are respectively weighted by w_1 and w_2, which are such that $w_1 + w_2 = 1$. These weights can be fixed by the user by employing the user interface. Next we show how c_1 and c_2 are calculated.

3.1 Year-Based Similarity

The criterion c_1 measures the similarity between a song s and a picture p taking into account both the year when the song was recorded and the year when the picture was taken. This similarity is defined as:

$$c_1(s, p) = \begin{cases} 0 & , \ d(s, p) < 0 \\ \frac{c - d(s,p)}{c} & , \ 0 \leq d(s, p) < c \\ 0 & , \ d(s, p) \geq c \end{cases}$$

where c is a constant value which defines a maximum number of years, such that if the picture p was taken c or more years later than the song was recorded, there is no similarity between p and s. And $d(s, p)$ is defined as:

$$d(s, p) = year(p) - year(s)$$

where $year(p)$ is the year when the picture p was taken and $year(s)$ is the year when the song s was recorded.

Thus, by employing the criterion c_1, the similarity between a song and a picture is greater than zero if the picture was taken the same year that the song was recorded or after that year, but less than c years later (see figure 2). Moreover, the nearer both dates are the greater the similarity is.

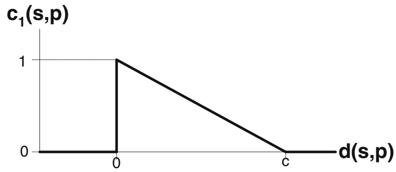


Fig. 2. Year-based similarity

3.2 Topic-Based Similarity

The criterion c_2 measures the similarity between a song s and a picture p by estimating the affinity between the topics of the song and the topics of the picture. The song s is automatically categorised. As a result, for each category of songs, sc, a value $vcat(s, sc)$ is obtained, such that this value indicates the degree the song s belongs to the category sc. Let $\{sc_1, ..., sc_n\}$ be the set of categories of songs sc_i such that $vcat(s, sc_i) > 0$. On the other hand, each picture p is annotated with a set of categories of pictures. Let $\{pc_1, ..., pc_p\}$ be the set of categories of pictures that the picture p belong to (i.e. the picture p is annotated with these categories).

The system utilises a table which stores the affinity between pairs (sc, pc), where sc is a category of songs and pc is a category of pictures. Let $affinity(sc, pc)$ be the value of the affinity between sc and pc stored in that table. This value indicates the degree with which we expect a picture of category pc to be triggered by a song of category sc. Thus, the topic-based similarity between the song s and the picture p is defined as:

$$c_2(s, p) = \sum_{i=1}^{n} \sum_{j=1}^{p} vcat(s, sc_i) \times affinity(sc_i, pc_j)$$

and values of $c_2(s, p)$ are later normalized between 0 to 1.

References

1. Ferber, J.: Multi-agent system: An introduction to distributed artificial intelligence. Addison-Wesley, Reading (1999)
2. Witten, I.H., Frank, E.: Data mining: Practical machine learning tools and techniques with java implementations. The Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann, San Francisco (1999)

An Architecture for Multimedia Content Publishing with GIS-Based Retrieval Facility

Michele Di Capua[1], Donatella Padovano[1], and Monica Sebillo[2]

[1] Unlimited Software, Naples, Italy
md@unlimitedsoftware.it, d.padovano@u-s.it
[2] Dipartimento di Matematica e Informatica (DMI), University of Salerno, Via Ponte Don Melillo, I-84084 Fisciano (SA), Italy
msebillo@unisa.it

Abstract. With the rapid growth of broadband network, distribution of multimedia content via internet portals has become a must way to go. This work addresses the creation of a simple web architecture based on an open source framework capable to provide a means to publish and retrieve multimedia content, also with some facilities based on data positioning, associated with multimedia files and subsequent retrieval.

Keywords: multimedia, streaming, digital assets, geographic information system, flash video format.

1 Introduction

The aim of our research is to design and realize a stable software architecture [1], able to manage typical multimedia content (audio, video and images), based upon open source software reuse. Our requirements includes capturing and describing digital materials using a flexible submission workflow, distribute an organization's (mainly institutional) digital assets [2] over the web through a search and retrieval system, also with geographic queries, based on positioning data, and, in addition, preserve digital assets over the long term [3]. Others general requirements are related to object oriented technology solution and software portability.

1.1 Open Source Frameworks Analysis

Starting from the previous requirements we have analyzed some available frameworks that fit, at a first glance, our needs. To achieve our goal we have analyzed some literature proposals concerning multimedia content management systems available in the current open source scenario. Among the possible solutions we restrict our choices to four frameworks: ContentDM [4], DSpace [5], Fedora [6] and WikiD [7]. In the following tables, we briefly report a comparison among these frameworks, that puts in evidence the availability offered by each single product of some fundamental features grouped by data support and user support. The values

M. Sebillo, G. Vitiello, and G. Schaefer (Eds.): VISUAL 2008, LNCS 5188, pp. 323–326, 2008.

Table 1. Comparison among selected frameworks concerning data support features

Feature	ContentDm	DSpace	Fedora	WikiD
Arbitrary Bitstreams	+	+	+	
Arbitrary Complex Object	+	+	+	+
Versioning			+	+
Locale Metadata Elements	+	+		+
Preservation Metadata		+	+	
Batch Input	+	+	+	-
Rich Metadata Searching	+	-	-	-
Full-text Searching	+	-		-

Table 2. Comparison among selected frameworks concerning user support features

Feature	ContentDm	DSpace	Fedora	WikiD
User Roles with Privileges	-	+		
Workflows		+		
Object Marshalling	+	+		
Arbitrary Bitstream Retrieval	-	-	-	
Arbitrary Object Retrieval			+	
Web Interface	+	+		+
Content Easily Integrated into Web Pages	+		-	+

reported in the columns are +, - and blank, that stands for good, sufficient and not present, respectively.

All the frameworks seem to work fine with Bitstream formats and complex objects. Focusing our attention on metadata management we can easily discover that ContentDm and DSpace are the well suited for this feature.

DSpace offers better possibilities to manage user features and most of all, it supports workflow approval definition, which is an important activity in institutional multimedia publishing. Object marshalling is the ability to compose a complex object (e.g. an album consisting of multiple mp3 files plus other information) for submittal, and possibly manage the relationships between the component objects. The results of our analysis have guided us to select DSpace as our starting point for the architecture to be developed. Additional benefits are related to the technology on which this tool is realized, namely Java (or better J2EE model), that gives us some guarantees on its portability and object oriented design. A final consideration is related to the DSpace support of PostgreSQL as its integrated RDBMS, that give us the possibility of using the GIS extension (PostGIS) of the database, in order to provide some simple retrieval features based on geographic information.

1.2 The Adopted Solution (DSpace)

DSpace is an open source software framework that provides tools for the management of digital assets. It is also used as a platform for digital preservation activities. DSpace is a product of the HP and MIT Alliance and it is shared under a BSD license. We focus our attention on the DSpace structure in order to analyze which components may be integrated or simply modified in our goal architecture.

1.3 The Customized Solution

To match our requirements we have modified or added some components belonging to both the 3 levels of the DSpace architecture, shown in fig.1. Starting from the application level, and focusing on video, we have developed some customized web pages (JSP) that integrates the multimedia content (video specifically) through a simple Adobe Flash Video Player, in order to play .FLV video format.

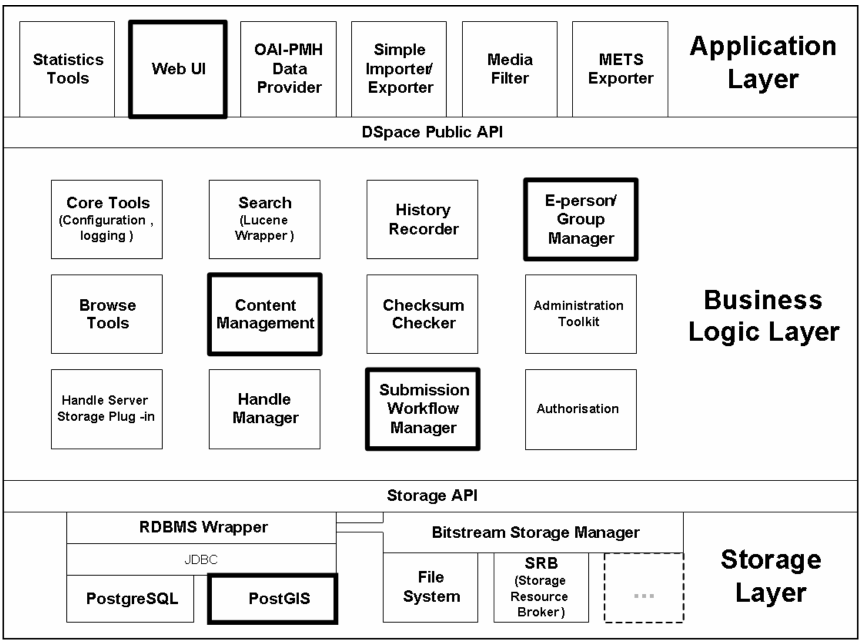


Fig. 1. The DSpace modified architecture layers (bold boxes)

In our customized solution all videos submitted to the web application are converted to FLV format (using FFmpeg utility). In this way we provide streaming features in reading video content, easily achieved with the FLV format. In the business logic layer we have developed a customized workflow model, based on the DSpace one, with simplified roles and rules. Finally in the Storage Layer, we have installed the PostGIS extension to PostgreSQL and extended the data model of DSpace also with geographic information (e.g. latitude and longitude).

2 GIS Data Integration

The GIS data integration has been achieved by simply extending the data model (and relative metadata) stored in the RDBMS. Then, using the native PostGIS API we have provided a visual query tool that retrieves multimedia content related to a specific geographic area, with the "distance" function provided with PostGIS, and mapping the results on a Google Map-based view (fig.2).

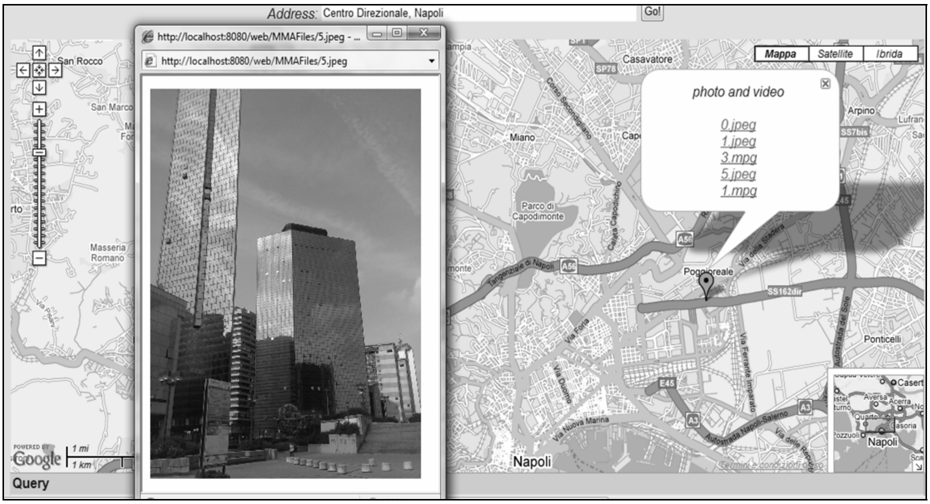


Fig. 2. The visual query tool integrated with Google Maps

References

1. Arms, W.Y., Blanchi, C., Edward, A.: Overly: An Architecture for Information in Digital Libraries, D-Lib Magazine (1997)
2. William, Y.: Arms: Key Concepts in the Architecture of the Digital Library, D-Lib Magazine (1995), `http://www.dlib.org/dlib/July95/07arms.html`
3. Hedstrom, Margaret.: Digital preservation: a time bomb for Digital Libraries. In: Symposium on Reconnecting Science and Humanities in Digital Libraries, Kentucky (1995)
4. Content, D.M.: Digital Collection Management Software, `http://www.contentdm.com`
5. Tansley, R., Bass, M., Smith, M.: DSpace as an Open Archival Information System: Current Status and Future Directions. In: van der Aalst, W.M.P., Best, E. (eds.) ICATPN 2003. LNCS, vol. 2679, pp. 446–460. Springer, Heidelberg (2003)
6. Payette, S., Lagoze, C.: Flexible and Extensible Digital Object and Repository Architecture. In: Nikolaou, C., Stephanidis, C. (eds.) ECDL 1998. LNCS, vol. 1513. Springer, Heidelberg (1998)
7. WikiD, OCLC Project, `http://www.oclc.org/research/projects/wikid/`

SISI Project: Developing GIS-Based Tools for Vulnerability Assessment

Bruno Della Rocca[2], Grazia Fattoruso[1], Sergio Locurzio[2], Francesco Pasanisi[1], Raffaele Pica[1], Alessandro Peloso[1], Maurizio Pollino[2], Carlo Tebano[1], Alfredo Trocciola[1], Davide De Chiara[3], and Genoveffa Tortora[3]

ENEA - Department of Environment, Global Change and Sustainable Development
[1] Portici Research Center, Località Granatello, 80055, Portici (NA), Italy
[2] Casaccia Research Center, Via Anguillarese, 301. 00123 S. Maria di Galeria (Roma), Italy
dellarocca@casaccia.enea.it
[3] Dipartimento di Matematica e Informatica, University of Salerno, Italy
Ponte don Melillo, Fisciano (SA) 84084, Italy
{ddechiara,tortora}@unisa.it

Abstract. In the framework of a wider research project, aimed at developing an high performance computing infrastructures in Southern Italy, the SISI project aims at developing customized GIS tools for vulnerable site selection in Italian territory. Vulnerable sites will be defined based on co-presence of different dangerous/vulnerable geographic features in given areas. Morphological and topological criteria will possibly be included in territorial diagnostics.

Keywords: GIS, Geoprocessing tools, Vulnerability Assessment.

1 Introduction

CRESCO (Italian acronym for "Computational Research Center on Complex Systems") is an ENEA Project, co-funded by the Italian Ministry of Education, University and Research (MIUR).

The aim of the CRESCO project is to develop a high performance computing infrastructure, located at ENEA Portici Research Center. The computing platform, with an estimated 10 Teraflops, will be the major node of the ENEA GRID computing facility which links HPC platforms located in the different ENEA Research Centers throughout Italy. Applications will focus on different matters, including life sciences, sociology and technological systems which may have a relevant impact on society. Namely, applications are organized in the following activity lines:

1. Implementation of innovative solutions in computing system architectures and GRID computing for the activities of advanced R&S in ENEA, requiring the use of computational resources highly performing (LA1);
2. Study of biologic objects in "systemic" point of view and study of natural systems (animal e social communities) according to paradigm of complex systems (LA2);
3. Study of technological complex systems with their mutual interactions and the creation of suitable tools for the modeling, simulating and control processes (LA3).

M. Sebillo, G. Vitiello, and G. Schaefer (Eds.): VISUAL 2008, LNCS 5188, pp. 327–330, 2008.

In the framework of LA3 activity lines, ENEA-ACS Dept. launched the SISI project (Italian acronym for “Information System for Protection of Infrastructures and Population”), focused on development of GIS tools for territorial diagnostics and modeling. SISI Project is currently being developed by ENEA-ACS Dept. in partnership with CeRICT (Regional Center for Information Communication Technology) – University of Salerno.

2 Overview of the SISI Project

The main aim of the SISI project is to develop a GIS capable of locating vulnerable sites on territory, considering human or natural factors.

Location and extension of vulnerable sites are based on the presence of one or more dangerous/vulnerable geographic features inside a specific area (Fig. 1). A geographic feature is regarded as “dangerous” if, due to human or natural causes, it can induce the partial or total failure or loss of functionality of other features in its neighborhood. Examples of dangerous sites are portions of land subject to landslides, chemical plants, dumps, *etc*. On the other hand, a geographic feature is regarded as “vulnerable” if, due to presence of dangerous features located in its neighborhood, it can be subjected to partial or total failure or loss of functionality. Typical examples of vulnerable features are civil infrastructures and human settlements.

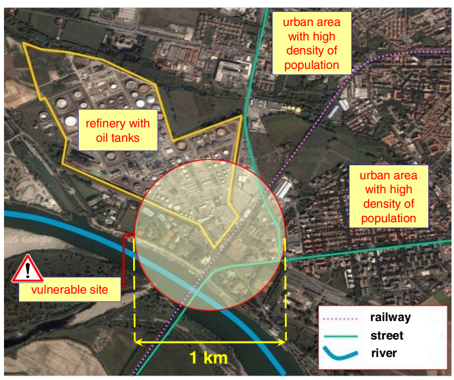


Fig. 1. Selection of a vulnerable site based on the co-presence of geographic features

It is noteworthy that the same geographic feature can be regarded either as dangerous or vulnerable feature. Moreover, depending on the nature of considered geographic features, dangerousness or vulnerability can be defined considering also morphological criterion or topological relation among features.

The project will be implemented using GIS technology, capable of fusing disparate data, analyzing trends and patterns, and providing actionable information that is easily shared and understood [1], [2].

3 Overview of the SISI Architecture

As for functionality, SISI will consist of three integrated modules (Fig. 2):

1. Geospatial database, containing data for the entire Italian territory;
2. Geoprocessing customized tools for territorial "diagnostics", aimed at identify vulnerable sites based on above described criteria;
3. Web-GIS for viewing and sharing the vulnerability maps on the web.

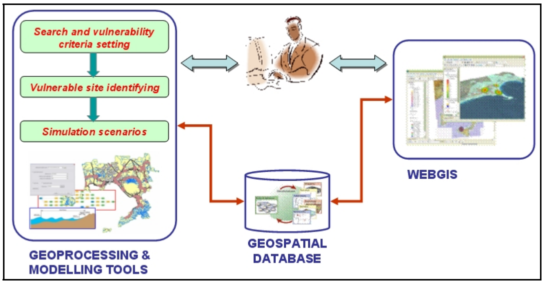


Fig. 2. Architecture of the SISI system

SISI diagnostic tools are based on advanced and customized geospatial functions that allow either to accurately pinpoint physical areas on the entire national territory, where user-defined vulnerability criteria are verified, or to map areas where exist defined morphological and topological relationships among the geographic elements.

These functionalities will be accessible via easy-to-use interfaces. The SISI user will be an expert with knowledge of territory (e.g. local authorities or administrators), but not necessarily an expert GIS analyst or user.

The diagnostic tools will allow the following functions, at national scale extension:

- Selecting investigated geographical features;
- Defining extension of research area (e.g. circle with radius 1 km);
- Defining possible morphological conditions or topological relation among features which influence dangerousness and/or vulnerability.

Eventually, at a smaller scale, prototype modeling tools will be developed, considering specific hazard factors.

4 Implementation of Diagnostics Tools

The SISI project is actually in progress. It will be developed in subsequent phases, whose incremental outcomes will make up the expected final result.

The aim of the first step is to identify areas containing elements or phenomena of interest, i.e., it is required to verify element/phenomena co-location in user-prefixed size zones. Facing this problem implies both to treat large-scale phenomena acting on layers which are dense of information, and to perform complex and time-consuming operations. In order to improve the efficiency of this phase, we aim to streamline the

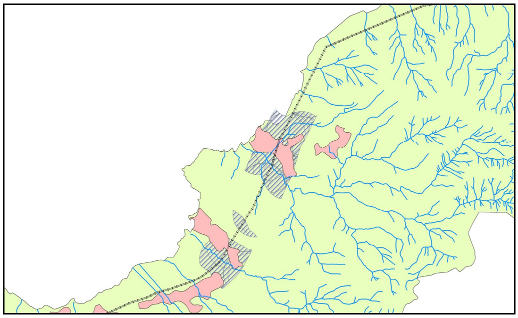


Fig. 3. Example of selected areas resulting from the first step

whole process by designing algorithms meant to reduce the computational complexity without increasing the possibility of error refund. The basic idea is to exploit properties of two geometric operations: *buffer* and *intersect*. The former draws boundaries to a specified distance around the input features, the latter detects features that geometrically shares a part with the source feature or features. By applying the intersection on the buffered zones, calculated on the input layers, a set of polygons is obtained, which match condition of contemporary presence within a given distance of features belonging to layers. In fact

- Each buffered zone contains points that meet the condition of being within a radius of a given distance from the features of each layer;
- The intersection of these buffered zones results in a polygon that contains items matching all the conditions, that is to say, items are within a range of predetermined distance from the features of each layer in input.

In order to optimize the performance of *buffer* and *intersect* we have applied filters meant to reduce the cardinality of the input features. The idea is to avoid computing buffer and intersection for features that certainly would be excluded from the final result. In practice, the first operation is a spatial query, namely *select by location*, which selects features based on their position relative to other features. This operation is applied on two input layers used in order to reduce the cardinality of features layer input. Once data have been properly filtered, the buffer and intersect can be applied to layers (Fig. 3). Then, by combining filters, buffer and intersection operations, the whole first step is optimized both in terms of computation time and costs.

As for the second phase of the SISI project, we are investigating a solution capable to optimize the matching of both topological and morphological properties on the basis of the results obtained in the previous step.

References

[1] Longley, P.A., Goodchild, M.F., Maguire, D.J., Rhind, D.: Geographical Information Systems and Sciences, 2nd edn. Wiley, Chichester (2005)

[2] O'Sullivan, D., Unwin, D.: Geographic Information Analysis. John Wiley & Sons, Chichester (2003)

GISTURISMO Massa-Carrara – Advanced Territory's Dynamic Fruition Service for a Conscious Tourism

Antonio De Luca

SIT Service and Cartography Manager Provincia di Massa-Carrara
http://gisturismo.provincia.ms.it

1 The Concept

Providing detailed and comprehensive information on everything a region can offer in terms of tourist, environmental, cultural, architectural, sporting or eno-gastronomic supply, making use of the most advanced technologies in order to satisfy the ever more increasing requirements of visitors either occasional or not who want to consciously tackle a territory.

The challenge is to satisfy at the same time the needs of who wishes to gather information before visiting a place, who isn't familiar with the region and asks to be accompanied and who demands information during an itinerary either to deepen his knowledge or to modify the destination real time.

Everything is done providing information quality, sources integration, easy management and updating, and the highest degree of accessibility.

Thus GISTURISMO Massa-Carrara is created, an open multiplatform territorial information system which from a single central GeoDB gives both static and dynamic information by means of Web, PDA (Personal Digital Assistant), Smartphone, Cellular Phones and GPS Navigators.

Before planning a trip, an excursion or any other pleasure, sports or cultural activity, it is possible through internet (either from a personal computer or from public totems placed in strategic locations such as Tourist Boards - APT, InfoPoints and Hotels) to gather all the needed information; once provided with a PDA or a smartphone, it can be possible for the user to travel across the province of Massa-Carrara by train, by car, by bicycle or walking, and to get information from the GISTURISMO Massa-Carrara service. Instead, if you don't know the region, you can rely on a customised multilingual GPS satellite navigation system, which is integrated with itineraries, interest points, and complete with specific tourist guides.

It's a case of make full use of the up-to-date technologies about data storage and management, region monitoring and management, and of their peculiarities allowed by GIS tools in order to create and populate databases accessed from both static and moving stations, with the primary aim to assist the user, whether he's tourist or resident, sportsman or arts-loving, excursionist or nature-loving, in choosing the opportunities offered by the provincial territory giving information about accessibility, accommodation capacity and any other available supporting facilities.

M. Sebillo, G. Vitiello, and G. Schaefer (Eds.): VISUAL 2008, LNCS 5188, pp. 331–337, 2008.

2 Objective

Providing tourist information or in any case dynamic services for a better enjoyment of the surrounding places is today an objective on which many public agencies work.

Multinationals, communication businesses and big service companies have invested and still invest in this sector because users ask for data, geographic information and advanced services of regional promotion. This is supported by the ever more completed web-based experiences, the widespread diffusion of GPS navigators, the mobile services development, the strong penetration of the mobile communications and the fast equipment turnover, the emerging of voice information services.

At the starting up of GISTURISMO project the platform choice encompassed the target selection:

- The provision of GPS navigation routing services involves the use of devices provided with resident memory and hence a more difficult updating;
- The provision of services for the access to a central and easy-to-update database requires the use of devices connected to internet, but routing-performing and web-connected mobile equipment is scarcely widespread;
- The provision of multimedia services requires the use of data receiving broadband which is still scarcely widespread for mobile applications: there is an ongoing development of new platforms such as wi-max, confining them to internet which however doesn't allow routing on site, namely on your means of transport.

The Province's aim is to offer region promotion services which are advanced, dynamic, mobile and multimedia, thus fostering the development of a single data framework (the central database founded on an already owned technology which can be reused for the territory management) able to supply platforms, both static and mobile, in charge of giving information, advertising events, suggesting itineraries, supporting users in their time planning according to their interests and guiding them directly through the region.

3 Implementing Methodology

In order to implement the service, the following steps have been taken:

- Screening of the available databases from the Province and APT's relevant departments for territory promotion;
- Planning of the survey campaign with GPS and CTR10K;
- Extrapolation of basic geographic data from CTR 10K and 2K (Tuscany Region) to make a technical-tourist cartographic base;
- Implementation of the technical features which make it possible to connect several information layers and the descriptive and multimedia cartographic components up, thus tuning the new spatial data with the regional recommendations for geographic data preparation (INSPIRE Directive, GIS Agreement, Thematic DB instructions);
- Detection of historical, cultural, environmental, eno-gastronomic and sports routes according to the selected targets and their Interest Points (PDI);

- Implementation of the servlet connected with central DB to supply the single platforms with data selection, coordinate system conversion and specific xml formatting;
- Validation phase;
- Operating phase.

4 Supplied Information

Information are basically tabular and they are linked with either point or line features, depending on the detected object; though, polygon features are not excluded.

In regard to Interest Points (PDI), information usually concern geographic positioning (point feature GPS calculated and CTR10K straightened up), data about accessibility (features, timetable, disabled accesses, addresses, etc.), laid on services (special and teaching activities, spoken languages, guides, etc.), descriptive data (history, descriptions, monographs, particular features, etc.), multimedia data (pictures, videos, animations, etc.); each object typology is characterized by a number of extensions depending on its kind.

With regard to routes both geographic data about the layout (GPS detection of 3D polylines) and nearby objects and the approaching ways, technical information about the administrator, accessibility and availability information are collected; it is also possible to link more general tourist information about surroundings, environment, landscape, accommodation and shopping facilities nearby the route itself.

The degree of GeoDB's accessibility is established by the option on the platform, because each type of platform offers a specific multimedia and interactivity level according to technologies, connectivity potential or capabilities of interface and firmware.

5 Surveyed Information

During the development of the project 10 macro-categories have been identified in order to get together the region's whole tourist supply grouped in homogenous territorial objects:

- Architecture & Territory
- Arts & Culture
- Sport & Leisure
- Environment (Nature)
- Eno-gastronomy
- Accommodation
- Tourist facilities
- Entertainment
- Tourist help
- Access facilities

The project framework is open and wide enough anyway, since 100 macro-categories have been included and each of them can contain up to 100 categories each of 99.999 objects, making possible to deal with any new need or characteristic brought about by the project development.

6 A Standard Architecture

It's a long time since Regione Toscana undertook an ambitious path which has led to the formulation of technical specifications for the implementation of thematic DBs (risen as Implementing Rules of art. 4 of L.R. 5/95 and today expression of the Regional Geographic Information System Regulation (SIGR) under the clauses 28 and 29 of L.R. 1/05); these meet the specifications Intesa –GIS and INSPIRE Directive, as they have been a base platform for the working group State-Regions.

Within the SIT framework in Tuscany, Provinces have been given a central role regarding the implementation, the coordination of the Local Authorities and the archives' widespread circulation, being "GIS centres" localized across the region; from this perspective, Provincia di Massa-Carrara has laid as the core of this project a unique geodatabase which conforms to existing regulation and is designed in collaboration with ESRI Italia and Regione Toscana which manages the whole project and allows both to draw on existing databases and reuse the new data structures for sports and tourist planning activities or regional programming.

What is peculiar to this geodatabase indeed is to be thought and managed as a usual group of territorial information layers which complete and enlarge the Territorial Information System of the province normally used by the Authority to manage the territory and design non-local planning instruments, but at same time source of data for context analysis of local urban plans and programmes.

A geodatabase ESRI Enterprise has been chosen: it is hosted on the Province's Map Server equipped with Oracle 10g R2, ArcGIS Server 9.2, ArcIMS 9.2 and ArcINFO 9.2, but the project is fully operational also on a FileGeodatabase data framework. This architecture manages both data acquisition through already in usage ArcGIS, the publication of the map services on WEB portal and PDA or smartphone accesses, the descriptive contents of the 3 portals' pages, and the extra contents of the satellite navigation platform. The data selecting servlet for the platforms is integrated in a special new ARCGIS menu.

7 Multiplatform System

7.1 Web Desktop Platform

A multilingual website in which all the necessary information for the platform use are gathered. Development of many registration forms aimed at recording and characterizing the users for newsletter or sms.

The principal section of the portal in cms standard format is realized with DotNetNuke asp.net 2.0 open source structure and it is completed by map server ArcIMS and Ajax technologies in order to offer a better netsurfing which allows the publication of all the geographic information and enables interesting synergies among the thematic DB hosted on the Oracle DB server with which is linked.

Developing special servlets and customizing the geodatabase, the web platform is shared with a minimum resolution of 1024x768 and it is the most rich in contents:

- Cross research engines for all kind of collected categories by means of a dynamic DB reading;
- Research lists for each category of objects by means of a dynamic DB reading;
- From the engine and the research lists access to synthetic information files about each collected object by means of a dynamic DB reading;
- For each synthetic file link to the map, to institutional web-portals and to dedicated web-pages;
- Geographic research engine equipped with routing functions to plan the journey;
- Cross geographic query for all the collected categories by means of a dynamic DB reading;
- Itineraries' description;
- Panoramic pictures;
- Regional guides;
- Guided itineraries;
- Free copy of the main publications.

7.2 Web Mobile Platform

The service is implemented on palmtops and smartphones equipped with operating system Microsoft Windows Mobile and Symbiam, by means of a dedicated portal with a resolution of 240 pixel.

Special applications, similar to those web-distribuited for desktop pc, allow the user to get to the whole project's database in a dynamic way.

The main architecture is in html with aspx functions for the DB queries and .net functions to get to the cartography.

This platform, to which the user is automatically redirected once the portal detects the specific browser typology, has been widely tested on windows mobile 2003 se devices, windows mobile 5 and 6 devices and Symbian platforms.

The mobile platform supplies the following services:

- Cross research engine for all kind of collected categories by means of a dynamic DB reading;
- Access to synthetic information files about each collected object by means of a dynamic DB reading;
- For each synthetic file link to the map, to institutional web-portals and to dedicated web-pages;
- Regional guides;
- Guided itineraries;
- Access to the events news through the integration with the portals of Provincia di Massa-Carrara and APT;
- Access to the newsletter.

7.3 Wap XHTML Platform

The service is implemented on any cellular phone provided with wap xhtml connection and colour screen, through a dedicated portal with a resolution of 240 pixel. The access will be possible also from low resolution terminals but getting poor images quality.

Special applications, similar to those web-distributed for desktop pc, allow the user to get to the database of recorded territorial categories of objects in a dynamic way.

The main architecture is in html with functions for the DB queries.

This platform, to which the user is automatically redirected once the portal detects the specific browser typology, has been widely tested on cellular phones from several producers.

The wide range of available phone terminals makes it impossible to test all the models, anyway the taken tests highlight that the only difference is about the character replacement.

The wap platform supplies the following services:

- Cross research engine for all kind of collected categories by means of a dynamic DB reading;
- Access to synthetic information files about each collected object by means of a dynamic DB reading;
- Access to the events news through the integration with the portals of Provincia di Massa-Carrara and APT.

7.4 PND EASYMAP DeAgostini Platform

The platform is made available through GPS satellite navigators tailored with the project's contents and which are put at disposal as commodatum at the Massa-Carrara APT information points and a selection of accommodation facilities involved in the testing. All the recorded and georeferenced categories are listed within the interest points, together with the information files also available on other platforms.

The navigator supplies the following services:

- Italia Teleatlas road graph;
- Opportunity of setting the language of the user interface: Italian, English, German and French;
- Satellite navigator equipped with auto-routing and vocal guide functions;
- Research for complete addresses, interest points or simply places;
- Places, addresses, interest points, GPS obtained points storage;
- Research for interest points either in general or just along the itinerary or only nearby the destination;
- Selection of interest points either by name or by proximity;
- Route calculation;
- Management of the tourist guides system interconnected with the routing functions.

7.5 GPS MOBILE DeAgostini Platform

Any user equipped with a palmtop or a smartphone, whose operating system is either Windows Mobile or Symbian, as long as provided with an integrated or external satellite aerial, will be able to download a software developed by Istituto Geografico DeAgostini (Novara) which can replicate the Easymap © system functions, allowing routing functions within the whole province's territory through the Teleatlas © road graph.

The navigation system makes for loading all the interest points recorded within the project together with all the other already available information, allowing to freely use the territory guides which are linked to the routing functions.

The mobile navigation software supplies the following functions:

- Italia Teleatlas road graph;
- Opportunity of setting the language of the user interface: Italian, English, German and French;
- Satellite navigator equipped with auto-routing and vocal guide functions;
- Research for complete addresses, interest points or simply places;
- Places, addresses, interest points, GPS obtained points storage;
- Research for interest points either in general or just along the itinerary or only nearby the destination;
- Selection of interest points either by name or by proximity;
- Route calculation;
- Management of the tourist guides digital system interconnected with the routing functions.

7.6 GPS Platform

The user who already has his own satellite navigator equipped with updating PDI, can free download from the portal each project recorded points categories already prearranged according to the standards of the main available producers: Tomtom, Garmin, Navio, Mio, Destinator and Navman.

Further exchange formats are also free available: Google Earth KML, GPS Exchange Format GPX and Smart Path Binary OVR.

Author Index