George A. Tsihrintzis
Maria Virvou
Lakhmi C. Jain
Robert J. Howlett
Editors

Intelligent Interactive Multimedia Systems and Services

Proceedings of the 4th International Conference
on Intelligent Interactive Multimedia Systems
and Services (IIMSS´2011)

KES

Springer

George A. Tsihrintzis, Maria Virvou, Lakhmi C. Jain, and Robert J. Howlett (Eds.)

Intelligent Interactive Multimedia Systems and Services

Smart Innovation, Systems and Technologies 11

Editors-in-Chief

Prof. Robert J. Howlett
KES International
PO Box 2115
Shoreham-by-sea
BN43 9AF
UK
E-mail: rjhowlett@kesinternational.org

Prof. Lakhmi C. Jain
School of Electrical and Information Engineering
University of South Australia
Adelaide, Mawson Lakes Campus
South Australia SA 5095
Australia
E-mail: Lakhmi.jain@unisa.edu.au

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

George A. Tsihrintzis, Maria Virvou, Lakhmi C. Jain,
and Robert J. Howlett (Eds.)

Intelligent Interactive Multimedia Systems and Services

Proceedings of the 4th International Conference on
Intelligent Interactive Multimedia Systems and
Services (IIMSS 2011)

Springer

Prof. George Tsihrintzis
University of Piraeus
Department of Informatics
80 Karaoli & Dimitriou St.
18534 Piraeus
Greece
E-mail: geoatsi@unipi.gr

Prof. Maria Virvou
University of Piraeus
Department of Informatics
80 Karaoli & Dimitriou St.
18534 Piraeus
Greece
E-mail: mvirvou@unipi.gr

Prof. Lakhmi C. Jain
University of South Australia
School of Electrical and Information
Engineering
Mawson Lakes Campus
Adelaide
South Australia SA 5095
Australia
E-mail: Lakhmi.Jain@unisa.edu.au

Prof. Robert J. Howlett
KES International
PO Box 2115
Shoreham-by-sea West Sussex
BN43 9AF
United Kingdom
E-mail: rjhowlett@kesinternational.org

Preface

This volume contains the Proceedings of the 4th International Conference on Intelligent Interactive Multimedia Systems and Services (IIMSS-2011). IIMSS-2011 comes as a sequel to IIMSS-2008 (Piraeus-Athens, Greece, July 9, 10 and 11, 2008), IIMSS-2009 (Mogliano Veneto (near Venice), Italy, July 15, 16 and 17, 2009) and IIMSS-2010 (Baltimore, USA, July 28, 29, and 30, 2010). This fourth edition of the IIMSS Conference was organized jointly by the Department of Informatics of the University of Piraeus, Greece and the School of Electrical and Information Engineering of the University of South Australia, in conjunction with KES International. IIMSS is a new series of international scientific conferences aimed at presenting novel research in the fields of intelligent multimedia systems relevant to the development of a new generation of interactive, user-centric services.

At a time when computers are more widespread than ever and computer users range from highly qualified scientists to non-computer-expert professionals and may include people with special needs, *interactivity*, *personalization* and *adaptivity* have become a necessity in modern multimedia systems. Modern intelligent multimedia systems need to be interactive not only through classical modes of interaction where the user inputs information through a keyboard or mouse. They must also support other modes of interaction, such as visual or lingual computer-user interfaces, which render them more attractive, user friendlier, more human-like and more informative.

On the other hand, the solution of "one-fits-all" is no longer applicable to wide ranges of users of various backgrounds and needs. Therefore, one important goal of many intelligent multimedia systems is their ability to *adapt dynamically* to their users.

To achieve these goals, intelligent multimedia systems need to evolve at all levels of processing. This includes further research and development of low-level data processing for information security, compression, transmission, clustering, classification and retrieval. This research leads into the development of new and more efficient intermediate-level intelligent multimedia systems for such applications, as information tracking, human and object monitoring, home and office automation, environmental information systems, or systems for rights management and licensing. Such intermediate-level intelligent multimedia systems are the building blocks of high-level intelligent multimedia services for such application areas, digital libraries, e-learning, e-government, e-commerce, e-entertainment, e-health, or e-legal services.

Multimedia Services based on multimedia systems have made significant progress in recent times, as they arise in various areas including, but not limited to, advertisement, art, business, creative industries, education, entertainment, engineering, medicine, mathematics, scientific research and other applications. The growth rate of multimedia services has become explosive, as technological progress is forced to match consumer hunger for content.

A typical example of multimedia service growth is the activity in mobile software. The term *mobile services* refers to services requested by a user of a mobile network through his/her mobile device. Early stage mobile services were centered on voice

communication (i.e., telephony) and were available only in telecommunication networks. However, with consumers requesting that modern mobile devices offer more than just voice communication, mobile software has already begun to provide access to a vast array of data and services. In general, mobile services, can be classified in several ways such as *voice/call services* (e.g. call forwarding, call waiting, calling line identification, missed call notification, or conference call service), *location-based services* (e.g. locating a street address or business or providing navigation information), transaction services (e.g. mobile banking or payments made directly over mobile phones), *advertising services*, and *multimedia mobile services*.

Mobile multimedia services provide means for delivering multimedia content and information between two mobile stations or between a mobile station and an operator. Today, mobile technology is capable of offering its users commercial opportunities for creation and sharing of multimedia. Thus, a variety of mobile multimedia services give users the ability to use their device not just for telecommunication purposes, but also for a variety of other purposes, such as entertainment, learning, medicine, or advertisement.

IIMSS conferences, following the general structure of KES events, aim at providing an internationally respected forum for presenting and publishing high-quality results of scientific research while allowing for timely dissemination of research breakthroughs and novel ideas via a number of autonomous special sessions and workshops on emerging issues and topics identified each year.

IIMSS-2011 is co-located with the 3rd International Conference on Intelligent Decision Technologies (IDT-2011).

We are very satisfied of the quality of the program and would like to thank the authors for choosing IIMSS as the forum for presentation of their work. Also, we gratefully acknowledge the hard work of IIMSS international program committee members and of the additional reviewers for selecting the accepted conference papers. Finally, we are indebted to the staff at Springer for their wonderful working in publishing the IIMSS-2011 proceedings.

General Co-Chairs

Prof.-Dr. George A. Tsihrintzis, University of Piraeus, Greece
Prof.-Dr. Maria Virvou, University of Piraeus, Greece
Prof.-Dr. Lakhmi C.Jain, University of South Australia, Australia

Executive Chair

Prof.-Dr. R. J. Howlett, University of Bournemouth, UK

Liaison co-Chairs

Prof. Toyohide Watanabe , Nagoya University, Japan
Prof. Valentina Zharkova, University of Bradford, UK
Dr. Lyudmila Mihaylova , Lancaster University, UK

Programme Coordinator

Dr. Marco Anisetti, University of Milan, Italy

General Track Coordinator

Dr. Valerio Bellandi, University of Milan, Italy

International Program Committee

IPC Co-chairs

Dr. George A. Tsihrintzis

Professor
Department of Informatics
University of Piraeus
Piraeus 185 34, Greece
geoatsi@unipi.gr

Dr. Maria Virvou

Professor
Department of Informatics
University of Piraeus
Piraeus 185 34, Greece
mvirvou@unipi.gr

Prof. Lakhmi C. Jain

PhD, ME, BE (Hons), Fellow (Engineers Aust),
Professor of Knowledge-Based Engineering,
Founding Director of the KES Centre,
School of Electrical and Information Engineering,
University of South Australia,
Adelaide,
Mawson Lakes Campus,
South Australia SA 5095,
Australia
Lakhmi.jain@unisa.edu.au
www.unisa.edu.au/kes

IPC Members

Anisetti, Marco	University of Milan, Italy
Asakura, Koichi	Daido University, Japan
Banerjee, Amit	Pennsylvania State University, USA
Bannore, Vivek	ThinkingSpace IT Solutions, Australia
Bianchini, Monica	University of Siena, Italy
Blanco-Fernandez, Yolanda	University of Vigo, Spain
Choras, Ryszard S.	University of Technology & Life Sciences, Poland
Correa da Silva, Flavio Soares	University of Sao Paulo, Brazil
Cunningham, Stuart	Glyndyr University, UK
Cuzzocrea, Alfredo	ICAR-CNR & University of Calabria, Italy
Dascalu, Sergiu	University of Nevada, USA
De Pietro, Giuseppe	ICAR-CNR, Italy
Gallo, Luigi	ICAR-CNR, Italy
Georgieva, Petia	University of Aveiro, Portugal
Grout, Vic	Glyndwr University, UK
Hatzilygeroudis, Ioannis	University of Patras, Greece
Horng, Mong-Fong	National Kaohsiung University of Applied Sciences, Taiwan
Kabassi , Katerina	Technological Educational Institute of the Ionian Islands, Greece
Kountchev, Roumen	Technical University of Sofia, Bulgaria
Lopez-Nores , Martin	University of Vigo, Spain
Matijasevic, Maja	University of Zagreb, Croatia
Matsubara, Shigeki	Nagoya University, Japan
Montani, Stefania	Universita' del Piemonte Orientale , Italy
O'Shea, James D.	Manchester Metropolitan University, UK
Ogiela, Marek	AGH University of Science and Technology, Poland
Patel, Nilesh	Oakland University, USA
Pazos-Arias, Jose J.	University of Vigo, Spain
Ronsin, Joseph	INSA/IETR-Rennes, France
Szmidt, Eulalia	Polish Academy of Sciences, Poland
Tateiwa, Yuichi	Nagoya Institute of technology, Japan
Tiako, Pierre F.	Langston University, USA
Watanabe, Toyohide	Nagoya University, Japan

Invited Speakers

Prof. Peter Brusilovsky	School of Information Sciences, University of Pittsburgh, USA
Prof. Petar Djuric	Stony Brook University, New York, USA
Prof. Janusz Kacprzyk	Polish Academy of Sciences, Warsaw, Poland
Prof. George Pavlou	University College London, UK
Dr. Arvind Ramadorai	iRobot Corporation, USA

Contents

Managing Collaborative Sessions in WSNs

Laura M. Rodríguez Peralta, Lina M.P.L. Brito, and João F.F. Santos

Competence Centre of Exact Sciences and Engineering
University of Madeira (UMa)
Madeira, Portugal
{lmrodrig,lina}@uma.pt, joao_santos87@hotmail.com

Abstract. In Wireless Sensor Networks (WSNs), the sensor nodes are typically resource limited. This fact fosters the nodes to collaborate in order to implement their tasks. Therefore, WSNs are, inherently, collaborative networks. In this paper, we propose and implement a model that represents the various types of collaboration relationships that can be established in a WSN. This involves identifying and analyzing the different types of collaboration that may occur in any WSN. As a result, we propose a hierarchy composed by all types and different levels of collaboration, and we propose a collaborative session management tool, called WISE-MANager. This tool allows bringing these concepts into practice, more precisely to the establishment of collaborative sessions. WISE-MANager increases the WNS's flexibility and the interaction between the user and the network.

1 Introduction

In this paper, we enhanced the CWSN model by defining a hierarchy of collaboration for WSNs. By bringing the main CSCW (Computer Supported Cooperative Work) concepts into the area of WSNs, this work includes identifying and describing the different types and levels of collaboration, as well as the collaborators that can exist within a WSN (which can be mainly classified in two types: humans and nodes).

By collaboration we refer to any collaboration relationships that may be established between any two components of a WSN. Aiming to bring these concepts into practice, we have programmed ZigBee-based sensor nodes in order to establish collaborative sessions in a real WSN.

This paper is organized as follows. Section 2 presents the related work. In section 3, the different types of collaboration are identified and described, and a hierarchy of collaboration is proposed. Then, a definition of session and its classification is proposed. Section 4 describes the WISE-MANager tool and its implementation. Section 5 provides some conclusions and some perspectives of future work.

2 Related Work

Most of works in the literature, that specifically focus collaboration in WSNs cover a specific type of collaboration, which is associated with the accomplishment of a certain

task, namely: signal processing [2], sensing [3], routing, localization [4], security [5], task scheduling [6], etc. Other works concern collaboration between wireless sensor nodes and other kind of devices (heterogeneous groupware collaboration) to support some specific applications.

The work proposed by Liu et al. [7], called SNSCW (Sensor Networks Supported Cooperative Work), is the only one which presents a model for collaborative work in sensor networks. It is a hierarchical model that essentially divides cooperation in sensor networks in two layers; the first one relates to cooperation between humans and sensor nodes; the second one relates to cooperation between the sensor nodes.

However, the SNSCW model only allows the modeling of collaboration itself. On the contrary, the CWSN model, which has been presented in [1], is a formal model that was created specifically to describe WSNs. However, the CWSN model allows not only the modeling of collaborative work (based in CSCW concepts), but also the modeling, formalization and visual representation of all the entities that can constitute a WSN (different types of nodes, clusters, relationships, sessions, obstacles, etc.), as well as its properties. Moreover, it considers more than two layers of collaboration.

The CWSN model formalizes all the properties of each entity through first-order logic. Even though it is a graph-based model, it includes other objects [1] in order to make the modeling of all the entities of a WSN possible. This is of paramount importance to completely represent a WSN.

3   Collaboration Hierarchy in WSNs

In this paper, we present a new approach that brings some of the fundamental CSCW concepts into the world of WSNs. The CWSN model is, then, improved in order to represent not only all the entities that can compose a WSN and its attributes, but also to fully represent collaboration in a WSN. So, on one hand, we enrich the CWSN model with the most important CSCW concepts, such as: participants, relationships, roles, tasks, sessions and groups. On the other hand, we identify several levels of collaboration, going beyond the two levels defined by the SNSCW model [8]. Thus, we extend the CWSN model with a hierarchical representation of collaboration.

By collaboration we denote any collaboration relationships that may be established between any two entities of a WSN. It may refer to collaboration involved in transmission of data between any two entities of the network, or to the collaboration required so that nodes can perform the majority of their tasks, which is a consequence of their severe resource limitations. However, the types of collaboration are determined by the types of nodes that exist in a WSN, since each type of node has its specific tasks. Consequently, the different types of collaboration that exist in a WSN are a natural result of its inherent hierarchy. For example, only the sink node can send data to the user; consequently, all the other nodes have to forward data towards the sink node; therefore, the sink node is naturally on the higher levels of the WSN hierarchy.

Even though the collaboration activities in a WSN depend essentially on the type of nodes involved in a collaboration relationship, which results in different levels of collaboration, other reasons for collaborating can be identified. In WSNs, collaboration typically occurs among nodes located in a certain region, which means that the group of nodes may not always be the same (location-centric collaboration), opposing to

traditional networks (node-centric collaboration). Nevertheless, besides location-based collaboration, it is possible to identify other ways to collaborate, based whether in monitoring a common phenomenon or in the hardware characteristics of the nodes themselves.

3.1 Types of Collaboration

Analyzing collaboration in WSNs from the point of view of relationships and interactions established among collaborators (or participants), we can identify essentially three main types of collaboration [8]: 1) collaboration between the user and the WSN, 2) collaboration among nodes, and 3) collaboration between sessions.

3.2 Collaboration Hierarchy

Fig. 1 illustrates these different types of collaboration. Analyzing the figure, it is rather intuitive to conclude that WSNs present a hierarchy of collaboration relationships. This hierarchy can be composed by different levels of collaboration, as represented in Fig. 1: the node level; the cluster level; the session level; the network level; and the user level. Besides, according to the types of collaboration presented above, collaboration can occur either within a certain level (horizontal collaboration) or between each two consecutive levels (vertical collaboration). In the case of heterogeneous WSNs, even though it is not represented in Fig. 1, the wireless devices will be represented in the Node Level of this collaboration hierarchy.
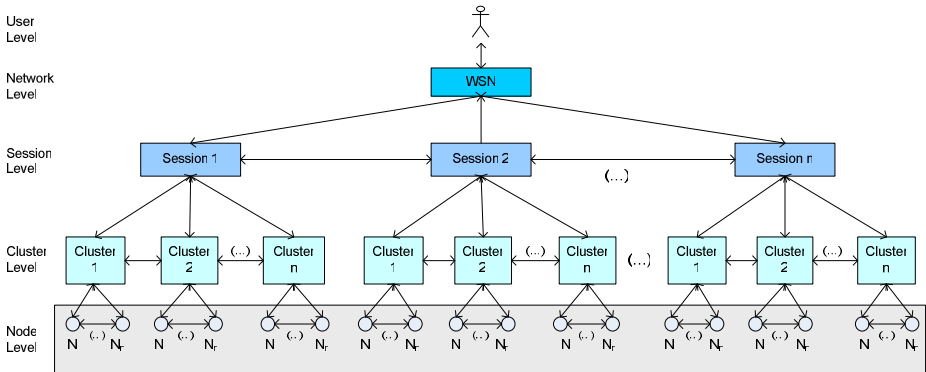
Fig. 1. Multi-level collaboration hierarchy within a WSN

3.3 Sessions

In this work, we propose a definition of session as the essential unit of a collaborative activity in WSNs. A session is created based on different objectives defined by the user, such as: the type of phenomenon to monitor, the geographical area to monitor, the monitoring period, etc.

In the context of WSNs and considering CSCW definitions, sessions are composed by participants, the collaboration relationships and the respective data flows established among them, and the tasks of each participant.

In a session, all types of collaboration relationships mentioned before can exist; therefore, several different collaborative groups can be established inside a session.

Concerning the state of the nodes that constitute a session, a session can be classified in one of four states, as represented in Fig. 2:

- **Created** – The session has been created but not initiated; i.e., the session is not in the open state yet. This is the first state of a session.
- **Open** – While the objective of the session is not fulfilled and some nodes are active, the session maintains its activity, i.e., it is open.
- **Close** – A session can become closed due to one of three possible motives: 1) when all nodes go into sleep mode; or 2) when all nodes are damaged or fail; or 3) when a temporary interruption of the session occurs (i.e., the session stops for a certain time interval that is settled by the user).
- **End** - A session ends due to one of three possible causes: 1) when the objective of the session is fulfilled; 2) when the predefined lifetime of a session comes to an end; or 3) when the session is aborted by the user (through the transmission of some command).
- **Deleted** – This state occurs when the user deletes the session and all the respective data.
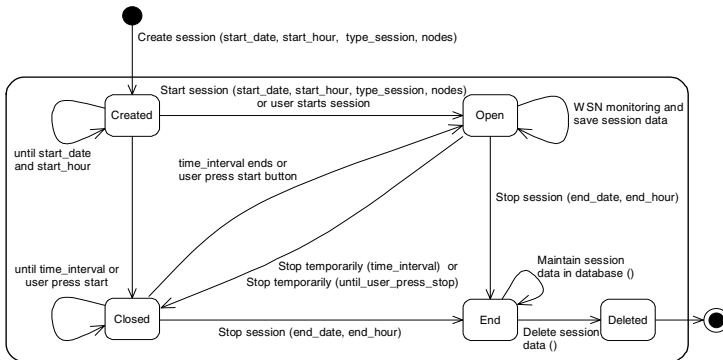
Fig. 2. State transition diagram of a session

Depending on the WSN specific application, sessions can be classified according to their temporal characteristics, as:

- *Parallel sessions* – Sessions that take place at the same time.
- *Sequential sessions* – A particular session starts only after another session ends.
- *Synchronous sessions* – The occurrence of these sessions is planned by the network manager. Parallel and sequential sessions can also be classified as synchronous sessions.

- *Asynchronous sessions* – The occurrence of these sessions is not planned by the network manager. Rather, they can be started by some action (user initiated or node initiated), by the detection of an unexpected change in a certain phenomenon, etc.

Defining a collaborative session with all its possible states, and establishing temporal relationships between collaborative sessions is important to allow other researchers to better understand the operation of a WSN. These concepts can also help researchers to develop management tools that optimize the network management and that can make it more flexible.

4 WISE-MANager

In order to implement and validate the CWSN model, we have implemented a collaborative sessions' management tool, called WISE-MANager (WIreless SEnsor networks MANager for collaborative sessions). The WISE-MANager tool allows creating, monitoring and managing collaborative sessions using the Zigbee protocol. The purpose of using collaborative sessions is to provide a better interaction between the user and the WSN, since the user can customize the type of monitoring to be carried out (sensor node, phenomenon, or time interval of monitoring), and query the network and its components. This way, the WISE-MANager tool increases the flexibility of the WSN.

It is important to note that the proposed tool was developed in the context of the WISE-MUSE project [9].

Also note that the ZigBee protocol defines three types of devices: end devices, routers and coordinators. End devices correspond to sensor nodes with sensing capabilities, routers are sensor nodes can also sense data but they are essentially responsible for routing data collected by the end devices in the their WPAN to the coordinator, and, finally, the coordinator corresponds to the sink node.

The WISE-MANager tool was implemented in Java and it is Zigbee-compliant. It is composed of two main modules: (i) Collaborative Sessions Management; and (ii) WSN Management.

The first module, collaborative sessions management, allows creating and managing collaborative sessions inside a WSN. Users can configure the session's parameters (id, description, etc.), the sensor nodes that will make part of that session, the monitoring period, and the phenomena to be monitored (Fig. 3).After creating the session, the user can visualize and change the session's parameters. Moreover, he can also start and stop the session's monitoring at any moment, monitor the sessions that are in an "open" state, and delete them. Thus, sessions can be opened manually by the user or automatically according to the session's monitoring schedule.

Moreover, the user can export the session's data to a MS Word document, choosing the session and the monitoring time interval. The document will contain the session's parameters as well as the data received during the session.

In the second module, named WSN Management, the user can choose a serial port where the WSN's coordinator is connected. Using this module, the user can see the WSN information, like the PAN ID, the network channel, and the network components (routers, end devices, coordinator, etc.) and its parameters. Moreover, the user can modify the device's identifier (Fig. 4).
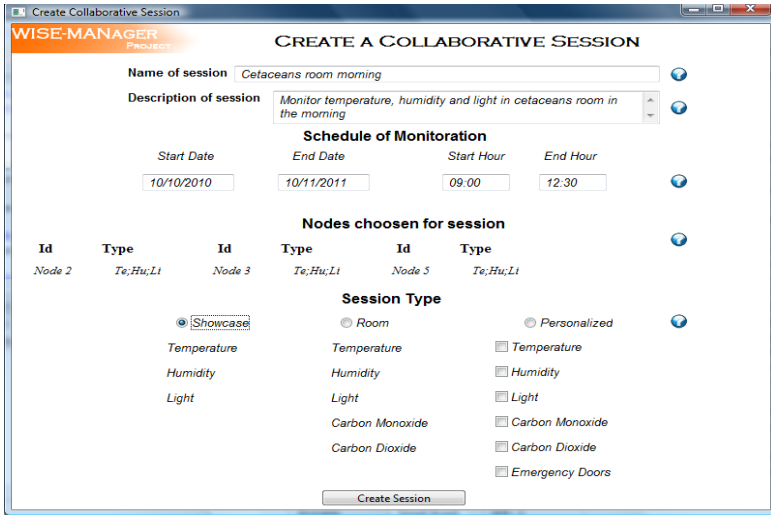
Fig. 3. Creating collaborative sessions using the WISE-MANager tool
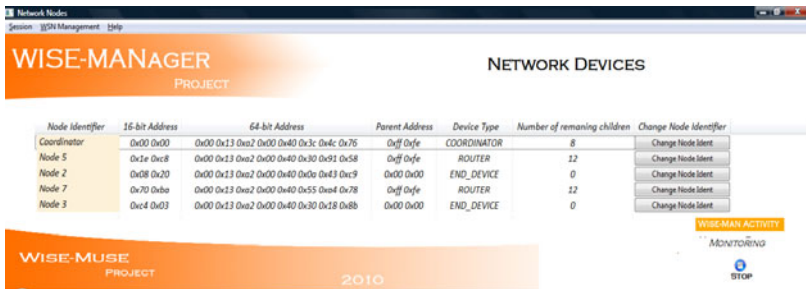
Fig. 4. Detecting WSN devices

4.1   Case Study

In order to validate the WISE-MANager tool, we have applied it to a heterogeneous network, which is composed of sensor nodes and other wireless devices that detect the state of the emergency doors at the Whale Museum situated in Madeira Island, Portugal.

Several experiments were carried out to validate the proposed tool.

One of the experiments consisted in the deployment of a WSN composed of six nodes (one coordinator, three routers, and two end devices). The main goal of this experiment was the monitoring of two separated exhibition rooms, as illustrated in Fig. 5.

In this experiment, two different sessions were created, one for each exhibition room. The first session is composed of sensor node 3 while the second one is composed of sensor node 2. Both sessions monitor temperature, light and humidity, and they were executed in parallel for nine hours, more precisely from 23:00 to 8:10 am. The sensor nodes were programmed to measure and send data once each 10 seconds.

Another experiment conducted at the museum was made to test the emergency door device inside the WSN. Thus, we created two sessions inside the WSN: (i) session 3 composed of node 3; and; (ii) session 4 composed of node 2 and node 6. This network is shown in Fig. 6. Fig. 7 illustrates physical location of the WSN inside the museum. In this figure, the whole WSN is represented using the CWSN model.
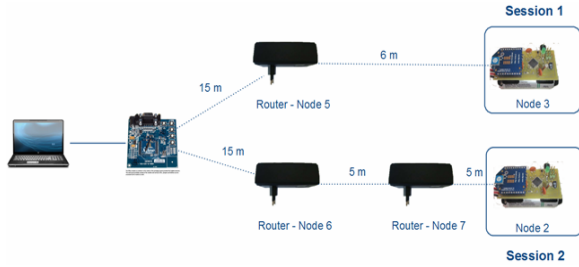
Fig. 5. WSN devices for the first experiment conducted in the Madeira Whale Museum
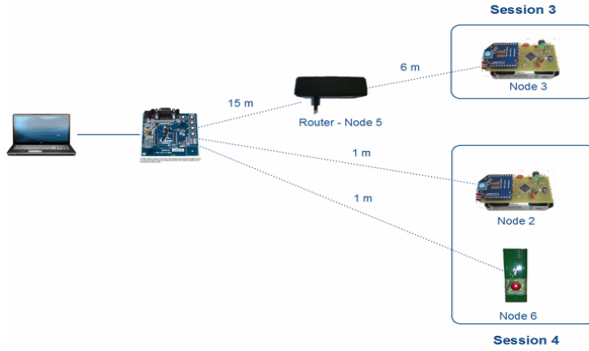
Fig. 6. WSN devices for the second experiment conducted in the Madeira Whale Museum
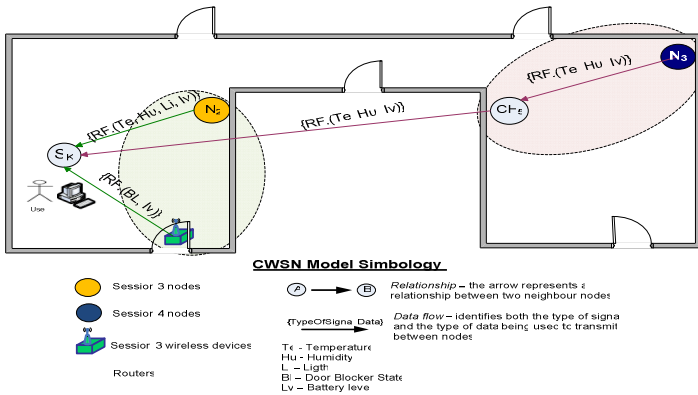
Fig. 7. Physical location of the second heterogeneous WSN

In this experiment, both sessions were executed in parallel, monitoring two different exhibition rooms. Session 3 monitored temperature and humidity, while session 4 monitored temperature, humidity, light, and the emergency door state changes (emergency, open or close) from node 6, which was installed inside the emergency door's blocker, as depicted in Fig. 7.

After these experiments, we verified that the WISE-MANager tool was able to create, start, close, end and delete sessions. Furthermore, all the data sent by the sessions was collected by the tool without any packet loss.

4.2   Advantages and Disadvantages

Analyzing Table 1, we can verify that in terms of querying the WSN, most of the tools are able to do it. On the other hand, none of these tools can create or use collaborative sessions to manage the WSN. The WISE-MANager tool allows customizing the monitoring activity and defining the session's parameters.

Moreover, the WISE-MANager tool, allows the user to control the network and inquire the WSN, getting information like communication channel, network ID, PAN ID, etc. It is also possible to detect the network's devices and change their identifiers.

Through collaborative sessions, the WISE-MANager tool enhances collaboration between the user and the network. Thus, the network is more flexible since the user can customize the collaboration, choosing different nodes to monitor different phenomena, and the monitoring time interval. Therefore, the network topology can be dynamic, since nodes can be active or inactive, according to the collaborative session's state. Additionally, this feature allows the energy saving of the network nodes.

Table 1. Comparing WISE-MANager with related solutions

Tools	Query WSN	WSN Management	Create Sessions	View Sessions	Monitor Sessions
Tiny DB [10]	No	No	No	No	No
MonSense [11]	Yes	Yes	No	No	No
Mote-View [12]	Yes	Yes	No	No	No
MANNA [13]	Yes	Yes	No	No	No
BOSS [14]	Yes	Yes	No	No	No
WISE-MANager	Yes	Yes	Yes	Yes	Yes

5   Conclusions

The CWSN model, published in [1], was specifically designed for WSNs and it is based in the CSCW methodology. It is a graph-based model, which great advantage is that, besides modeling collaboration, it also allows for modeling the entire WSN, all its entities, properties, relationships, states, etc. The CWSN model can be used, for example, as a framework for high-level configuration or programming tools (usually know

by macro-programming). Moreover, the CWSN model can be used to automatically generate some graphs of the WSN that will allow for identifying routing paths, detecting damaged/failed nodes or links, etc.

In this paper, we have enhanced the CWSN model by proposing a hierarchy of collaboration that identifies the different types and levels of collaboration that might exist within a WSN. We also correlated these types of collaboration with the different entities that can constitute a real WSN. Moreover, we have proposed a hierarchical model of collaboration that brings the CSCW to WSNs. This work allowed us to conclude that the collaboration hierarchy, which is composed by distinct collaboration levels, is a result of the distinct roles that the different entities play in a WSN. A major advantage of the hierarchical modeling of collaboration is that it can be used by other researchers as a framework to describe the collaboration relationships established in any WSN, despite its particular application.

References

[1] de Brito, L.M.P.L., Rodríguez Peralta, L.M.: A collaborative model for wireless sensor networks applied to museums' environmental monitoring. In: Luo, Y. (ed.) CDVE 2008. LNCS, vol. 5220, pp. 107–116. Springer, Heidelberg (2008)

[2] Ramanathan, P., Saluja, K., Hu, Y.: Collaborative Sensor Signal Processing for Target Detection, Localization and Tracking. In: 23rd Army Science Conference, Orlando, USA (2002)

[3] Wang, K.-C., Ramanathan, P.: Collaborative Sensing Using Sensors of Uncoordinated Mobility. In: Prasanna, V.K., Iyengar, S.S., Spirakis, P.G., Welsh, M. (eds.) DCOSS 2005. LNCS, vol. 3560, pp. 293–306. Springer, Heidelberg (2005)

[4] Dardari, D., Conti, A.: A Sub-Optimal Hierarchical Maximum Likelihood Algorithm for Collaborative Localization in Ad-Hoc Networks. In: IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks (SECON 2004), Santa Clara, USA, pp. 425–429 (2004)

[5] Chadha, A., Liu, Y., Das, S.: Group Key Distribution via Local Collaboration in Wireless Sensor Networks. In: IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks (SECON 2005), Santa Clara, USA, pp. 46–54 (2005)

[6] Sanli, H., Poornachandran, R., Cam, H.: Collaborative Two-Level Task Scheduling for Wireless Sensor Nodes with Multiple Sensing Units. In: IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks (SECON 2005), Santa Clara, USA, pp. 350–361 (2005)

[7] Liu, L., Ma, H., Tao, D., Zhang, D.-M.: A hierarchical cooperation model for application self-reconfiguration of sensor networks. In: Shen, W., Luo, J., Lin, Z., Barthès, J.-P.A., Hao, Q. (eds.) CSCWD. LNCS, vol. 4402, pp. 33–42. Springer, Heidelberg (2007)

[8] Rodríguez Peralta, L.M., Leão Brito, L.M.P., Teixeira Gouveia, B.A.: The WISE-MUSE Project: Environmental Monitoring and Controlling of Museums based on Wireless Sensors Networks. EJSE – Electronic Journal of Structural Engineering, 46–57 (2009); Special Issue on Sensor Network for Building Monitoring: From Theory to Real Application; ISSN:1443-9255

[9] Madden, S., et al.: TinyDB: An Acquisitional Query Processing System for Sensor Networks. Massachusetts Institute of Technology, USA (2005)

[10] Pinto, J., et al.: MonSense-Application for Deployment, Monitoring and Control of Wireless Sensor Networks. Faculdade de Engenharia da Universidade do Porto, Portugal
[11] WinnieL., et al.: Network Management in Wireless Sensor Networks. School of Computer Science & Software Engineering, The University of Western Australia, Australia
[12] Ruiz, L.B., Nogueira, J.M., Loureiro, A.A.F.: MANNA: A Management Architecture for Wireless Sensor Networks. IEEE Communications Magazine (2003)
[13] Song, H., Kim, D., Lee, K., Sung, J.: Upnp-Based Sensor Network Management Architecture. In: Proc. 2nd Int. Conference on Mobile Computing and Ubiquitous Networking, pp. 85–92 (2005)

OGRE-Coder: An Interoperable Environment for the Automatic Generation of OGRE Virtual Scenarios

Duarte M. Fernandes[1], Duarte J.O. Teixeira[2], and Paulo N.M. Sampaio[3]

Madeira Interactive Technologies Institute (M-ITI)
University of Madeira (UMa)
Funchal, Madeira, Portugal
dmf9000@yahoo.com, duartejoaoornelas@hotmail.com,
psampaio@uma.pt

Abstract. Currently, there are several languages and tools that enable the creation of virtual scenarios. However, existing approaches are not intuitive and require a thorough knowledge of the user. The work presented in this paper aims at addressing this gap by the development of an intuitive framework for the creation of OGRE virtual scenarios with high graphical quality, customized multimedia presentations and supporting the distributed navigation and use of collaborative communication tools, such as VoIP (Voice over Internet Protocol), chat, etc. The proposed framework relies on the utilization of a generic metadata for the description of virtual scenarios which can be applied by different tools for authoring, and which can facilitate the subsequent automatic generation of OGRE code. With this approach OGRE developers can focus on the codification of the dynamics and strategies of the application being developed which helps reducing considerably the development time of these applications.

Keywords: 3D Modeling, OGRE, Virtual Reality, XML, Authoring Tool, Virtual Scenarios, OGRE-CL, OGREML, OGRE-Coder.

1 Introduction

Currently, there are different languages and tools available for the development of virtual worlds. Nevertheless, the existing solutions are not intuitive and require from the developer a deeper knowledge of their components, representation and structure in order to be able to create a virtual environment. Furthermore, most of the existing solutions still lack the possibility of creating more complex virtual scenarios with: a high graphical quality; the possibility to integrate multimedia presentations inside the virtual world; the development of distributed applications which are able to provide remote navigation or communication tools such as *VoIP* (*Voice over Internet Protocol*), *chat*, etc., among other advanced features.

One of the existing solutions for the development of Virtual Reality applications is the utilization of the graphical rendering engine OGRE (Object-Oriented Graphics Rendering Engine) [1]. OGRE is a scenario oriented 3D engine written in C ++ which

provides a library of classes (APIs) for the description of virtual worlds and objects in a high abstraction level and graphical quality.

Based on a literature review, some of the most promising OGRE editors available are Ogitor [2], OGRE Editor Multi Scene Manager Project Environment [3] and OGRE – MOGRE Editor [4]. These applications are in general well conceived and simple in order to allow the creation of virtual scenarios and they can be found in different states of completeness, sometimes being able to adapt themselves to new APIs and libraries. However, most of these tools have a proprietary representation for the virtual scenario created, do not allowing exporting the respective OGRE code for further implementations.

Therefore, the main goal of this paper is to introduce and discuss the development of OGRE-Coder [5] which is a graphical tool for providing the visualization of OGRE virtual scenarios and the automatic generation of its respective OGRE source code. Another important contribution of this paper is outlined by the presentation of an XML-based language called OGREML [5] which was proposed for providing the interoperability among authoring tools. Thus, virtual scenarios can be created with any OGRE-compliant tool, and then exported to OGREML for the subsequent automatic generation of OGRE code. This approach allows developers to optimize the prototyping of their 3D applications since the time dedicated to the design of their graphical interface is reduced, allowing developers to focus on the programming of the dynamics and strategies of their applications. Thus, this framework is helpful for reducing considerably the development time of OGRE applications.

2   OGRE Markup Language (OGREML)

Based on the literature review, it is possible to verify that there are a few authoring tools for OGRE 3D environments. Nevertheless, these tools have a proprietary code and do not provide the automatic generation of OGRE code (C++ or C#). For this reason, it was important to provide a solution for generating OGRE code automatically, thus optimizing the prototyping of OGRE applications. Indeed, with a graphical interface for creating OGRE 3D scenarios, the OGRE developer is able to build his virtual scenario quickly and, after that, to generate all the respective OGRE code to the scenarios automatically.

In this context the XML-based language *OGREML (OGRE Markup Language)* was proposed as a common representation for OGRE 3D scenarios. Therefore, the main goals for the proposal of OGREML were: (1) to describe completely all the components of an OGRE scenario and; (2) to provide the interoperability among different OGRE tools.

OGREML is an XML-based language to describe OGRE virtual scenarios and it is composed of four major components: *Ambient*, *Objects*, *Multimedia* and *NetworkConfiguration*. All these components are gathered in the main block *SceneConfiguration*, as depicted in Figure 1.
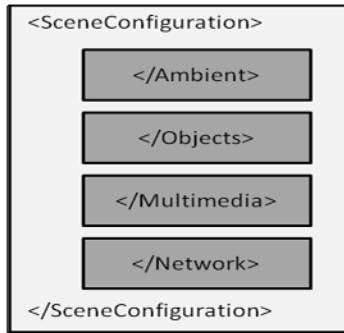
Fig. 1. Main components of OgreML

For instance, the component *Ambient* contains all information about the environment of the scenario, such as the color of ambient light (position, direction, values of the diffuse color and specular color, and its type), fog, sky, plane and camera. This component is illustrated in Figure 2.
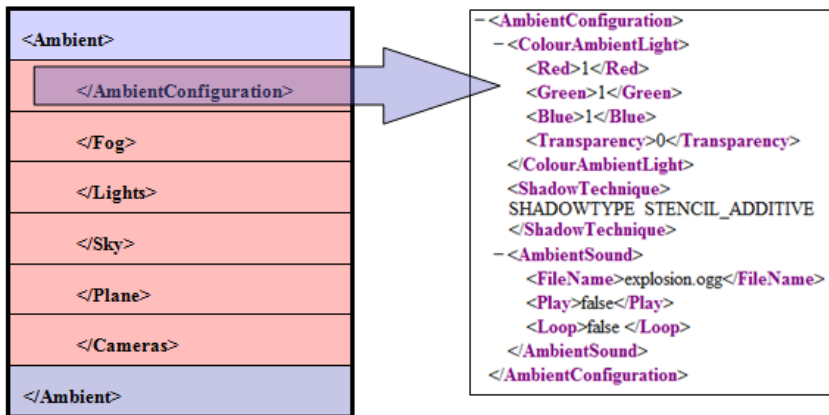
Fig. 2. Illustration of the component *Ambient*

The *Objects* component contains a description of all objects in the scene and their attributes such as: id, position in the scenario, scale and rotation, mesh file and if it can cast shadows.

The *Multimedia* component defines synchronized multimedia presentations that may consist of videos, images and sounds, which will be displayed in the virtual world after the activation of a trigger [6].

The *NetworkConfiguration* represents the configuration of different network services among distributed OGRE applications, such as chat, VoIP and distributed navigation [7].

OGREML uses a simple description to represent all the components of an OGRE 3D scenario, being at the same time expressive supporting most of the OGRE development components. In the next section we present further details on the design and implementation of OGRE-Coder.

3   OGRE-Coder: Design and Implementation Issues

In this section we present the main issues related to the implementation of OGRE-Coder such as requirements, use cases, architecture, and implementation aspects.

3.1   Requirements and Use Cases

The **requirements** reflect the needs that a client wishes to be fulfilled by the application being developed [8]. These requirements serve as guidelines, and most of the times, they may represent the objectives to be achieved in the project development. The requirements can be defined as functional and non-functional.

The functional requirements describe the functionalities or services provided by the system (main functionalities). The functional requirements for the developed application were:

- Provide the user with a free navigation inside the virtual environment;
- Allow the user to open and visualize any virtual scenario described through the OgreML;
- Allow the automatic generation of C# code based on all the scenarios described using OgreML;
- Allow the user to visualize a scenario based on the generated C# code, and;
- Allow the visualization of multimedia content inside the virtual environment.

The non-functional requirements are those which are not related directly to the main functionalities of the system such as, for instance, reliability, robustness, efficacy and security. The main non-functional requirements for the OGRE-Coder are:

- Provide a good performance for the automatic generation of C# code;
- Provide continuous and steady presentation of multimedia content inside the virtual environment, and;
- Allow future enhancements and expansion in the application, if needed.

A **use case diagram** allows the understanding of all the possible interactions with the system, under the standpoint of the user (actor) [9]. After the definition of the requirements, only one actor was identified for the system, which is one responsible for:

- Choosing the scenario to visualize and map – The user can choose the respective OgreML file related to the scenario he wants to visualize and generate the respective C# code;
- Visualizing a scenario – The user can visualize a scenario from an already generated C# code, or from an OgreML file;
- Navigating in the scenario – The user can use the keyboard and/or the mouse to navigate and interact within the virtual environment, and;

- Visualizing multimedia content – The user can navigate within a virtual scenario and interact with a trigger (virtual object) in order to start the presentation of multimedia content (audio, video and image) inside this environment.

3.2 Architecture

The OGRE-Coder architecture is divided into five sub-modules: *User Interface*, *Main Module*, *OGRE*, *XML-Manager* and *Multimedia Manager*. Figure 3 illustrates the architecture of OGRE-Coder.
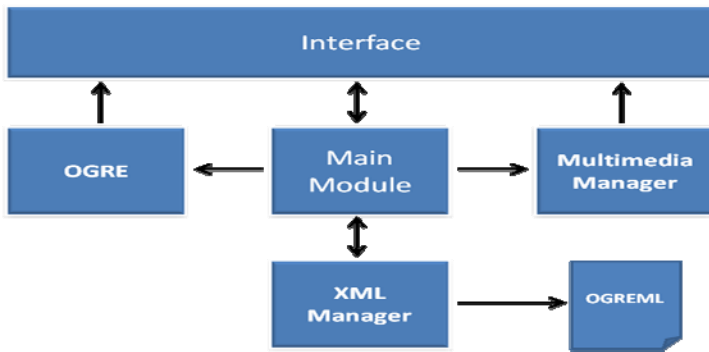
Fig. 3. Main architecture of OGRE-Coder

As depicted in Figure 3, *Main Module* and *Interface* have a major role in the application. *Main Module* is responsible for the integration and coordination of all the other modules. *Interface* is responsible for receiving and handling all the user interactions. These two modules work in a coordinated way in order to respond to users interactions and to provide the presentation of the virtual environment.

In particular, the **Main Module** can still be refined into 5 sub-modules, as illustrated in Figure 4(a):

- *OGRE-Manager* – module responsible for managing the user interactions, and initializing different components such as multimedia, scenario, data input and XML;
- *Input Manager* – module responsible for initializing the libraries related to the data input devices and events management (keyboard and mouse inputs);
- *XML Manager* – module responsible for managing the XML parsing requirements and the respective C# generation;
- *Multimedia Initialization* – responsible for initializing the multimedia libraries (audio and video), the definition of triggers and the execution of user interactions handlers, and;
- *Scenario Initialization* – module responsible for the initialization of all the objects which will be presented in the virtual scenario and for the definition of multimedia triggers.

(a) **Main Module**   (b) **XML Manager**   (c) **Multimedia Manager**
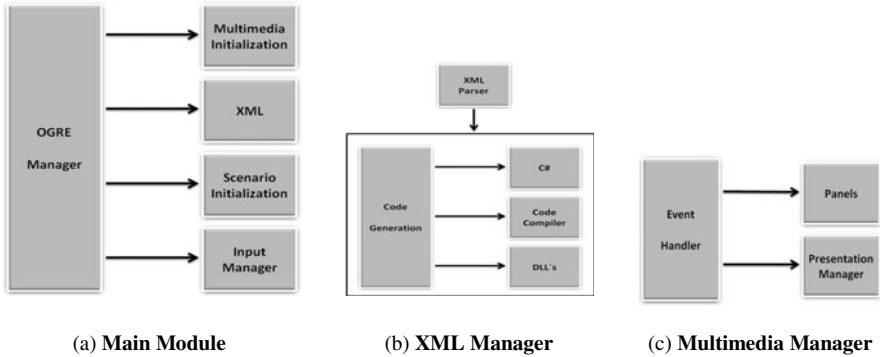
Fig. 4. Main Module Architecture

The *XML Manager* is responsible for reading the XML file indicated by the user, and for the automatic generation of its respective C# code. This module is depicted in Figure 4(b). This module is refined in the following sub-modules:

- *XML Parser* – component that carries out the parsing of the XML file, following the description of the virtual scenario, and storing its objects in typed lists. These objects can be of the type *ambient*, *sky*, *object*, *camera*, *fog*, etc.;
- *Code Generation* – component that provides different functionalities related to the automatic generation of C# code;
- *C#* - component that reads the lists provided by the XML parsing, and generates one object-oriented class for each type of object in the respective list;
- *Code Compiler* – component that compiles the description of the objects present in each class in order to create dll´s (dynamic-link libraries), and;
- *DLL´s* – component that applies the dll´s generated by the *Code Compiler* module in order to instantiate all the objects described by the C# classes for the visualization of the OGRE virtual scenario.

The *OGRE Module* (depicted in Figure 3) represents the graphic engine OGRE which is the main module of the system being responsible for creating, managing and updating the tri-dimensional model. Briefly, this module is responsible for rendering and managing the presentation of the virtual environment. Further details on the OGRE´s architecture and hierarchy of classes can be found in [1].

The *Multimedia Manager Module* is responsible for handling the multimedia presentation requests. These requests take place when the user interacts with a trigger object, which can be associated with any 3D object in the virtual scenario. As depicted in Figure 4(c), this module is composed of the following sub-modules:

- *Event handler* – component responsible for managing the events produced by the user interaction in order to trigger a multimedia presentation;
- *Panels* – component responsible for creating the visible panels where multimedia content (image and video) will be presented inside the virtual environment, and;
- *Presentation Manager* – component responsible for managing the multimedia presentation, begin and end of a media object presentation, presentation duration of each media object, panel management, etc.

3.3   Implementation Tools

OGRE-Coder has been implemented using Visual Studio C# 2008 [10]. Visual Studio is a software development environment created by Microsoft. This software is very flexible and allows creating from simple to more complex applications.

Visual Studio proved to be an asset for the development of this application since some of its features enable a faster coding and debugging.

Visual Studio 2008 C# allows the utilization of Visual .NET, C#, J# and ASP.NET, while providing the possibility to develop applications from scratch for different types of devices such as PCs, pocket PC and smart phones.

4   OGRE-Coder Functionalities

In this section we present some of the main functionalities of OGRE-Coder and we illustrate these functionalities with some snapshots of the application.

The development of OGRE applications is in general a programming activity since all the code must be developed in C++ or C#. For this reason, the development process of these applications is rather complex and time consuming to the OGRE developer since he has to code all the virtual scenarios to be applied in his application.

In order to facilitate the development of OGRE virtual scenarios a generic framework was proposed for promoting the interoperability among different OGRE authoring tools, and further automatic generation of the respective code with the
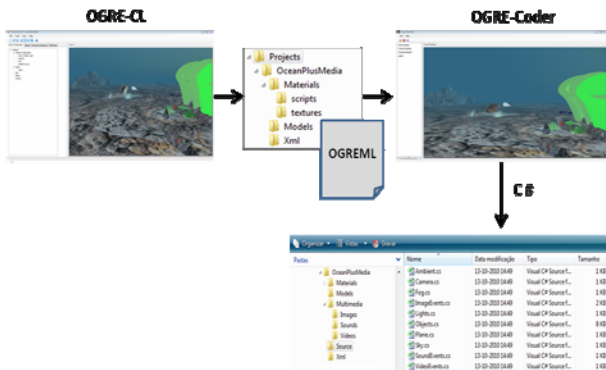
Fig. 5. Generating C# Automatically with OGRE-Coder

description of the virtual scenarios. This framework relies on the utilization of OGREML for the complete description of OGRE virtual scenarios.

For instance, virtual scenarios can be created easily using OGRE-Creative Laboratory (OGRE-CL) [11], and by means of OGREML, be exported to Ogre Coder for the automatic generation of C# code, as illustrated in Figure 5. The integration of an authoring tool for the creation of virtual scenarios (OGRE-CL) with OGRE-Coder for the automatic generation of C# code will allow a more rapid prototyping of OGRE applications.

Next section presents some characteristics of the authoring tool for OGRE virtual scenarios, OGRE-CL.

4.1   Authoring OGRE Virtual Environments

OGRE Creativity Laboratory (OGRE-CL) is an easy to use graphical tool for optimizing the creation of OGRE applications [11]. With this tool OGRE developers are able to quickly design the main virtual scenarios of his applications, being able to further generate the respective C# code related to the virtual scenarios conceived. Among its different functionalities, OGRE-CL allows the user to: create and manage a virtual scenario from scratch; configure parameters for the virtual environment (*Ambient*); add/remove mesh files; add and manage 3D objects; add multimedia content; allow conversation between different participants of a session, either by text message or by Voice over Internet Protocol (VoIP), and; allow the distributed navigation between remote OGRE applications.

OGRE-CL provides a simple interface for the composition of virtual scenarios based on OGRE. An important characteristic of OGRE-CL is that it relies on an extensible library of objects (meshes) that can be used to create the virtual scenarios. Some snapshots of OGRE-CL are depicted in Figure 6 (a) and (b).

(a)   (b)
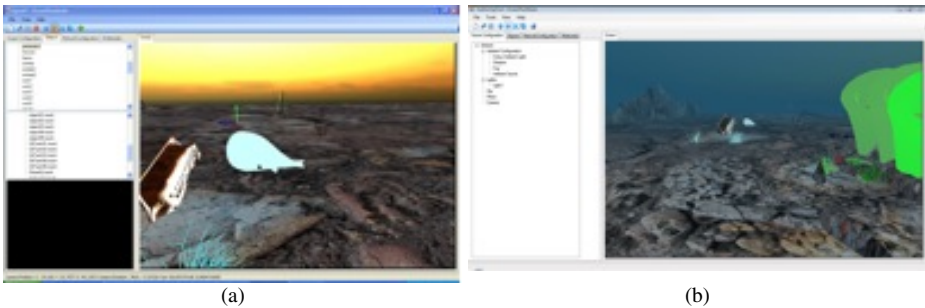
Fig. 6. Ogre Creative Laboratory (OGRE-CL)

OGRE-CL is an optimal solution for OGRE developers since they can minimize the time for designing the graphical interface of their applications. With the utilization of OGREML, they can export the description of their virtual scenarios to OGRE-Coder to further generate automatically the respective OGRE code.

As a case study, we applied OGRE-CL to create a thematic virtual scenario related to the undersea life. This scenario presents different 3D objects such as sunken boats, whales, fishes, sea stars, rocks, anchors, etc., as illustrated in Figure 6 (a) and (b).

When this scenario is saved, it is stored as an OGREML file. This OGREML file and all of its respective files (meshes, materials, etc.) are stored in a directory structure which is used to provide interoperability between OGRE-CL and OGRE-ML, as illustrated in Figure 5.

4.2   Generating OGRE Code with OGRE-Coder

When the user executes OGRE-Coder and opens the OGREML file related to the virtual scenario he wishes to visualize and convert to OGRE code, the application will automatically generate all the C# code associated with this scenario, and also it will render this scenario in an OGRE window so that the user can verify the final graphical quality of his scenario, as depicted in Figure 7.
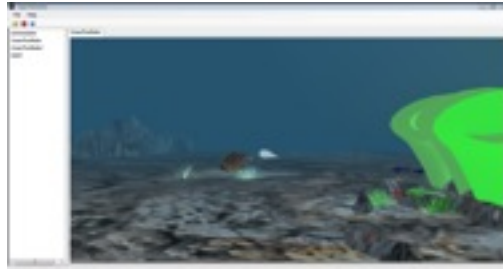
Fig. 7. Visualization of a virtual scenario with OGRE-Coder

OGRE-Coder also allows users to navigate within the virtual environment and to interact with all the objects during the visualization of it. Furthermore, the user is also able to interact (click) with some objects considered as triggers for multimedia presentation. If he clicks on these objects, the multimedia presentation is automatically started, as illustrated in Figure 8 (a) and (b).
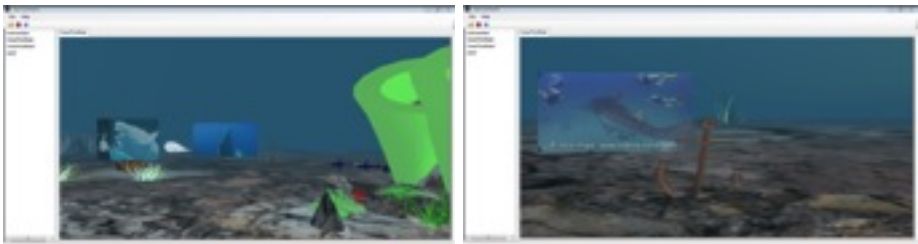
(a)   (b)

Fig. 8. Presentation of multimedia content with OGRE-Coder

In general, OGRE-Coder provides a good performance for the visualization of virtual scenarios and automatic generation of their respective C# code. When applied jointly with OGRE-CL, these tools represent an expressive environment for the intuitive authoring of high-quality virtual environments supporting a variety of 3D models and textures.

5   Conclusions

In this paper we presented the main aspects of the development of a useful tool for the rapid development of 3D scenarios, OGRE-Coder. An important contribution presented in this paper is the proposal of an XML-based language, called OGREML, for the complete description of OGRE virtual scenarios, and for promoting the interoperability among OGRE authoring tools.

With the utilization of OGREML, OGRECL becomes an optimal solution for OGRE developers since they can minimize the time for designing the graphical interface of their applications, and after that they can apply OGRE-Coder to generate automatically the OGRE code related to his scenarios. This approach allows developers to quickly prototype the virtual scenarios of their OGRE Applications, concentrating their efforts on the development of the dynamic aspects of the application such as artificial intelligence, strategies, events handling, etc.

References

1. OGRE – Open Source 3D Graphics Engine (2001), `http://www.ogre3d.org/`
2. Ogitor SceneBuilder, `http://www.ogitor.org/HomePage`
   (last visited in November 2010)
3. OGRE Editor Multi Scene Manager Project Environment,
   `http://www.youtube.com/watch?v=TU_Tc4EBLHQ`
   (last visited in November 2010)
4. OGRE – MOGRE Editor,
   `http://www.youtube.com/watch?v=xXJDvKzY1Us&feature=related`
   (last visited in November 2010)
5. Fernandes, M.D.: Basis for the Automatic Generation of OGRE Virtual Scenarios. M.Sc. Dissertation in Informatics Engineering – University of Madeira, Madeira, Portugal (2010) (in Portuguese)
6. Freitas, R.: Multimedia Presentation in Virtual Environments. B.Sc. Dissertation in Informatics Engineering – University of Madeira, Madeira, Portugal (2007) (in Portuguese)
7. Cardoso, G.: Uma Virtual – Basis for the development of complex virtual environments. B.Sc. Dissertation in Informatics Engineering – University of Madeira, Madeira, Portugal (2007) (in Portuguese)
8. Oberg, Roger. Probasco, Leslee. & Ericsson, Maria Applying Requirements Management with Use Cases (2010),
   `http://www.wthreex.com/rup/papers/pdf/apprmuc.pdf`
   (last visited on November 2010)
9. Bredemeyer Consulting Functional Requirements and Use Cases (2010),
   `http://www.bredemeyer.com/use_cases.htm` (last Visited on November 2010)
10. Visual Studios 2010 Editions – Microsoft Visual Studio,
    `http://www.microsoft.com/visualstudio/en-us/products/`
    `2010-editions` (last visited on November 2010)
11. Teixeira, D.J.O.: OGRE Creativity Labs (OGRE-CL) – Basis for the development of an authoring tool for virtual scenarios (In Portuguese). M.Sc. Dissertation in Informatics Engineering – University of Madeira, Madeira, Portugal (2010) (in Portuguese)

Natural and Intuitive Video Mediated Collaboration

Joona Manner and Juho-Pekka Virtanen

Aalto University School of Art and Design, Industrial and Strategic Design
Hämeentie 135 C 00560 Helsinki Finland. 2011
`joona.manner@aalto.fi, juho-pekka.virtanen@aalto.fi`

Abstract. This paper presents a project which introduces a novel concept and technical implementation of a video mediated collaboration system. The starting point was to provide tools suitable for professionals particularly in the field of design and engineering. The core of the concept is to create a permanent connection between two remote locations and to combine live video screen and interactive table surface. The aim is to use the best features of existing videoconferencing solutions, smart whiteboards and collaborative file sharing systems. The technical solution is based on utilisation of existing components and low cost software. The first prototypes suggest that this type of system has more intuitive and natural user interface compared to current solutions. With future field testing the challenges concerning the application environment can be better met.

Keywords: Video Mediated Collaboration, VMC, Natural User Interface NUI, intuitive, video mediated communication, computer mediated communication, multimodal systems, remote collaboration, computer supported cooperative work, user interfaces, videoconference, human centered systems, whiteboard, smartboard, disappearing computer, telepresence, face-to-face, interactive table

1 Current Systems for Enabling Remote Collaboration

There are several tools facilitating video mediated communication over distance, ranging from carefully designed telepresence environments to desktop videoconferencing applications. Smart whiteboards and interactive table surfaces have also been introduced into the office environment. In the following, we briefly review these systems, and point out some of the problems and limitations found in them.

1.1 Videoconferencing and Telepresence Systems

Videoconferencing systems have their historical background in videophone solutions. They have been developed as a visual version of a conventional telephone. Modern videoconference solutions thus share the basic operational logic of a telephone, where a connection is established on user's request and terminated at the end of the discussion.

Current systems have not been able to solve all challenges of having a viable video mediated collaboration environment. Full scale telepresence systems (e.g. Cisco, Polycom) feature carefully designed room environments with special lighting,

cameras, microphones and large displays [5, 15]. These systems usually have their own reserved internet connections and specific expensive hardware solutions to ensure the stability and quality of the connection.

Telepresence solutions may have accessibility problems [14]. Usually an organisation has a limited amount of telepresence capable rooms, this may hinder collaborative work in general. Arranging meetings to telepresence room and establishing a video conference there also creates an extra step to the creative design process. Furthermore telepresence rooms are usually not suitable to any other kind of work than meetings. An environment created for executive level meetings does not encourage casual interaction. In design process it would be useful to be able to review a physical prototype and sketch new solutions together [4, 8].

1.2 Desktop Video Conferencing

Laptops and modern smart phones with integrated web-cameras enable video calls and limited videoconferencing at nearly any location. One of the disadvantages is that current desktop videoconferencing setups are designed for individual use, and thus work poorly when there are several people by the same computer device. They also restrict the work to be conducted with the computer and do not encourage the use of conventional, not computer related working methods. Video, which represents the face of the remote user in a small display screen with low resolution and poor image quality restricts the modes and aspects of communication [1, 17].

Document sharing features of current desktop videoconferencing solutions and collaborative file sharing systems enable distributed work on the same document. However, they focus solely on digital documents and tools. This is considered as a limitation to work and forces users to use tools not entirely natural to them [8].

1.3 Interactive Tables and Smart Whiteboards

Once the large LCD and Plasma computer displays became available in early 2000's, there have been several interactive tables developed (Microsoft Surface, EyeBoard, Aurora, atracTable, Illuminate). Their most common use is viewing and using interactive multimedia for various purposes. The interactive surface of these devices is usually horizontally positioned [2,3,9,10,13].

Smart whiteboards that enable the use of interactive multimedia are provided by several different companies, most commonly for education and business environments [16,19,21].

There is also research and development of a smart whiteboard that enables collaborative work over distance, sharing tangible post-it notes and drawings [8].

All smart whiteboard solutions have the same limitations as ordinary whiteboards: they do not allow attaching objects on the vertical surface. In the worst case they can be used with an electronic pen only, which makes the drawing experience less natural. These boards are usually vertically positioned and allow digital content to be added via projector and the outcome to be captured as a digital image and in some systems shared with another similar device.

2   The Importance of Usability

Usability is one of the main challenges of all existing videoconference systems. These systems are often found too complex and thus require training [1,6,14]. Even making a simple conference audio call requires some knowledge about telephone operator system.

Videoconference systems vary and have dissimilar user interfaces [5, 15] which makes using them even more difficult. Establishing connections between different systems is in the worst case, impossible. All interruptions that are caused by the technology are found to prevent the actual communication [1]. These problems may limit the overall use of videoconferencing.

Videoconference systems should be designed to be intuitive and easy to use, so that any person without any specific training could operate them. Creating a connection with a videoconference system usually requires previous experience. Establishing a connection between two places should be as easy as making a conventional telephone call. Unfortunately, in many cases starting a videoconference can fail already at this stage [14].

More usability and technical challenges arise once the connection has been established. Poor quality nonsynchronous video and audio can cause problems in understanding, turn taking etc. [11]. Sharing documents might require the use of some other system, which may have a completely different user interface [1].

We find that making the technology unnoticeable would be the easiest way to tackle these problems. Instead of concentrating on technical issues and problems users could focus on the actual work with the tools they prefer and are familiar with.

3   Building Prototypes

3.1   First Prototype

This project started as an experiment to connect two Aalto University buildings located in remote campuses with a good quality live video link (Design Factory, Espoo and Industrial design programme, Helsinki).

Our first prototype (see figure 1) was a high quality (FullHD) permanent video connection between these two locations. Both ends were connected to FUNET (Finnish University and Research Network) which allowed us to use 50 Mbit (25 Mbit in / 25 Mbit out) bandwidth. During this experiment we found out that our system had used as much data bandwidth as the rest of our campus. This connection was round the clock sending unpacked raw camera data directly through the network using DVTS [7] for WindowsXP and receiving it in the other end by using the same software. The first prototype was used for a period of six months.

The purpose of this system was to connect design professionals and trainees, and therefore system providing only live video connection was considered to be too limited. Collaboration between these two locations required more than a permanent video connection. This gave us an idea of a simple interactive table that could serve as a shared drawing board.
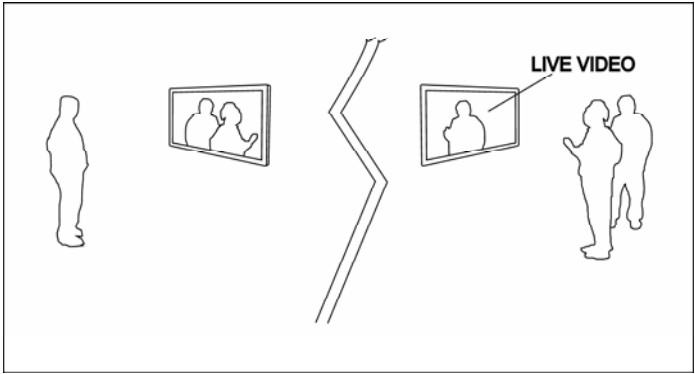
Fig. 1. First prototype

3.2   Second Prototype

In the second prototype we added another video connection for the table, still maintaining the first video connection (see figure 2). Since the first prototype required too much bandwidth to be feasible, we started experimenting with Skype [18]. The use of Linux operating system enabled us to control Skype using Python programming language. We were able to create an automatically re-connecting system, eliminating the need of all user intervention. In addition, Linux was considered more stable operating system than WindowsXP.

One prototype unit consisted of two desktop PCs and two displays installed to a single casing that had wheels. The casing was required to hide the computers and their controls completely. Furthermore this casing allowed embedding the monitor to the table surface, mounting the screen for video connection and installing all necessary cameras.
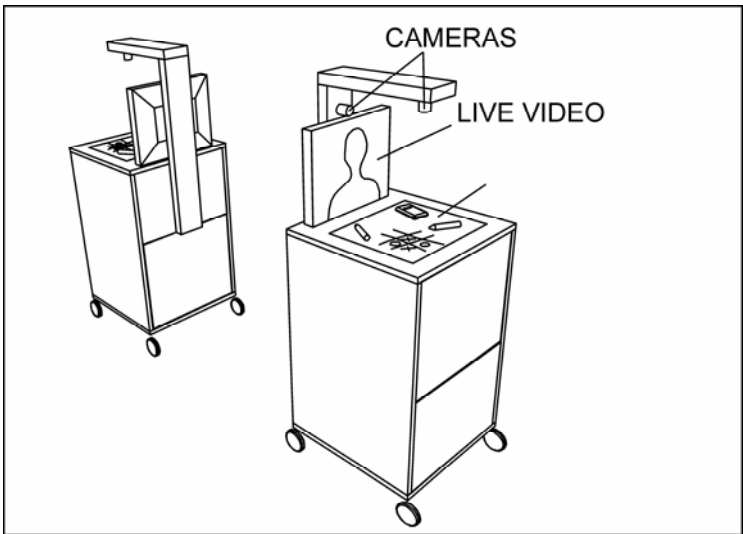
Fig. 2. Second prototype

Table surface consisting a monitor was plated with glass that enabled drawing on it with whiteboard markers and supporting objects placed on top of it. Camera was installed above the table surface capturing the table top. Two identical copies of this setup were constructed to be used in field studies later.

As we found that our interactive table surface requires a higher resolution than Skype can provide, we plan to replace Skype with a set of Linux programs better suited for this kind of application. We also plan using video cameras equipped with better optics, offering higher image quality compared to current webcams.

4 Our Design Concept

Our concept combines a high quality video and audio connection on large vertical screen and horizontal high resolution large video table. see figure 3. This table can be used as a shared whiteboard and surface that can transfer image of any object that is placed on it to remote location and vice versa. Information can be shared just by drawing directly on the table surface or placing an object on to the table (i.e. IPad, smartphone, laptop, paper, coffee cup, scale model).
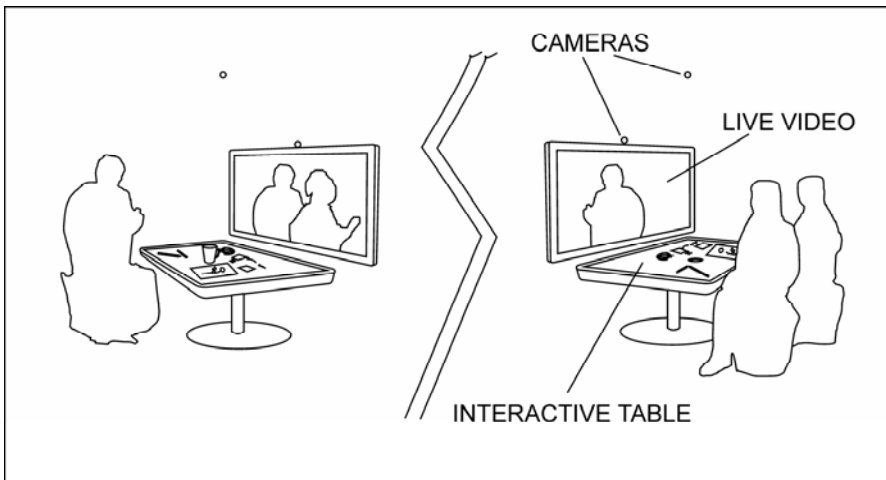
Fig. 3. Final concept

4.1 Permanent Connection, Invisible User Interface

In the era where Internet connections allow users to access digital media around the clock, establishing video calls separately for each discussion can be considered old fashioned. Creating a permanent connection removes this user task.

We propose that the video and audio connection is active all the time which eliminates the problem of establishing a connection. Consequently, the conventional user interface becomes unnecessary and can be hidden from the user. Users are able to focus on their collaborative work intuitively, naturally and without struggling with technology. The solution resembles the philosophy of ubiquitous computing where

the user needs no longer focus on using a computer system, but simply performs the normal tasks, while the computer facilitates them on the background [6,20].

Having non-stop video connection in office environment changes the whole culture of collaborative video mediated work from formal meeting to an informal gathering around a coffee table.

4.2 Natural Collaborative Tools

In creative collaboration, drawing and writing to a shared table with regular whiteboard pens is more intuitive than using mice, cursors or computer drawing tablets. Our interactive table allows the use of these tools in video mediated collaboration.

Sharing information by placing it on a table is less complicated than scanning in documents, sharing them via file sharing systems and struggling with different software incompatibilities. Intuitive tools and working methods also remove the need of any system training.

5 Discussion

5.1 Is Permanent Connection Too Limited?

Our concept is useful if there is a constant need of communication between two locations. By eliminating the "call function" we remove the possibility of making individual video calls to more rarely contacted recipients. The proposed system cannot completely replace current videoconferencing systems. Are the usability requirements met closely enough to justify the existence of partially overlapping systems? Could the development of an intuitive connecting interface solve this problem?

5.2 Where Could This Concept Be Applied?

The potential use-cases of the proposed system include remote offices, established subsidiaries and other long-time partners of organisations. There are also other situations where easy video communication could be applied, that haven't been at the scope of development of current video conferencing systems i.e. a live connection from your local restaurant to a restaurant in Athens as new interior decoration element.

5.3 Challenges of Permanent Video Connections

One of the potential problems in permanent video connections is privacy. Research has been conducted on video communication systems in home environments. One of the challenges found has been defining which areas are shared by the video communication system, which are private, and how this is indicated [12].

Closed videoconferencing environments ensure privacy, but also create accessibility problems. If open videoconferencing is introduced to the office environment, it may cause privacy issues. Therefore the video communication areas should be clearly indicated, so that they can be avoided if necessary.

Permanent live video connection wastes bandwidth when nothing is happening in the viewed area. Implementing software based motion detection to activate data transfer when persons appear into field of camera would prevent sending data in vain.

5.4 Future Development Challenges

The overall concept and the implementation should be tested with the users in the field. This would enable further study of potential problems regarding usability, privacy and technical aspects. Through extensive field testing it can be verified whether the proposed system genuinely supports video mediated collaboration and social interaction when compared to existing solutions.

References

[1] Agius, H., Angelides, M.: Desktop video conferencing in the organisation. Inf. Manage 31(6), 291–302 (1997), doi:10.1016/S0378-7206(97)00009-8

[2] atracTable, http://www.atracsys.com (accessed January 12, 2011)

[3] Aurora, http://mindstorm.com (accessed January 12, 2011)

[4] Barthelmess, P., Kaiser, E., Lunsford, R., McGee, D., Cohen, P., Oviatt, S.: Human-centered collaborative interaction. In: HCM 2006: Proceedings of the 1st ACM International Workshop on Human-Centered Multimedia. ACM, New York (2006)

[5] Cisco, http://www.cisco.com (accessed January 12, 2011)

[6] Cooperstock, J.: Making the user interface disappear: the reactive room. In: Bennet, K., Gentleman, M., Johnson, H., Kidd, E. (eds.) CASCON 1995: Proceedings of the 1995 conference of the Centre for Advanced Studies on Collaborative Research. IBM Press (1995)

[7] DVTS, http://www.sfc.wide.ad.jp/DVTS/software/win2000/ (accessed January 12, 2011)

[8] Everitt, K., Klemmer, S., Lee, R., Landay, J.: Two worlds apart: bridging the gap between physical and virtual media for distributed design collaboration. In: CHI 2003: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, New York (2003)

[9] EyeBoard, http://www.eyeclick.com (accessed January 12, 2011)

[10] Illuminate, http://www.gesturetek.com (accessed January 12, 2011)

[11] Isaacs, E., Tang, J.: What video can and can't do for collaboration: a case study. In: MULTIMEDIA 1993: Proceedings of the First ACM International Conference on Multimedia. ACM, New York (1993)

[12] Junestrand, S., Tollmar, K.: Video Mediated Communication for Domestic Environments - Architectural and Technological Design. In: Streitz, N., Siegel, J., Hartkopf, V., Konomi, S. (eds.) CoBuild 1999: Proceedings of the Second International Workshop on Cooperative Buildings, Integrating Information, Organization, and Architecture. Springer, London (1999)

[13] Microsoft Surface,
    `http://www.microsoft.com/surface/en/us/Pages/Product/`
    `WhatIs.aspx` (accessed January 12, 2011)
[14] Olson, G., Olson, J.: Distance matters. Hum.-Comput. Interact. 15(2), 139–178 (2000),
    doi:10.1207/S15327051HCI1523_4
[15] Polycom, `http://www.polycom.co.uk` (accessed January 12, 2011)
[16] Polyvision, `http://www.polyvision.com` (accessed January 12, 2011)
[17] Sebe, N., Lew, M., Huang, T.: The State-of-the-Art in Human-Computer Interaction. In:
    Sebe, N., Lew, M., Huang, T. (eds.) ECCV/HCI 2004. LNCS, vol. 3058, Springer, Hei-
    delberg (2004)
[18] Skype, `http://www.skype.com` (accessed January 12, 2011)
[19] Smartboard, `http://smarttech.com` (accessed January 12, 2011)
[20] Streitz, N.: From human–computer interaction to human–environment interaction: Ambi-
    ent intelligence and the disappearing computer. In: Stephanidis, C., Pieper, M. (eds.)
    ERCIM Ws UI4ALL 2006. LNCS, vol. 4397, pp. 3–13. Springer, Heidelberg (2007)
[21] Teamboard, `http://teamboard.info` (accessed January 12, 2011)

Evolutionary System Supporting Music Composition

Rafal Drezewski and Przemyslaw Tomecki

Department of Computer Science
AGH University of Science and Technology, Krakow, Poland
`drezew@agh.edu.pl`

Abstract. Evolutionary algorithms are heuristic techniques for finding approximate solutions of hard optimization and adaptation problems. Due to their ability to create innovative solutions, evolutionary algorithms are very well suited for applications in the domain of art. In this paper the system supporting human composer in the process of music composition with the use of evolutionary algorithm is presented. The architecture of the system as well as implemented algorithms and results of selected experiments will be presented.

1 Introduction

Evolutionary algorithms (EAs) are heuristic techniques for finding approximate solutions of hard optimization and adaptation problems [2]. The evolutionary algorithms can be applied to many different problems, like multi-modal optimization, multi-objective optimization, combinatorial optimization, etc.

Because of their ability to explore huge solution spaces and to generate/ propose new and original (previously not even known to experts in the field) solutions to the given problems they can be applied in domains which require innovative approaches and creativeness, for example designing of engineering components, architecture, etc. Such capabilities were supposed to belong only to the human beings but the evolutionary algorithms demonstrate that computer program can also be creative and propose new, so far unexplored or unknown, solutions and even create art. One of the domains of art in which the evolutionary algorithms can support (or compete with) humans is the process of music composition. This paper presents one of the possible approaches to support the human composer.

Almost every musician is to some extent skeptical about the concept of using machines (computers) in the process of music composition—of course what we mean here is not using electronic musical instruments or some software supporting the process of music composition by the human composer, but composing musical pieces by computer itself. Probably something like complete replacement of the human composer with some music software or hardware in the process of music composition is impossible. However, humans can easily interact with the system (especially when we use evolutionary algorithms) and guide the process of music composition.

There have already appeared several approaches to supporting music composition with the use of evolutionary algorithms. J. A. Biles is the author of GenJam—system for generating jazz solos on the basis of input data from MIDI interface [3, 4]. GenJam was generating sample bars or phrases, which then were presented to the so called "mentor" who was assigning them fitness values.

B. Budzyski in his work [5] presented different approach to music composition with the use of evolutionary algorithms. There was no input data—initial population was generated at random. There was also human factor present during evolution, but composer decided only whether to continue or to stop the evolution process—music pieces were evaluated automatically, since there existed the fitness function.

Tree-based representation for musical pieces and the appropriate mutation operator was proposed in [1].

Authors of [6] proposed Harmony Search Algorithm. Musical composition was generated on the basis of vectors of notes improvised by n musical instruments. These vectors were classified on the basis of intervals between theme and variations.

R. De Prisco and R. Zaccagnino presented the idea of bass harmonization with the use of genetic algorithm [9]. Their system took bass line as the input and on its basis generated three additional voices.

In this paper we will present evolutionary system supporting music composition—*EvoMusComp*. We will use different approach than those presented above. Firstly, in the presented system user can generate the whole new piece of music, or variation on chosen theme, with the use of evolutionary algorithm. User can interact with the system by setting values of parameters and stopping the process of evolution. The music pieces are evaluated with the use of fitness function, which is constructed from several components by the user at the beginning of evolution. In the following sections we will describe the architecture of the system, evolutionary algorithm used to generate the music and selected results of preliminary experiments.

2   Evolutionary System Supporting Music Composition

In this section we will describe the architecture of the evolutionary system supporting music composition (EvoMusComp) as well as the genetic algorithm used for music generation.

2.1   Architecture of the System

The implemented evolutionary system supporting music composition is based on the Java technology. The eclipse RCP platform was used to create user interface (see Fig. 1a).

Basic components of the system are presented in the Fig. 1b. *System* component is responsible for merging information and controlling the information flow
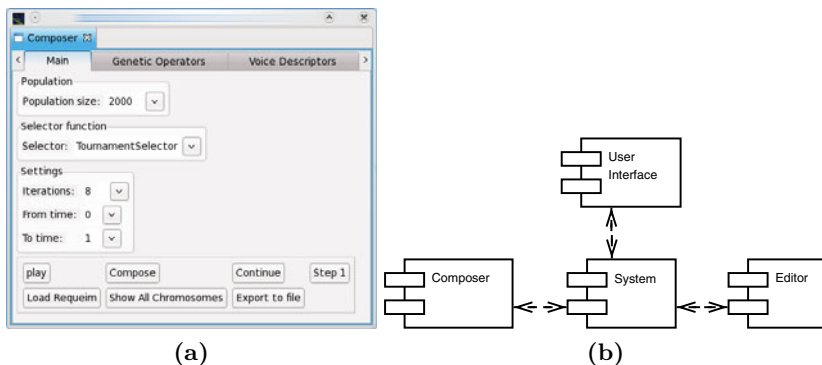
Fig. 1. GUI (a) and basic components (b) of the EvoMusComp

between other components. *User Interface* component is responsible for interaction with the user. *Composer* component is responsible for composition process and the evolutionary algorithm is a part of this component.

2.2   Genetic Algorithm

The general structure of the evolutionary algorithm used in the system is based on the well known J. Holland's and D. Goldberg's genetic algorithm [7].

In the classical genetic algorithm the binary representation is used in order to represent the encoded solution. In the genetic algorithm used in the presented system such representation is adapted to the music notes description.

The attributes of the *Genotype* class, that represents the genotype of individuals include:

- *name*—*String* type—represents the name of the note in the specific octave;
- *octave*—*integer* type—represents the octave that the note is in. In the algorithm it takes values from the range 3 to 7, depending on the key signature;
- *noteKind*—*double* type—represents the duration of the note. Acceptable values are: 1.0 (whole note), 0.5 (half note), etc.

In many research papers in the area of music composition, the structure of the chromosome is based on the single voice line. In the implemented system, the chromosome is built from *n*-elements voice array. This construction allows to compose not only one voice melodies, but also the whole phrase, or even whole piece of polyphonic music.

There are two implemented genetic operators: crossover and mutation. Below we will describe them in more details because they are quite different from the standard operators used in genetic algorithms.

The recombination operator was implemented as one-point crossover. When we try to exchange genes from the one chromosome with the genes from the other chromosome there may appear some problems. We can face the situation when the new chromosomes, created after the crossover, can be invalid.
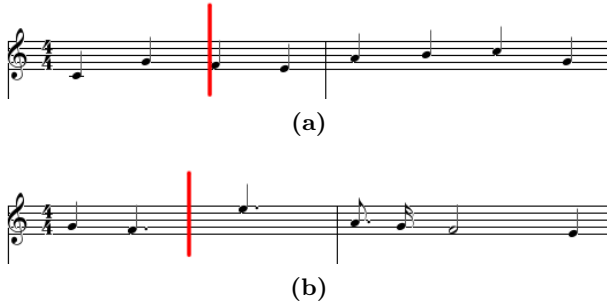
Fig. 2. First (a) and second (b) chromosome before crossover

Such situation is illustrated in the Fig. 2. The example shows that there is no note in the second chromosome in the place where the random point is placed before performing crossover. The sum of the notes' duration in the new chromosome would be longer or shorter than allowed.
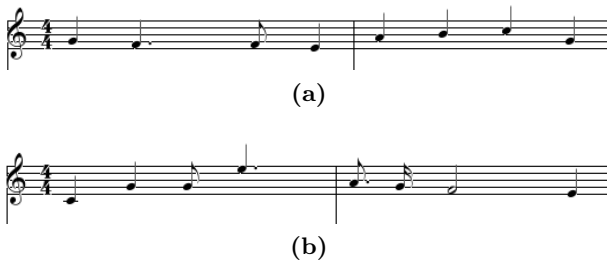
Fig. 3. First (a) and second (b) chromosome after crossover

Such error has to be fixed by changing the border note's values—it is shown in the Fig. 3.

There are three variants of mutation, which can be chosen by the user:

- *basic*—value of a note is changed half tone up or down;
- *extend*—two neighboring notes are joined;
- *split*—one note is split into two notes.

The probability of applying the mutation operator is also set by the user. The *basic* mutation of sample chromosome is shown in the Fig. 4.

There are three selection methods implemented in the system:

- roulette selection;
- tournament selection;
- ranking selection.

(a)

(b)
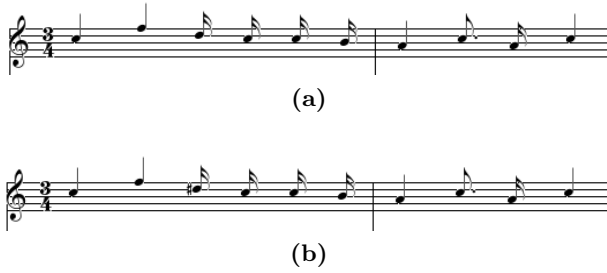
Fig. 4. Chromosome before (a) and after (b) mutation

As the result of applying of each of the above mechanisms the next generation consists of the same number of chromosomes (the number of the individuals in the population is constant).

There are three basic components of the fitness function, in which user can:

- define preferred kind of note and distance between neighboring notes;
- define harmony functions;
- define the chromosome that will serve as a reference point for the adaptation function.

With the use of the first and the second component we can compose a completely new piece of music, but with the use of the third one, the user is able to compose a music variation on a given theme.

2.3   The Process of Music Composition

The process of music composing with the use of implemented system is illustrated in the Fig. 5. First, we define parameters connected with the music notation:
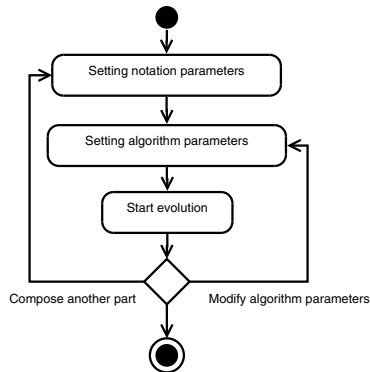
Fig. 5. Process of music composing with the use of EvoMusComp

measure count and key signature. Then parameters of the evolutionary algorithm are set. Evolutionary algorithm can be interrupted by the user and parameters can be modified interactively.

3   Experimental Results

In this section we will present the results of experiments carried out with the use of system presented in the previous section. The experiments were conducted in order to verify whether the system generates sensible results (music pieces) and which operators (and to what extent) give more satisfying results.

The results of this experiments are very promising. With the use of implemented system user can create an original piece of music. With the given melody or music fragment (theme) one can also compose the variation.

We can also mix settings of the system. At the beginning we can start composing using some basic parameters describing specific kind of note and preferred distance between notes. Then we can take satisfying results as a basis chromosome, which will be used in the next part of the music composition process as a reference point. If we want to extend our composition with an extra voice that wasn't specified at the beginning and restart the process of evolution, we can do it easily. So the implemented system is quite flexible and can serve as a composer of completely new pieces of music or variations on the given theme. Also the user can interact with the system during the process of music composition.

During experiments different aspects were taken into consideration. One of them was the question whether the algorithm is able to generate reasonable results when it runs completely without the user intervention (changing values of parameters) for a given number of steps. Such experiments were conducted with all three selection mechanisms implemented. The results are presented in the Fig. 6. The whole experiment was carried out with the parameters' values shown in the Table 1 (column *selection mechanism*).
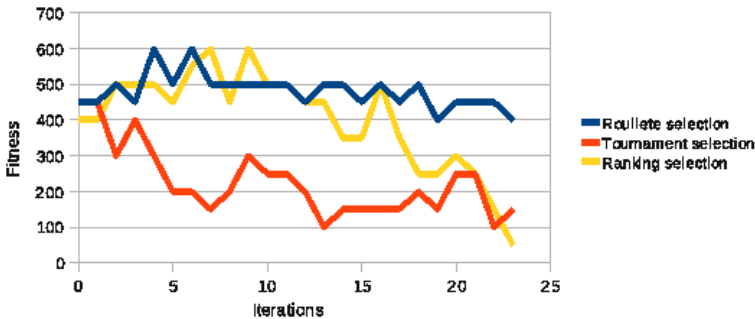
Fig. 6. Fitness of the best individual in consecutive steps of the algorithm for three types of selection mechanisms used and without user interaction (average values from 30 experiments)

Table 1. Parameters' values used in three types of experiments

	Selection mechanism	Time of computations	Mutation probability
Number of iterations (steps)	24	24	11
Number of measures	8	8	8
Genetic operators	crossover, mutation	crossover	crossover, mutation
Number of voices	1	1	1
Fitness function components	1	1	1

The results show that when there is no user interaction the greater number of steps for which the evolutionary algorithm is run do not necessary increase the value of the fitness function when tournament selection is used. In the case of the other two selection mechanisms there is a little progress only during the first 10 steps. These results show that interaction with the user is necessary—the best way of music composing with the use of evolutionary algorithm is to run it for a few steps, then stop the algorithm, modify the values of some parameters and resume the run of algorithm.

The number of individuals present in the population has usually two contradictory effects. Better results are obtained when the population is larger, but it of course slows down computations. The influence of the number of individuals in the population on the time of computations is presented in the Fig. 7. Values of parameters used during these experiments are presented in the Table 1 (column *time of computations*).
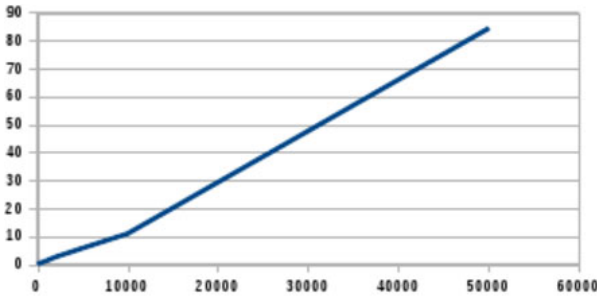
Fig. 7. Time of computations (in seconds) of 24 iterations versus number of individuals in the population. Average values from 30 experiments

The influence of mutation probability on the quality of obtained results is shown in the Fig. 8. Values of parameters used during these experiments are presented in the Table 1 (column *mutation probability*). It can be observed that mutation is very important operator in the presented system. Without mutation there is stagnation in the population. Aggressive mutation ($p = 0.2$) causes that better solutions appear in the population.
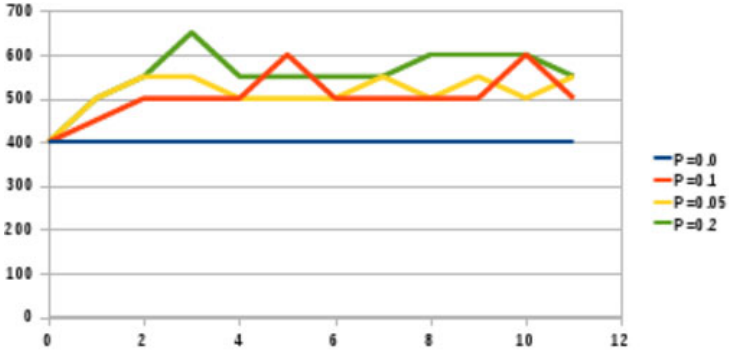
Fig. 8. Value of best individual's fitness for different probabilities of mutation versus number of iterations. Average values from 30 experiments
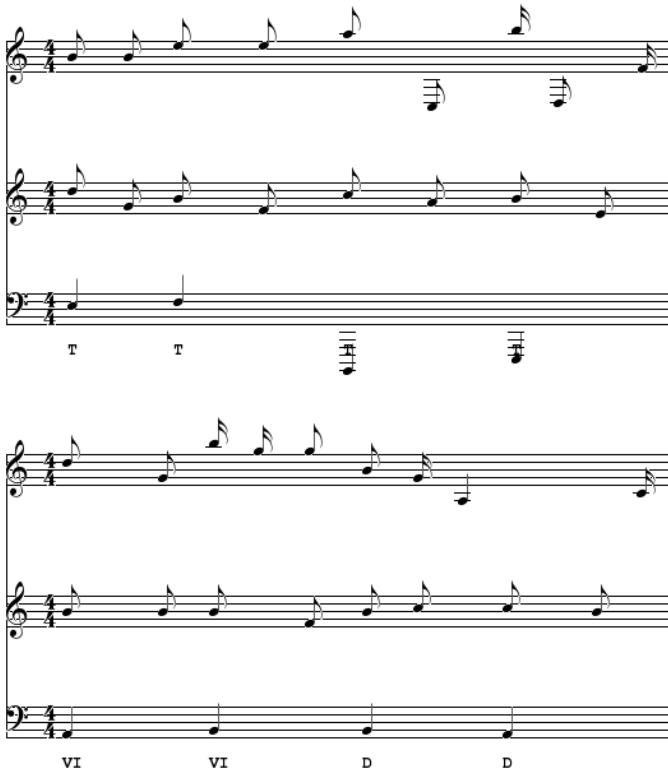
Fig. 9. "Requiem for dream" variation generated by EvoMusComp

The sample of the "real" results obtained with the use of presented system can be seen in the Fig. 9. Presented piece of music is the generated variation on the "Requiem for dream" theme with an extra voice that was also composed with the use of genetic algorithm during the first stage of composing process in the two voices context.

4   Summary and Conclusions

In this paper we have presented the system that can support the user (composer) during the process of music composition. The system can compose completely new pieces of music or generate variations on the given theme.

User can interact with the evolutionary algorithm during initial phase, by setting the values of parameters and by constructing fitness function from the given set of components. User can also interrupt the process of evolution at the selected moment, change the values of some parameters and continue the evolution. As presented experiments show the obtained results are much better when user interacts with the system during its run.

Preliminary experiments are quite promising—the system is able to generate completely new pieces of music or variations on the given theme, as it was presented in the previous sections. Future work will be focused on adding an expert system component which will play the role of human expert during music composition.

Acknowledgments

References

1. Ando, D., Dahlsted, P., Nordahl, M.G., Iba, H.: Interactive gp with tree representation of classical music pieces. In: [8], pp. 577–584
2. Bäck, T., Fogel, D., Michalewicz, Z. (eds.): Handbook of Evolutionary Computation. IOP Publishing and Oxford University Press, Oxford (1997)
3. Biles, J.A.: Genjam: A genetic algorithm for generating jazz solos. In: Proceedings of the International Computer Music Conference (1994)
4. Biles, J.A.: Genjam: evolution of a jazz improviser. In: Creative Evolutionary Systems, pp. 165–187. Morgan Kaufmann Publishers Inc., San Francisco (2002)
5. Budzyński, B.: Algorytmy genetyczne w rozwiązywaniu problemów generowanie muzyki (Genetic algorithms in problems solving—music generation). Zeszyt Naukowy Sztuczna Inteligencja (1) (2002) (in polish)
6. Geem, Z.W., Choi, J.Y.: Music composition using harmony search algorithm. In: [8], pp. 593–600

7. Goldberg, D.E.: Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley, Reading (1989)
8. Giacobini, M. (ed.): EvoWorkshops 2007. LNCS, vol. 4448. Springer, Heidelberg (2007)
9. De Prisco, R., Zaccagnino, R.: An evolutionary music composer algorithm for bass harmonization. In: Giacobini, M., Brabazon, A., Cagnoni, S., Di Caro, G.A., Ekárt, A., Esparcia-Alcázar, A.I., Farooq, M., Fink, A., Machado, P. (eds.) EvoWorkshops 2009. LNCS, vol. 5484, pp. 567–572. Springer, Heidelberg (2009)

Usability Inspection of Informal Learning Environments: The HOU2LEARN Case

Eleni Koulocheri, Alexandros Soumplis, Nektarios Kostaras, and Michalis Xenos

Hellenic Open University, School of Sciences & Technology
Tsamadou 13-15, 26222 Patras, Greece
{ekoulocheri,a.soumplis,nkostaras,xenos}@eap.gr

Abstract. It is argued that a social network can support knowledge transfer to its members with an informal way. It is also argued that a usable, in terms of HCI, social network can do this effectively and efficiently. Therefore, usability evaluation is a significant process that social networks, especially the education-oriented ones, have to undergo in order to be widely accepted and accomplish the task of knowledge transfer. This paper presents a methodology for the usability evaluation of social networks, that was applied to HOU2LEARN, an open educational social network setup in Hellenic Open University for educational and research purposes. The application of this methodology to HOU2LEARN is described and the evaluation results and outcomes are discussed.

Keywords: Informal learning, social networks, usability, heuristic evaluation.

1   Introduction

Granted the pervasiveness of informal learning, the latter is supported by social networks which are widely accepted along with the Web 2.0 proliferation. Informal learning occurs by interaction, sharing ideas, creating content, making new contacts and engagement in critical thinking [14, 21]. Such activities within a social network offer experiences to the users while providing them knowledge in an informal way. Hellenic Open University (HOU), in order to face the challenge of open education through innovative methods, has set up HOU2LEARN (H2L); an open educational platform that aims to identify the learning outcomes that the members gain through their activities.

H2L, as every environment that is used for educational purposes, has an increased demand, providing activities in an effective and efficient way, making the user feel engaged while interacting with it. To that extend, a usability evaluation may reveal usability issues that can be taken into consideration in a future redesign of the platform. In this paper, a detailed usability evaluation method, regarding H2L, is presented. More specifically, the heuristic evaluation method employed, was based on Nielsen's 10 heuristic rules [17], as well as on five additional heuristics specifically created, for such educational platforms, taking into consideration numerous studies using various heuristics. In the case of H2L, the main concern was on the usability of

its interface and not the learnability issues (as was the case in the majority of the studies considered). In brief, the evaluation depicted some aesthetic issues and some problems regarding the menus design and the navigation supported.

Section 2 includes a brief literature review that summarizes the theoretical background and the relation between informal learning, Web 2.0 and social networks. The special issues of learning platforms regarding usability are also discussed. In Section 3, a detailed description of the H2L platform including objectives and features is presented. In Section 4, the evaluation methodology is presented along with the five additional heuristics. Finally, Section 5 summarizes the results of the methodology application in the case of H2L.

2   Literature Review

Researchers in the field of education acknowledge that informal learning is a significant element of education for learners of all ages [3, 21]. While formal learning needs an organised and structured environment, informal learning has a pervasive nature; it is integrated with daily routines related to work, family or leisure with no limits regarding age or institutional structure. It is a low conscious activity of the learner that may be initiated by both internal and external triggers. It doesn't require any kind of structure or organizational support in terms of time and learning as it may occur by chance [14, 15]. Furthermore, while in formal learning, knowledge transfer takes place through teaching, in informal learning it takes place through experience catching and compiling; Davies [4] advocates that a range of learning stimulated by general interests which is "caught, not taught" is also involved in informal learning. Learning by linking to others also results informal learning, given that the latter is an inductive process of reflection and action [14].

In reference to experience catching, a current trend Web 2.0 is a great source of experiences for the user. The user is in control of the content, by being not any more a passive receiver/consumer of information but takes the role of being the actor who publishes his content by himself and also expresses freely [6]. The social aspect of Web 2.0 had a significant impact on the way people connected to each other and led to the bloom of social networks. Using social networks, users collaborate with each other through a variety of group interactions, exchanging information and experience with other users, while sharing ideas, thoughts and interests [23]. The accessibility to and the engagement in social networks have accepted a strong influence by the development and the proliferation of Web 2.0 technologies [14]. It is indicative that one of the most popular social networks, Facebook, has over 500 million active users who spend over 700 billion minutes per month as logged-in users [9].

The activities that a social network provides to the user, are based on three different axes [14]: a. *Personal Profile:* The social network permits the user to share personal information – in the level he wishes – through his profile, letting him decide how and which kind of this information will be visible to other users, b. *Networking:* The social network allows the user to connect with other users of the network, who might be already known as friends or just as contacts and initiate or join sub-sets of user groups based on common interests or pursuits. c. *Content:* The social network enables the user to interact with the content, i.e. to create, upload, tag and

share multimedia content that he has created and link other users to a variety of web-accessible content.

A social network fundamentally enhances information sharing, ideas and thoughts exchange, communication and thus experience assembling. It is evident that informal learning is facilitated by social networks and these networks are a medium that transfers knowledge in an informal way [14]. In order knowledge to be transferred in an effective and efficient way, usability issues of such environments should be deeply investigated, as interfaces with educational aspects are the medium for this kind of knowledge transfer and cannot be considered as a common interface. In order to focus to the medium, Feldestein [10] isolates usability from learnability issues and advocates that usability in learning interfaces concerns the way the content is presented and not just the content solely. Furthermore, for Zaharias et al. [29] an instructional interface should allow the user to focus on learning content rather than focusing on how to access it.

3   The HOU2LEARN Platform

HOU, in the context of software quality research, has set up H2L, an open educational social environment. H2L has been set up in order to monitor the correlation among Learning 2.0 challenges, social networking activities and informal learning. H2L has launched in September 2010, and is part of a wider ongoing research on informal learning environments under the auspices of Software Quality Research Group of the HOU.

H2L is powered by Elgg, an award-winning open source social networking engine, delivering the building blocks that enable educational or business bodies to create their own social networks and applications. It was the software of choice after an extensive evaluation of the Web 2.0 characteristics of existing learning environments [24] as well as its user base and existing community contributing to the project. H2L has been set up in 2010, and it is initially promoted by HOU to the students of the postgraduate course "MSc in Information Systems". H2L is running in combination with the "traditional" e-learning environments (i.e. Moodle, LAMS, IBM Lotus Learning Sametime) that HOU uses in the context of the open and distance education principles. It is vital to note that both tutors and students have the same access rights and this aims to enforce the informal aspect of the platform.

There are three main objectives of the H2L environment. First is to promote openness regarding content creation and sharing in an informal environment. This less typical kind of environment, makes students feel more comfortable to create and share content without the stress of competition or making an error. The second is to enhance communication among the members in a less formal way. This aims to create connections with members who share common educational and research interests. The third objective is to promote socialization among the members, as socialization potentially endorses the exchange of ideas and experience, and therefore the establishment of an informal way of learning. From scientific perspective and within a long term of use, H2L intends to investigate the knowledge transfer procedure, the assessment of knowledge acquired through such an informal environment and the

revelation of strong and weak parts of environments for informal learning based on social networks.

Moreover the activities within H2L can be classified into the following three categories/axes: a. *Personalization* which refers to the ability of users to create their profile as well as their personal ePortfolio, b. *Networking* which refers to the ability of users to follow the activities of other users and create or join groups of users with common interests or ideas. Moreover it is interoperable with other social networks such as Facebook and microblogging tools such as Twitter, c. *Content* which refers to the ability of users to customize the content presented in their dashboard through various widgets (small sub applications).

Another important feature of H2L is that most of the activities provide RSS for content syndication. This allows members of H2L to stay attuned to the latest feeds and along with RSS, enabling publish events directly to their Facebook or Twitter accounts. Pushing out content in such ways, promotes openness as well as ideas and experience exchange, leading to a holistic approach of Informal Learning through Web 2.0. The following figures depict the H2L login page (Fig.1) and a sample of a H2L dashboard (Fig.2):

Fig. 1. H2L Login Page   **Fig. 2.** H2L Dashboard sample

4   Usability Evaluation

4.1   General

Having presented H2L and before going into its evaluation details, it is vital to define usability. The ISO 9241-11 [11] defines usability as "*the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use*". This definition poses three parameters: Effectiveness, that inquires how accurately and completely user achieves specific goals, Efficiency, that measures the resources expanded in relation to the accuracy and completeness with which user achieves goals, and Satisfaction that is related to the freedom from discomfort, the willingness to reuse it and positive attitudes towards the use of the system [13].

According to Nielsen [16], usability is described by answering five fundamental questions, as following: 1. How easily and fast can the user learn the system? 2. Is the

system efficient to use? 3. Is the system easy to remember after some period of not having used it, and how much memory load does its usage require? 4. Does the system have a low error rate and how easily can the user recover from errors? 5. Does the user feel satisfaction -even subjectively satisfaction- by the system usage?

Many methods that measure usability have been developed. These methods require various resources (various numbers of users, users with different level of skills, equipment etc.), they are employed on various phases of the software life-cycle and they are applied in usability laboratory environments or as field studies. In the case of H2L, the method employed was the heuristic evaluation; Heuristic evaluation is a method based on a structured critique of a system using a set of relatively simple and general heuristics (i.e. guidelines or general principles or rules that can be used to critique a decision that has been taken or to guide a design decision) [5]. The main reason that this method has been chosen for the evaluation of H2L, lies in the fact that is a well-established method providing good results with relatively low resources (it is usually called "discount" method) [16]. Furthermore, this method can be applied to systems that are already operational such as H2L. Heuristic evaluation requires a number or 3 to 5 usability experts who, according to Nielsen [16] can reveal 75% of the overall usability problems. These experts (evaluators) judge the system under evaluation, checking whether it complies with the established usability principles mentioned above. Evaluators also use their own experience and point of view regarding the system they inspect.

4.2   The Method Applied

Heuristic evaluation is conducted in two phases [2, 5, 13, 22]: a. *Overall inspection:* The evaluator navigates through the interface for several minutes in order to get familiar with it, as well as with the flow of the interaction and the general scope of the system, and b. *Focused inspection:* The evaluator goes through the interface several times, inspecting various dialog elements. Usually, he is asked by the usability coordinator to follow a scenario and this facilitates the whole process. While the evaluator executes the scenario, he compares the system's behavior with the list of heuristics and notes any violation in an evaluation form.

A list of heuristics that fit into the majority of web-based systems and provide an effective coverage of the most common usability problems [5], has been suggested by Nielsen and Mack [17] as follows:

1. *Visibility of system status*
2. *Match between system and the real world*
3. *User control and freedom*
4. *Consistency and standards*
5. *Error prevention*
6. *Recognition rather than recall*
7. *Flexibility and efficiency of use*
8. *Aesthetic and minimalistic design*
9. *Help users recognize, diagnose, and recover from errors*
10. *Help and documentation*

At this point, it is vital to note that this list is not static; Researchers in the field of usability [5, 16, 22] advocate that it can be modified depending on the domain of the system under evaluation.

Therefore, in order to "customize" the evaluation according the three axes that an open educational social platform such as H2L is based on, and taking into consideration numerous evaluation studies using various heuristics [1, 8, 12, 18, 19, 20, 25, 26, 27] the abovementioned list has been enriched with five additional heuristics (specifically created for the usability evaluation of e-learning platforms):

11. Customisation: Customisation of the content: The presentation of the content should be accomplished through multiple ways, contributing to a customization, without confusing the user, however. Content should be provided with alternative forms and all representations and metaphors should be meaningful while the transition from one form to the other should be performed with a simple and visible way. Customisation of the platform: The platform should make user feel free to change the settings according his preferences, giving thus, a personal "look and feel" in the user interface. This heuristic also checks if the platform allows user to present and manage his personal profile effectively, having the control of what is published and to whom.

12. Navigation: This heuristic rule checks if the navigation in platform is conducted with a clear and comprehensive way. User should be able to know at any instant where he is and where he can go to. Ideally, the system should inform him visually, on the available option he has. Furthermore, navigation should be supported with fidelity; the representation of elements of real world and the provision of complicated options should be controlled. Lack of fidelity may lead to user attention destruction and consequently user disorientation.

13. Interactivity: Interaction with the content: The platform should support interactivity between user and content, with a clear but pervasive way. It should be obvious that the content is dynamic and everyone can contribute to its creation whenever he wishes to. Platform should allow the community of users to create content jointly, promoting thus, collaborative learning, while it provides all interaction options in the right location on the interface. Interaction with peers: In an open social platform, interaction with peers contributes to the ideas and experiences sharing. Therefore, it should be provided effectively and unobstructedly giving the user the feeling that he deals not only with the content but with a community of users ready to connect and collaborate with him.

14. Tools and Multimedia integration: This heuristic rule judges how easy is the installation of new tools and widgets that will increase interactivity both with content and in peers. The platform should allow user to manage the tools changing their location on the user interface. Furthermore, the platform should be efficiently and effectively integrate external media and tools allowing content exporting.

15. Motivation – Engagement: In order to keep a social network alive, user should feel satisfaction while interacting with it, engagement so as to use it for long time and motivation so as to use it actively. Therefore, through multiple but discreet ways, platform should encourage users to be actively logged in for longer.

4.3   The Experiment

During the evaluation of H2L, 4 evaluators were involved; 2 of them were usability experts with more than 7-year scientific experience in the domain and the rest of them had significant experience in heuristic evaluation. In order to avoid biased results, evaluators did not have any previous experience with H2L platform (setting up or interaction with the platform) before the evaluation.

Prior to the evaluation, the coordinator of the process, made an oral, brief presentation of the H2L, and distributed the evaluation scenario to each evaluator. The scenario had been designed by the usability coordinator is such way, so as the evaluation lasted between one and two hours for each evaluator. The scenario was also divided into three sections following the three axes (Personal profiling/Personalisation, Networking, Content) aforementioned.

The evaluators were encouraged to follow the tasks of the scenario and to validate the implementation of each heuristic rule. Each time a heuristic rule was violated, the evaluator identified the part of the scenario where the violation occurred, keeping notes in an evaluation form.

The couple of usability experts reported 23 and 19 usability problems respectively, that lead to 41 and 41 violations of the heuristic rules. The other two evaluators detected 19 and 12 problems respectively that lead to 33 and 27 violations. 52% of the reported problems were reported by more than one evaluator. Considering these duplications as a single problem, the aggregation of the evaluation forms, shows a total number of 35 usability issues that correspond to 68 violations of the heuristic rules, as the following table depicts:

Table 1. Number of violations per heuristic rule

Heuristic Rule	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Number of violations	5	9	1	11	0	7	5	10	1	2	2	10	2	2	1

Hereinafter, a number of characteristic usability issues per heuristic rule are presented and discussed:

1. Visibility of system status: The evaluators mentioned that the system didn't inform them about the progress of file uploading. Furthermore, when they visited a page of another user, they found that the "Follow this Person" button is not adequately visible and easy to find.

2. Match between the system and the real world: For the evaluators, it was not clear the difference between "Dashboard" and "Home". "Tools" is not an indicative term as it represents a menu, in fact. The "Tools" drop down menu and the menu next to "Home" include the same terms that they don't lead to the same actions, however. Moreover, the term "Wire post" is reported to be unfamiliar. Another problem detected is that the envelope icon that leads to the user inbox is not visible enough and this obstructs communication among the members.

3. User control and freedom: In general, H2L doesn't limit the user control, but a major problem reported by the evaluators, was the lack of "Back" button.

4. Consistency and standards: In this case, the evaluators found that "Home" button should be placed in a more usual place, such as up and left next to the "Dashboard". Furthermore, some actions do not lead to the same result; Clicking on a "Following" icon sometimes leads to his profile page and sometimes to content that he has created.

5. Error preventions: The evaluators didn't report any problem related to lack of error prevention.

6. Recognition rather than recall: For some actions, H2L requires some memory load as the evaluation reports showed; "ePortfolio" was expected to be in personal profile page. "Tools" is an 11-level dropdown menu with a structure similar but not the same as the menu that is next to "Home". Furthermore, it is not mentioned which menu tab is active.

7. Flexibility and efficiency of use: The evaluators reported lack of efficiency when integrating locked Twitter accounts. They also mentioned that "Dashboard" and "Tools" menu are not visible, decreasing flexibility. Moreover, problems on searching new contacts decrease efficiency on networking.

8. Aesthetic and minimalistic design: The evaluators found that upper menus design is not clearly visible and furthermore, it is unnecessary to divide the menu into two parts "Dashboard" etc on the left, and "Home" etc on the right. Some buttons such as the "Publish" buttons and some menu icons that appear as a column on the content creation pages don't have a uniform style confusing, thus, the user.

9. Help users recognize, diagnose and recover from errors: The only problem reported regarding help recovering from errors is that no help provided when setting up a locked Twitter account as already mentioned before.

10. Help and documentation: The evaluators remarked that within H2L, the word "Help" is missing. They reported that some support was expected, especially during tasks that they couldn't perform. Furthermore, the search tool for finding contacts didn't work with a consistent way.

11. Customization of the content: The evaluators reported that the "Edit" link on the "Home" page was confusing as it doesn't lead to content creation as expected. They also reported that any customization related to contacts, such as notifications settings, was expected to be found on every contact's page. The choices of the right menu ("Home") provide content in a different way as the "Tools" menu does, but they reported that this should be presented in a more clear way.

12. Navigation: After login, it wasn't clear for the evaluators, which page was open as there was not such a navigational indication. They expected the "Home" to be the anchor-page. Moreover, none of the evaluators managed to find the page which includes all latest feeds, remarking that the navigation options are not of much assistance, due to the division of the menu options into the left and the right part.

13. Interactivity: The evaluators expected that interaction among contents to be more active; in some cases, contacts' usernames didn't appear through mouse-over and sometimes, options for networking were not visible. The problems with contacts searching, influenced interactivity among users, as the evaluators advocated.

14. Tools and multimedia integration: The evaluators reported the lack of efficiency in the integration of non-public Twitter accounts; Dialogue boxes with technical terms were poped-up.

15. Motivation-Engagement: In general, the evaluators mentioned that they felt engaged to use the H2L platform, but they reported that the aforementioned problems related to networking facilities, influenced their motivation.

5 Conclusions – Future Goals

This paper presented the process and the results of the heuristic evaluation of an education-oriented social network platform named HOU2LEARN. The results of the conducted heuristic evaluation presented 35 usability problems. The heuristic rules that were used, were modified in order to cover the specific axes that social networks are based on. The rules were presented, along with the applied methodology description. The evaluation revealed lack of visibility for networking actions, as well as help support and documentation. Furthermore, it seems necessary to reposition the menus in a more concise way and rename of some menu options. Some navigation problems detected and need to be taken into consideration in a future redesign of H2L.

As a future work, on-going research of a further usability evaluation using a more objective method, such as eye-tracking, will lead to the comparison and integration of the results of these different methods. The usability team of the Software Quality Research Group of the Hellenic Open University will aggregate the findings of these two methods, and propose solutions, contributing to the enhancement of the H2L usability. From learning perspective, it is planned to quantize and assess the informal learning "produced" through such an environment, aiming to show that a usable informal learning environment such as H2L is able to provide knowledge that can be assessed even through formal procedures.

References

[1] Ardito, C., Costabile, M.F., De Marsico, M., Lanzilotti, R., Levialdi, S., Roselli, T., Rossano, V.: An approach to usability evaluation of e-learning applications. Univ. Access Inf. Soc. 4, 270–283 (2006), doi:10.1007/s10209-005-0008-6

[2] Avouris, N.: Human Computer Interaction. Hellenic Open University Publications (2003) (in Greek)

[3] Colley, H., Hodkinson, P., Malcolm, J.: Informality and Formality in Learning. In: Vorhaus, J. (ed.) Learning and Skills Research Center, Transforming Learning Cultures in Further Education, London (2003)

[4] Davies, P.: Formalising learning: the role of accreditation. The Policy Press (1998); Book chapter in The necessity of informal learning, Edited by frank Coffield (2000)

[5] Dix, A., Finlay, J., Abowd, G., Beale, R.: Human-Computer Interaction, 3rd edn. Pearson – Prentice Hall, London (2004)

[6] Downes, S.: The Buntine Oration: Learning Networks (2010), `http://www.downes.ca/cgi-bin/page.cgi?post=20` (accessed December 23, 2010)

[7] Elgg (2010), `http://www.elgg.org` (accessed December 23, 2010)

[8] Evans, C., Sabry, K.: Evaluation of the Interactivity of Web-based Learning systems: Principles and process. Innovations in Education and Training International 40(1), 89–99 (2003)

[9] Facebook (2010),
http://www.facebook.com/press/info.php?statistics
(accessed December 23, 2010)

[10] Feldstein, M.: Tutorial: What is "usable" e-learning? eLearn Magazine (9), 2 (2002)

[11] ISO 9241-11, Ergonomic requirements for office work with visual display terminals – Guidance on usability (2003)

[12] Karoulis, A., Pombortsis, A.: Heuristic Evaluation of Web-based ODL Programs. In: Ghaoui, C. (ed.) Usability Evaluation of Online Learning Programs. Hershey P.A. Information Science Publishing (2003)

[13] Kostaras, N., Xenos, M.: Assessing Educational Web-site Usability using Heuristic Evaluation Rules. In: Proceedings of 11th Panhellenic Conference in Informatics, Patras, May 18-20, pp. 543–550 (2007)

[14] Lockyer, L., Patterson, J.: Integrating Social Networking technologiews in Education: A Case Study of a Formal Learning Environment. University of Wollongong (2008), http://ro.uow.edu.au/edupapers/73 (accessed December 23, 2010)

[15] Marsick, V.J., Watkins, K.E.: Informal and Incidental Learning. New Directions for Adult and Continuing Education 2001(89), 25–34 (2001)

[16] Nielsen, J.: Usability Engineering. Academic Press, London (1993)

[17] Nielsen, J., Mack, R.L.: Usability Inspection Methods. John Wiley & Sons,Inc., New York (1994)

[18] Norman, D.A.: Thinks that make us smart – Defending human attributes in the age of the machine. Perseus Books (1994)

[19] Reeves, T., Benson, L., Elliott, D., Grant, M., Holschuh, D., Kim, B., Kim, H., Lauber, E., Loh, S.: Usability evaluation and instructional design heuristics for e-learning evaluation. In: Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2002, pp. 1615–1621 (2002)

[20] Schunk, D.H.: Learning Theories: An education perspective, 3rd edn. Prentice Hall, Englewood Cliffs (2000)

[21] Selwyn, N.: Web 2.0 applications as alternative environments for informal learning - a critical review. In: Background paper for the CERI-KERIS International Expert Meeting on ICT and Educational Performance. Cheju National University, South Korea (2007)

[22] Sharp, H., Rogers, Y., Preece, J.: Interaction Design, Beyond Human-Computer Interaction, 2nd edn. Wiley, Chichester (2006)

[23] Shirky, C.: Social Software and the Politics of Groups (2003),
http://www.shirky.com/writings/group_politics.html
(accessed 23 December 2010)

[24] Soumplis, A., Koulocheri, E., Kostaras, N., Karousos, N.: The evolution of e-Learning 2.0. In: Social Applications for Lifelong Learning, Patra, Greece, November 4-5, pp. 36–41 (2010); ISSN: 1792-586x

[25] Ssemugabi, S., de Villiers, R.: A Comparative study of two usability evaluation methods using a web-based e-learning application. In: Proceedings of the 2007 Annual Research Conference of The South African Institute of Computer Scientists and Information Technologists on IT Research in Developing Countries. ACM International Conference Proceeding Series, vol. 226 (2007)

[26] Squires, D., Preece, J.: Predicting quality evaluation in educational software. Evaluating for learning, usability and the synergy between them, Interacting with Computers 11, 467–483 (1999)

[27] Vrasidas, C.: Issues of Pedagogy and Design in e-learning systems. In: ACM Symposium on Applied Computing (2004)

[28] Wellington, J.: Formal and informal learning in science: the role of the interactive science centres. Physics Education 25, 247 (1990)

[29] Zaharias, P., Vassilopoulou, K., Poulymenakou, A.: Designing On-Line Learning Courses: implications for Usability. Scientific Journal on Applied Information Technology (2001)

Procedural Modeling of Broad-Leaved Trees under Weather Conditions in 3D Virtual Reality

Margarita Favorskaya, Alexander Zotin, and Anastasia Chunina

Siberian State Aerospace University, 31 Krasnoyarsky Rabochy,
Krasnoyarsk, 660014 Russia
favorskaya@sibsau.ru, zotin@sibsau.ru, chunina_aa@sibsau.ru

Abstract. The realistic trees modeling is required in many virtual systems such as architecture modeling packages, widely spread simulators and computer games, large-scale realistic maps in geographic information systems (GIS), and also 3D-modeling packages of landscapes and forest monitoring. Existing L-system (even stochastic L-systems) considers a growing effect without degenerating artifacts that leads to closely visual results when all trees are looked like as a sample. We suggest a procedural approach including an initial chose procedure, a growing procedure, and a degenerating procedure. Also we discuss an influence of some weather conditions and forest modeling for improvement of virtual results. We use the main advantage of L-systems (a quick process of 3D-visual "growing") and achieve a reality visualization in landscape applications.

1 Introduction

The design of 3D-objects of growth is one of complex modeling tasks because of geometrically complex structures and great amount of growth kinds. Existing methods of trees modeling are classified into two wide categories: (1) methods used on an initial set of 2D images, and (2) methods based on mathematical rules of object generation. In the first case the modeling process is divided into three stages: receiving of object's images, segmentation and modeling of foliage, at least segmentation and modeling of trunk and branches [1]. One needs to receive near 30 images of a single object from various points of view without shadows by digital non-calibrated camera. Then an image segment with well-defined leaf structure is manually chosen as a sample. Leaf modeling includes a creation of 2D-model, a contour deformation of 2D-model, a creation of 3D-model, a common form deformation, and a form texturing. Usually a special package is used for modeling of trunk and branches with such operations as arcs' drawing, displacement, editing, connections based on input image segmentation. After that broad-leaved mass is covered on 3D-model. The main disadvantage of such approach is a manually drawing of trunk and branches, but it is possible to model of really existing in nature growth. In the second case methods are subdivided on context-free, deterministic, stochastic, context-depended and parametric methods. The most of them are fractal based methods. The simplest context-free and deterministic systems use the determined values of angles, lengths and basic vectors in 2D- and 3D-spaces. More complex stochastic systems assign probability

values for decision rules. In context-depended systems the rule syntax is complicated by analysis of adjacent elements. In parametric systems we may store parameters of modeling branch (angle, length, thickness), check the conditions for rule application, and set a number of iterations. Advantages of such approach are compact mathematical models and modeling of trees growth considering biochemical processes and the environment. A disadvantage is a complex modeling for men who are not experts in such sphere.

2   Related Work

Realistic modeling is directly depended from possible interactions between various elements into a tree structure, and between trees structures in the environment (last modeling is a very complex task). That's why at present simpler models are used. One of the first famous models suggested by H. Honda was the model with following strong assumptions [2]:

- Tree segments are lines, squares of their crosscuts are not considered;
- During iteration a parent segment divides into two child segments;
- Lengths of two child segments are in r_1 and r_2 times more shooter then a parent segment;
- A parent segment and its two child segments are situated in one plane of branching, and child segments come out from a parent segment with branching angles α_1 and α_2;
- Under action of gravity force a plane of branching is closely to a horizontal plane except of branches connecting with a trunk, in this case a constant angle α is used.

In 1968 Hungarian biologist A. Lindenmayer suggested a mathematical model for development simulation of simple multicellular organisms which was later extended for modeling of complex branching structures – trees and flowers. Such model is known as Lindenmayer System (L-system or Lindenmayer grammar) [3]. The Lindenmayer grammar is similar to Chomsky hierarchy language but with simultaneous application of production rules. Each iteration process is represented by strings denoting by symbols X_n, X_{n+1},… in a formal language. The strings do not contain geometrical data: additionally drawing rules are necessary to translate such strings to morphological descriptors. Many well-known classical fractal curves are presented by L-system. An L-system is defined by a triple

$$\mathbf{G} = <\mathbf{V}, W, PR> , \qquad (1)$$

where $\mathbf{V}$ is a set of symbols; W is a starting string or axiom; PR is a production rule.

Expression (1) determines so-called deterministic L-system. Trees generated by such expression have identical shapes and artificial regularities in spite of modifications of geometrical parameters – trunk length and branching angles. In this case the topology of modeling tree remains unchanged. A stochastic L-system is constructed according to a stochastic grammar

$$\mathbf{SG}_d = <\mathbf{V}, W, \mathbf{PR}, d> , \qquad (2)$$

where $\mathbf{V}$ is an alphabet; W is an axiom, $\mathbf{PR}$ is a set of productions; a function $d : \mathbf{PR} \rightarrow (0,1)$ is a probability distribution, it converts a set of productions $\mathbf{PR}$ into the set of production probabilities. The sum of probabilities of all productions for any letter $a \in \mathbf{V}$ is equal to 1.

Another method for simulation of biological objects is based on Iterated Functional Systems (IFS) suggested by M. F. Barnsley [4]. A first component is a finite set of mappings in 2D- or 3D-space $\mathbf{M}=\{M_1, M_2, \ldots, M_m\}$; a second component is a set of corresponding probabilities $\mathbf{P}=\{P_1, P_2, \ldots, P_m\}$ where the sum of probabilities is equal to 1. Fractal objects are generated by randomly choosing mappings from set $\mathbf{M}$. The iteration process starts from a point z_0 and a mapping M_i with probability P_i resulting in $z_1 = M_i(z_0)$, and so on. The initial points z_0 are assigned from points of attraction in the case the mappings are limited and the process is converged.

Some authors consider a realistic modeling of plants as automatically rendering pen-and-ink illustrations of trees [5] which are usually created by combining brush or pencil strokes. Trunk and branches up to the second order are presented by an assembly of generalized cylinders which 2D-projections are well described by analytical silhouette algorithms with corresponding shadows and texture. Significant attention is concerned for realistic drawing of several thousand of leaves. The authors do not built a tree model but made a single 2D-illustration with the foliage conditionally divided into three areas. The top of a tree is usually under the direct Sun rays and must be visualized with maximum details. In the middle (half shadow) leaves are also drawn in detail. Bottom of a tree is a shaded part; it may be invisible in a single illustration. That's why this area is not elaborate drawing. Such approach connects with 3D visual effects of shadows on foliage for a single illustration but not for 3D-scenes.

P. Tan et al. developed a simple sketching method for realistic 3D-model of a tree from a single image [6]. They suggested a manual drawing of strokes on an image to create a tree model starting from crown, branches and finishing a main trunk. The branch thickness is computed by varying a circle radius to find a largest circle whose pixels are all branch pixels. This thickness computation is unreliable for small twigs. That's why authors simply set a branch radius to 75% of its parent branch. Interesting variant of foliage extraction was suggested. They use a Gaussian mixture model (GMM) in crown region to identify color crown from background. Ten Gaussian distributions are computed and only four the most green or red Gaussian components are covered to leaf cluster. GMM's $G(I_x, \Omega^F)$ and $G(I_x, \Omega^B)$ are composed foreground and background segments in crown where I_x is a function of brightness in RGB-space along axis OX, Ω^F and Ω^B are GMM parameters. For each pixel x label β_x is computed, $\beta_x \in [0,1]$ via graph cut where $\beta_x = 0$ represents leaf pixels and $\beta_x = 1$ represents background pixels.

T. Ijiri, S. Owada, and T. Igarashi suggested a tree modeling system based on L-system that allows the user to control the overall appearance and the depth of recursion during growth process [7]. A user determines a growth direction as a drawn stroke, and a system gradually advances the growth simulation. Also a sketching interface creates a rough model which is quickly rebuilt under a user's intentions. So, here we see an application of adaptive generating rules but with the help of user.

The remainder of this paper is organized as follows. In Section 3 we introduce some procedures which are improving a reality of a tree life-cycle. Section 4 contains a modeling of broad-leaved trees under some weather conditions, and Section 5 presents a forest modeling. In Section 6 we tell about our experimental program, and in Section 7 we give conclusion.

3   Procedural Modeling of Broad-Leaved Trees

Let's extend stochastic grammar using some procedures which include natural features of growth and aging of trees. Let's transform an axiom W from expression (2) to an initial chose procedure PR_W which calculates a set of parameters

$$PR_W = \{\mathbf{Pb}_0, \mathbf{Ab}_0, \mathbf{Tb}_0, \mathbf{Lb}_0\} \,, \tag{3}$$

where $\mathbf{Pb}_0$ is a central point of branch 0 with coordinates (x,y,z); $\mathbf{Ab}_0$ is a set of angle values in 3D-space in central point $\mathbf{Pb}_0$; Tb_0 is a set of thickness of branch 0; Lb_0 is a set of proposed length of branch 0. Extended set (3) maintains parameters of branch 0 in experimental interval estimations for modeling a hardwood containing a certain kind of trees.

It is well-known a monopodial and sympodial structures of trees [3]. In this paper we'll discuss a sympoidal structure of broad-leaved trees which supposes a uniform growth of all branches. (A monopodial structure typifies for firs, pines, cedars and other conifers.)

In fractal model a fractal "branching" is interpreted as a tree growth but it's not right because thickness and length of sub-branches are also increased. Let's introduce a growing procedure PR_G which is periodically restarted and recalculates parameters of all branches and trunk simulating annual tree volume:

$$PR_G = \{GF_i, GL_i, \mathbf{Pb}_i, \mathbf{Ab}_i, Tb_i, Lb_i\} \,, i \in (0,N) \,, \tag{4}$$

where GF_i is a growing factor of i branch (in a simple case it may be a constant); GL_i is a variable calling a foliage increasing; $\mathbf{Ab}_i$ is a set of angle values in 3D-space in central point $\mathbf{Pb}_i$; Tb_i, Lb_i are a thickness and a length of i branch increased by a growing factor GF_i correspondingly; N is a number of visible branches. Procedure (4) makes a growing tree more realistic; it's especially important for trees' modeling without foliage.

A process of growing is not a single process during trees life-cycle. Under some biological artifacts or weather conditions a tree changes its branch structure that leads to degenerate the parts of branches until a whole tree will not become a dry tree. Let's determine a degenerating procedure PR_D which recalling frequency is connected with the age of a tree. We may call such procedure non-periodically during a tree growing but during a tree drying procedure PR_D will stop a development of tree' branches and foliage. So,

$$PR_D = \{DF_i, DL_i, \mathbf{Pb}_i, Tb_i, Lb_i\} \,, i \in (0,N) \,, \tag{5}$$

where DF_i is a degenerating factor of i branch $DF_i = (0, 1]$, if $DF_i = 0$ then i branch and its foliage are implicity destroyed ; DL_i is a variable calling a foliage decreasing;

Tb$_i$, *Lb$_i$* are a decreasing of thickness and a length of *i* branch by a degenerating factor *DF$_i$* correspondingly in point **Pb$_i$**; *N* is a number of visible branches.

We apply procedures (3) – (5) for simulation of a broad-leaved forest, also procedures (4) and (5) are used for simulation of season foliage effects. More advisable approach for foliage imitation is an object approach for creation a sample of a leaf (for a given tree kind) with available morphed transformations during leafs appearance in spring. Leaf falling simulation in autumn is connected with a speed and character falling of leaves having various square values. Then foliage store on the Earth' surface and change a texture of a 3D-surface scene.

A single tree modeling is accomplished by algorithm 1:

Initial data: kind of tree, initial parameters of trunk, crown, samples of leaf texture.

Step 1. Run an initial chose procedure (3).

Step 2. Generate of tree structure (based on L-system) by using initial parameters of a tree kind and results from Step 1.

Step 3. Create foliage and cover foliage texture.

Step 4. Visualize an object in 3D-scene.

Step 5. If it is necessary, run a growing procedure (4) and rebuild an object (Step 2 – Step 4).

Step 6. If it is necessary, run a degenerating procedure (5) and rebuild an object (Step 2 – Step 4).

Output data: 3D visualization of a single tree.

Algorithm 1 is a simple case of a tree modeling. Let's discuss a modeling of trees under weather conditions and a forest modeling.

4   Modeling of Tree under Weather Conditions

Many natural effects are concerned to weather conditions such as wind, rain, fog, snow, also a day luminance [8]. The often used natural effect is a wind simulation with various forces. In this paper we do not consider unusual natural disasters [9]. We'll discuss a modeling of wind and luminance influence on a tree structure generation during some years. Also we'll consider a simulation of branches' flexure under a wind.

There are some woodland on the Earth' surface which are characterized by a stable wind rose. We may interpret such influence as a set of local geometrical barriers. The barriers distort a trunk, branches and their locations along a prior direction of stable wind rose. In this case a distorted tree model is actively generated by user corrections. When we have received a basic distorted tree model we replicate it as a fractal structure with an initial chose procedure (3). Thereby we simulate woodland with inclined trunks of trees and clearly defined directions of branches.

Usually a wind is not strong, that's why we'll discuss in detail a branches' flexure (if a wind is very weak then it is enough to make small changes in leaves texture from frame to frame – a simulation of leaves blinks). It is ought to calculate a controlled correction simulating a flexure in branches coordinates. One approach consists in calculation of rotating matrix round an axis passing via branch origin and normal to a wind force vector. However a branch consists from sub-branches (according to fractal modeling), and a number of such matrixes makes impossible the calculations in real

time. We suggest a simpler but less exact approach connecting with displacement of branch in a required direction. The main task is to find such direction.

We may suggest that a vector of wind speed can not have a vertical direction. That's why it is required to find projections of such vector (wind of "unit force") for every branch along directions *OX* and *OY* in a horizontal plane. A direction of displacement is a normal relatively direction of branch because a branch can not change its length. So, if we know a direction of a wind we'll find a direction of branch displacement as a product of vectors. A length of displacement vector is proportional to a length of branch. Thereby we found own oscillations of branch. In common case a branch has oscillations of load-bearing branches which are calculated similarly and then summarized (a trunk is stationary). We may take into account a thickness of branches setting a constant which is depend from branch level in recursion. For realistic effect let's add a wind randomness. This parameter is well described with a noise suggested by K. Perlin [10]. Imitation of oscillations lag of branches (during a change of a wind direction) may be realized as a gradual change of a wind direction by linear interpolations between adjacent frames. Also we can add a resonance random moving of wands depending from a wind force as an additional constant in calculations.

Generalized algorithm of tree modeling under weather conditions has following steps (algorithm 2):

Initial data: L-model of a tree, parameters of main natural effects.
Step 1. Choose a natural effect.
Step 2. Determine barriers distorting a tree development if it is necessary.
Step 3. Recalculate L-model of a tree by using parameters of a chosen effect.
Step 4. Visualize an object in 3D-scene.
Output data: 3D visualization of a single tree.
In Sections 3 and 4 we considered only a model of a single tree but a forest modeling is based on another model – a model of interactions.

5   Forest Modeling

We consider that a forest modeling is based on following assumptions. Model-based space in divided into 3D-cellular structure so that a crown of one tree may situate only in one cell or as exception in adjacent cells in horizontal or vertical planes. The sizes of such cells are chosen by analysis the experimental data of realistic forest. As a result each tree is associated with any cell, has its space coordinates that permits to simplify essentially a forest model and decrease an execution time of algorithm. A day luminance for each tree is well modeled by space division on such 3D cells. Spatial structure of forest model is presented on fig. 1.

The basic structure unit is a tree, and forest modeling is accumulated the results of dynamic modeling of a separate trees. According to this assumption we consider an individual development of some tree with a spatial distribution of other trees. In such model we set local barriers (as a part on planes under different angles) near growth tree if we have a closely located an old tree with a big crown. So we use L-system with geometrical limitations when a tree can not well develop in restricted zones. For such weak trees a degenerating procedure (5) is recalled often when for normal trees.
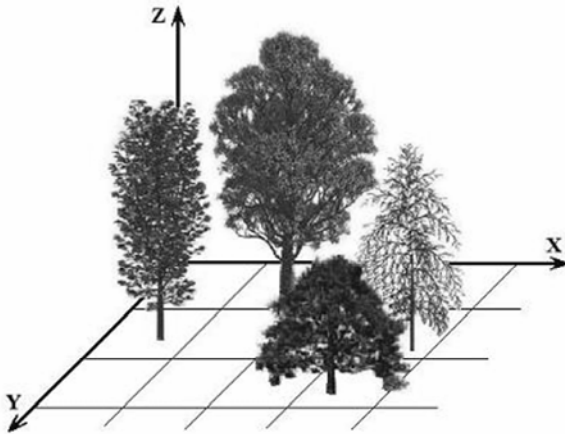
Fig. 1. An example of a forest modeling as a spatial structure

The dynamic modeling of a single tree requires the assessment of the influence of other closely situated trees; otherwise we watch a competition for external resources: light, water, minerals, location. So the state of tree in some instant depends from kinds and age-specific groups of closely located trees. These parameters are enough for receive a model of continuous varying in time forest' discontinuities. A model of single tree includes three stages: (1) a tree growth, (2) a competition for external resources, and (3) reproduction, aging and collapse. A model of forest includes spatial interactions between trees under possible biological, weather, and human activities. A modeling of forest is not a well investigated process but the experimental results show that processes of chaotic self-organization are occurred even in uniform external conditions. Appearance of such discontinuities is explained by causes of inner spatial competition for life resources.

A forest modeling is accomplished by algorithm 3:

Initial data: kind of trees, L-models of trees and their coordinates, 3D spatial space.
Step 1. Choose sizes of cells and label used cells.
Step 2. Determine barriers distorting a tree development if it is necessary.
Step 3. Recalculate dynamically L-model of trees.
Step 4. Simulate a life-cycle of trees.
Step 5. Visualize dynamically 3D-scene.
Output data: 3D visualization of forest scene.

6   Experimental Program

Experimental program "Trees Editor" includes three main modules: a module of a single tree generation (6 kinds of broad-leaved trees), a module of weather conditions (imitation of wind, luminance, fog, rain), and a module of forest generation. A tree generation is based on L-system with proposed procedures concerning a life-cycle of a tree. A tree model is saved as an object in 3D-scene, and we may apply standard

operations such as rotation, scaling, translation. If it is necessary we call a point light source in scene, and program automatically forms a tree shadow. In module of weather conditions the user may locate "semitransparent" barriers back from which only wands penetrate. We gave consideration a sub-module of wind imitation and received good visual results in real time. A module of forest generation represents a dynamic of growing closely located trees but it is needed in following development for more realistic results. We imitate a Sun moving across a scene with shadow effects and balancing day illumination. At present many functions of this module are accomplished manually. On fig. 2 the main screen of our experimental program is presented.
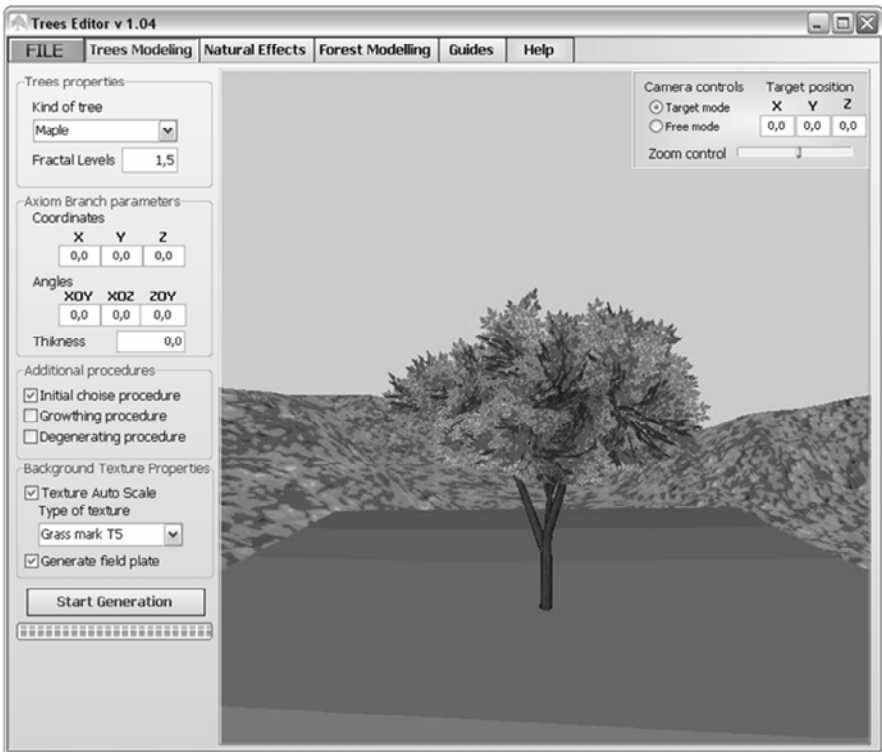
Fig. 2. Main screen of experimental program "Trees Editor"

For program realization we used a graphical jet GLScene applying library OpenGL as application programming interface in development environment RAD Studio 2010. We used some standard objects from the jet GLScene such as TGLCamera (a camera object), TGLSceneViewer (a 3D viewer object), TGLMaterialLibrary (a library of materials), TGLFreeForm (a static 3D model), TGLLightSource (a light source), and some others. Also we designed original procedures and components for forest modeling. Program "Trees Editor" has user-friendly interface but some complex functions need a preliminary instructions reading.

We made experiments for modeling of a single tree in 3D scene (6 kinds of trees – a white birch, a broad-leaved maple, a green ash, a white poplar, a silky willow, a red oak) with various initial parameters and growth algorithms. We tested trees models imitating a wind with various force values and Sun lighting. Modeling of fog and rain is additional function of this program. Also we received relatively simple dynamic examples of forest modeling. Generated virtual scenes have a good visual effects and likely nature image by experts' estimations.

7   Conclusion

Analysis of existing methods of trees modeling shows that more constructive approach connects with application of L-system. We extended a stochastic L-grammar by proposed procedures including natural features of trees' life-cycle. Also proposed growing and degenerating procedures imitate a dynamic development of broad-leaved forest and season foliage effects. We considered a modeling of wind and luminance influence on a tree structure generation during some years and discussed a simulation of branches' flexure under a wind by simpler and less exact approach.

We developed three algorithms for a single tree modeling, for tree modeling under weather conditions, and for forest modeling. Three basic modules of our experimental program "Trees Editor" are built on these algorithms. We generate 3D scenes by a graphical jet GLScene (applying the library OpenGL) and by own designed components. Generated virtual scenes have likely nature visualization and may be use in many virtual reality systems.

References

1. Quan, L., Tan, P., Zeng, G., Yuan, L., Wang, J., Kang, S.: Image-based Plant Modeling. ACM Transactions on Graphics 25(3), 599–604 (2006)
2. Honda, H., Tomlinson, P.B., Fisher, J.B.: Two geometrical models of branching of botanical trees. Annals of Botany 49, 1–11 (1982)
3. Prusinkiewicz, P., Lindenmayer, A.: The Algorithmic Beauty of Plants. Springer, New York (1996)
4. Barnsley, M.F.: Fractals everywhere. Academic Press, Boston (1988)
5. Wilson, B., Ma, K.-L.: Rendering Complexity in Computer-Generated Pen-and-Ink Illustrations. In: NPAR, pp. 103–111 (2004)
6. Tan, P., Fang, T., Xiao, J., Zhao, P., Quan, L.: Single Image Tree Modeling. ACM Transaction on Graphics (TOG) Proceedings of SIGGRAPH Asia 27(5), 1–11 (2008)
7. Ijiri, T., Owada, S., Igarashi, T.: The Sketch L-System: Global Control of Tree Modeling Using Free-Form Strokes. In: Butz, A., Fisher, B., Krüger, A., Olivier, P. (eds.) SG 2006. LNCS, vol. 4073, pp. 138–146. Springer, Heidelberg (2006)
8. Favorskaya, M., Zotin, A., Danilin, I., Smolentcheva, S.: Realistic 3D-modeling of Forest Growth with Natural Effect. Advances in Intelligent Decision Technologies SIST4, 191–199 (2010)
9. Singh, P.A., Zhao, N., Chen, S.-C., Zhang, K.: Tree animation for a 3D interactive visualization system for hurricane impacts. In: ICME, pp. 598–601 (2005)
10. Perlin, K.: An Image Synthesizer. ACM SIGGRAPH 19(3), 287–296 (1985)

New Method for Adaptive Lossless Compression of Still Images Based on the Histogram Statistics

Roumen Kountchev[1], Vladimir Todorov[2], and Roumiana Kountcheva[2]

[1] Technical University – Sofia, Department of Radio Communications and Video Technologies, Boul. Kl. Ohridsky 8, Sofia 1000, Bulgaria
rkountch@tu-sofia.bg
[2] T&K Engineering Co., Mladost 3, POB12, Sofia 1712, Bulgaria
todorov_vl@yahoo.com, kountcheva_r@yahoo.com

Abstract. The method for adaptive lossless image coding is aimed at efficient compression of grayscale or color digital still images. The method is based on the analysis of the processed images histograms, followed by modified Huffman and Run-length coding. The coding is performed in two consecutive stages. In the first one, the input data (i.e., the digital information about the image brightness and color components) is transformed without affecting their volume in such a way, that to obtain sequences of same values (in particular, zeros) of maximum length. The transform is reversible and is based on the analysis of the input data histograms and on the differences between each number in the processed data and the most frequent one. In the second stage of the processing the transformed data is analyzed and sequences of same numbers are detected. Every such sequence is substituted by a shorter one, corresponding to the number of same values which it contains. The method is extremely suitable for images of graphics or texts (for example: fingerprints, contour images, cartoons, medical signals, etc.). The method has low computational complexity and is suitable for real-time applications.

Keywords: Lossless image compression, adaptive run-length coding, histogram-adaptive lossless image compression.

1 Introduction

With rapidly advancing computer technologies, there is huge amount of visual data which should be archived and stored, retaining its quality as high as possible. Lossless image compression is required by medical imaging, satellite/aerial imaging, preservation of special art work and documents, or any applications demanding ultra-high image fidelity. Furthermore, the lossless image coding is the last step of many lossy image compression systems, needed for example, for the lossless compression of the transform DCT coefficients or for the wavelet sub-band coding. The basic approach for archiving of documents and other imagery, which requires retained quality, is to use the famous standards JPEG and JPEG 2000, which answer these requirements to a very high degree. In some cases however, the image quality should be retained unchanged and the images should be losslessly coded. The methods developed to solve

this task are based on some kind of lossless data coding, the most important and widely used of which could be classified as follows: Coding based on dictionaries [1]: Lempel-Ziv; Statistical coding: Run-length [7], Shannon-Fano [3], Huffman [4], Arithmetic [5] (Binary arithmetic coding, QM-coder); Coding based on data trans-forms [2, 6]: Burrows-Wheeler Transform; Predictive coding; Structured universal coding schemes [8] (Elias codes, Exponential-Golomb codes): these schemes generate variable-length code words with a regular structure; Hybrid coding [7, 8] - a mixture of some of the methods, mentioned above: LOCO-I, adopted in the JPEG-LS stan-dard, and others. The coding algorithms, whose parameters are adaptive to some char-acteristics of the input data, are called "adaptive" [9-15].

In this work is presented one new method for adaptive lossless compression of still images based on the histogram statistics, called Adaptive Run-Length (ARL) Coding. The work is arranged as follows: Section 2 introduces the basic principles of the new method; Section 3 focuses on the evaluation of the method efficiency; Section 4 gives some experimental results for lossless still image compression and comparison with other contemporary methods and Section 5 is the Conclusions.

2   Basic Principles of the ARL Coding Method

The Adaptive Run-Length (ARL) coding method is aimed at the compression of im-age data, represented as N-dimensional sequence of n-bits binary words (numbers) x_k for $k=1,2,..,N$, whose values are in the range: $(-2^{n-1}) \leq x_k \leq (2^{n-1}-1)$. The method com-prises two basic steps. The first step should be assumed as a pre-processing, because the data is transformed in such a way, that to enhance the coding, performed in the second step. The image compression is based on new run-length coding of sequences of equal or regularly changing values. In result, each N-dimensional sequence repre-senting the image data is substituted by a shorter one, containing a header and com-pressed data.

The method is described below.

Preliminary processing of the input data

The input data sequence x_k for $k = 1,2, ..., N$ is transformed into the sequence v_k of same length N so, that to obtain sequences of maximum length of numbers with same value (in particular, this value is equal to zero). For this are performed the following operations:

❖   Analysis of the histogram of the processed image (the input data), x_k:
   - The histogram $H(x)$ is calculated for $x = -2^{n-1}, -2^{n-1}+1, ..,-1, 0, 1,.., 2^{n-1}-1$. Here $H(x)$ is the count of the numbers whose value is equal to x.
   - The count $L(x)$ is defined, representing the values in the histogram which are not used (free), and for which $H(x)=0$, when $x = -2^{n-1}, -2^{n-1}+1,..,-1,0,1,.., 2^{n-1}-1$:

$$L(x) = \sum_{x=-2^{n-1}}^{2^{n-1}-1} f(x) \text{ for } f(x) = \begin{cases} 1, & \text{if } H(x)=0, \\ 0, & \text{if } H(x)\neq 0. \end{cases} \tag{1}$$

- Positions $p_i = x_i$ are defined, which correspond to the start of the intervals of free values. The lengths $(\Delta l_i + 1)$ of these intervals in the histogram $H(x)$ are defined, in correspondence with the relation:

$$x \in [p_i, p_i + \Delta l_i] \text{ for } i = 1, 2, \ldots, T(x), \text{ when } L(x) > 0 \text{ and } H(x) = 0. \tag{2}$$

- The interval of free values and of maximum length is detected, as follows:

$$p(x) = p_i \text{ and } l(x) = \Delta l_i = max \text{ for } i = 1, 2, \ldots, T(x). \tag{3}$$

In case that there is more than one interval of free values and of maximal length, then $l(x)$ corresponds to the one, whose start position has smallest value.

❖   The data x_k are transformed into y_k using a special coding, called "size-saving prediction coding" (SSP), specially developed for the lossless coding method. The processing is performed in correspondence with the relation:

$$y_k = \begin{cases} (x_k - x_{k-1}), & \text{if } (-2^{n-1}) \le (x_k - x_{k-1}) \le (2^{n-1} - 1), \\ -(x_k + 1) & - & \text{in all other cases,} \end{cases} \tag{4}$$

for $k = 1, 2, \ldots, N$ and $x_0 = 0$. In result, the sequences of same numbers in x_k are transformed into sequences of zeros in y_k.

❖   The histogram $H(y)$ of the data y_k is calculated for $y = -2^{n-1}, -2^{n-1} + 1, \ldots, -1, 0, 1, \ldots, 2^{n-1} - 1$ analyzed in the way, performed for x_k, and the positions $p_i = y_i$ are defined, which point at the start of the intervals of free values. Their lengths $(\Delta l_i + 1)$ in the histogram $H(y)$ are calculated in correspondence with the relation:

$$y \in [p_i, p_i + \Delta l_i] \text{ for } i = 1, 2, \ldots, T(y), \text{ when } L(y) > 0 \text{ and } H(y) = 0.$$

The longest interval of free values is detected, for which:

$$p(y) = p_i \text{ and } l(y) = \Delta l_i = max \text{ for } i = 1, 2, \ldots, T(y). \tag{5}$$

❖   The conditions $L(x) = 0$ and $L(y) = 0$ are checked, which are satisfied if only there are no intervals of free values in the two histograms. A special flag is set, which indicates if the image data is suitable for compression. This flag is a special bit in the control word in the header of the losslessly coded data, named F_1. In case that conditions $L(x) = 0$ and $L(y) = 0$ are satisfied, the stage of the preliminary transform of the input data is stopped, and the coding ends. In all other cases the flag $F_1 = 1$. After that is checked which data sequence x_k or y_k is more suitable for the lossless coding (i.e. which one will ensure higher compression) The choice is done in accordance with the relation:

$$y_k = \begin{cases} x_k; & F_2 = 0, H(y) = H(x), L(y) = L(x), p(y) = p(x), l(y) = l(x) \text{ if } L(y) = 0 \text{ or } L(x) > 2^n - 4; \\ y_k; & F_2 = 1 & - & \text{for SSP coding.} \end{cases} \tag{6}$$

Here F_2 is a flag (another bit in the control word), which indicates the selected sequence ($F_2 = 1$ if the sequence y_k better suits the coding method).

❖   The value $y = r(y)$, is defined, for which $H(y) = max$, when:

$$y = -2^{n-1}, -2^{n-1} + 1, \ldots, -1, 0, 1, \ldots, 2^{n-1} - 1.$$

❖ The data are modified transforming every word of the sequence y_k into v_k subtracting $r(y)$ without setting carry in correspondence with the relation:

$$v_k = \begin{cases} [\, y_k - r(y)\,] & if \quad [\, y_k - r(y)\,] \in [\,-2^{n-1},\, 2^{n-1}-1\,]; \\ [\, y_k - r(y) - 2^n \,] & if \quad [\, y_k - r(y)\,] > 2^{n-1}-1\,; \text{ for } k=1,2,..,N. \\ [\, y_k - r(y) + 2^n \,] & if \quad [\, y_k - r(y)\,] < -2^{n-1}. \end{cases} \tag{7}$$

❖   The histogram $H(v)$ for $v = -2^{n-1}, -2^{n-1}+1,...,-1,\ 0,\ 1,..,2^{n-1}-1$ of the data sequence v_k is calculated and analyzed in the already described way for x_k. The count $L(v)$ of the free histogram values is also calculated and the positions $p_i = v_i$ are defined, which point at the start of the intervals of free values. The lengths $(\Delta l_i + 1)$ of these intervals in the histogram $H(v)$ are defined in correspondence with the relation:

$$v \in [p_i,\, p_i + \Delta l_i] \text{ for } i = 1, 2, \ldots, T(v), \text{ when } H(v) = 0 \tag{8}$$

The maximum interval of free values is defined as follows:

$$p(v) = p_i \text{ and } l(v) = \Delta l_i = max \text{ for } i = 1,2,..,T(v). \tag{9}$$

❖   The flag F_3 is set, which indicates that the length of the interval of free values is bigger than 1. It is set in the coded data control word, in accordance with the length $[l(v)+1]$ of the detected interval of free values, as follows:

$$F_3 = \begin{cases} 1, if\ l(v) > 0; \\ 0,\ if\ l(v) = 0. \end{cases} \tag{10}$$

With this, the first step of the processing is finished. In result is obtained the transformed data sequence v_k for $k=1,2,..,N$ and is defined the additional information about $r(y)$, $p(v)$ and $l(v)$, which are processed in the second step of the coding.

Data coding

In this part of the processing is composed the header of the coded data, needed for the proper coding/decoding of the transformed data sequence, v_k. This part of the processing comprises the following basic operations:

Composition of the coded data header

The header is necessary for the proper decoding of the compressed data and consists of control word w_{00}, which comprises the three control flags F_1, F_2 and F_3 and additional information, presented below. The header contains additional information, in accordance with the conditions:

- for $F_1 = 0$ the header does not contain additional information;
- for $F_1 = 1$ the header contains additional information, which comprises the numbers $w_{01} = r(y)$ and $w_{02} = p(v)$. *Here $r(y)$ is the most frequently met value in the data sequence y_k, and $p(v)$ is the start position of the longest interval of free values in the data sequence v_k, which was detected first;
- for $F_1 = 1$ and $F_3 = 1$ the additional information in the header includes one more number $w_{03} = l(v)$, which defines the length of the interval with start position $p(v)$, decreased by 1.

Coding of the transformed data

➢ For $F_1 = 1$ the transformed data are processed in accordance with the new method for adaptive lossless coding. For this, each sequence of numbers of same value in v_k is substituted by w_s, and in result is got a shorter data sequence:

- Each sequence of zeros $v_d = v_{d+1} = ... = v_{d+P-1} = 0$ of length P in the range $1 < P \le l(v)$ $+1$, which had been detected in v_k is substituted by one n-bit word $w = p(v) + P-1$, i.e.:

$$(v_d, v_{d+1}, .., v_{d+P-1}) \Rightarrow (w = p(v) + P\text{-}1). \tag{11}$$

For $P = 1$ the sequence contains one zero only ($v_d = 0$) and the coding is not performed;

- Each zero sequence $v_d = v_{d+1} = ... = v_{d+P-1} = 0$ of length P which is in the range $2^{mn} \ge$ $P > l(v) + 1$ for $m \ge 1$, detected in v_k, is substituted by $2m$ words, of n bits each. In the first word is stored $p(v)$, in the next $(m-1)$ words is stored zero, and in the remaining m words – the number $(P-1)$, i.e.:

$$(v_d, v_{d+1}, .. , v_{d+P-1}) \Rightarrow (w_1 = p(v), w_2 = 0, . . , w_m = 0, w_{m+1} = P_1,.., w_{2m} = P_m), \tag{12}$$

where $P_1, .., P_m$ corresponds to the number $(P-1)$, represented by m words of n bits each (here P_1 is the *MSW*,… and P_m - the *LSW*);

- Each sequence of same numbers, not equal to zero $v_d = v_{d+1} = ... = v_{d+P-1} = v$ of length P in the range $2^{mn} \ge P > 4$ for $m \ge 1$, detected in v_k, is substituted by $(2m+2)$ words, n bits each. In the first two words are stored the numbers $p(v)$ and "1", in the next $(m-1)$ words – zeros, in the next m words – the number $(P-1)$ and in the last word – the number, which is different from zero, $v \ne 0$ i.e.:

$$(v_d, v_{d+1}, .., v_{d+P-1}) \Rightarrow (w_1 = p(v), w_2 = 1, w_3 = 0, .., w_{m+1} = 0, w_{m+2} = P_1, .., w_{2m+1} = P_m,$$

$$w_{2m+2} = v), \tag{13}$$

where $P_1, .., P_m$ present the number $(P-1)$, coded as m words of n bits each (here P_1 is the *MSW*,… and P_m - the *LSW*).

Sequences of non-zero values of length $P \le 4$ are not losslessly coded, because this does not enhance the method efficiency.

In result of the so presented coding the v_k data is transformed into the compressed sequence w_s for $s = 1,2,..,S$ and $(-2^{n-1}) \le w_s \le (2^{n-1} - 1)$. Here S is the count of the words in the sequence w_s, which is smaller than the total count N of the words in the original sequence x_k.

➢ For $F1 = 0$ the input sequence x_k is not compressed and after the header follows the original data: $w_s = x_k$ for $s = k = 1, 2, ..., N$ $(S = N)$.

Data decoding

The decoding of the w_s sequence comprises the following operations:

• The flags in the control word of the coded data header w_{00} are analyzed consecutively. Two main cases are possible:

If $F_1 = 0$, this means that the input data sequence x_k had not been losslessly coded and the decoded data u_s is connected with w_s by the relation $u_s = w_s$ for $s = 1, 2, 3, ..., S$ $(S = N)$;

If $F_1=1$, this means that the input sequence x_k had been coded and the following steps are then performed:

Step 1: The flag F_3 is analyzed (for $F_1=1$) and the decoding is performed. In the case, when $F_3=1$ is analyzed the available additional information from the header of the compressed sequence w_s: the numbers $r(y) = w_{01}$, $p(v) = w_{02}$, $l(v) = w_{03}$. In the case, for $F_3 = 0$, this information comprises only numbers $r(y) = w_{01}$ and $p(v) = w_{02}$ (for the decoding $l(v)=0$). After that, every value of w_s for $s=1, 2,..., S$ and $F_3=0$ is compared with the number $p(v)$. Depending on the difference $\delta_s = w_s - p(v)$ when the value of w_s is decoded, it is retained, or substituted by a sequence of numbers of same value, $v_p = v$ for $p = 1, 2,..., P$ in correspondence with one of the followings procedures:

- For $\delta_s > l(v)$ or $\delta_s < 0$, the value of w_s is not changed. In this case is performed the substitution $w_s \Rightarrow (v = w_s)$;

- For $0 < \delta_s \leq l(v)$, the value of w_s is substituted by the sequence v_p, of length $P=w_s-p(v)+1$, which consists of zeros only, and $w_s \Rightarrow (v_1=0 , v_2=0 ,.., v_P=0)$;

- For $\delta_s = 0$ and $w_{s+1} \neq 0$, the data w_s and w_{s+1} are substituted by the sequence v_p of length $P=w_{s+1}+1$, which consists of zeros only, and in result is obtained $(w_s, w_{s+1}) \Rightarrow (v_1=0 , v_2=0 ,.., v_P=0)$;

- For $\delta_s = 0$, $w_{s+1} = 0,..,w_{s+m-1} = 0$, $w_{s+m} \neq 0,..,w_{s+2m-1} \neq 0$, the data in w_s up to w_{s+2m-1} are substituted by the sequence v_p, consisting of zeros only, with length $P = (w_{s+m},..,w_{s+2m-1}) + 1$ (w_{s+m} is the MSW, and w_{s+2m-1} is the LSW) and then $(w_s ,..., w_{s+2m-1}) \Rightarrow (v_1=0 , v_2=0 ,.., v_P=0)$;

- For $\delta_s = 0$, $w_{s+1}= 1$ and $w_{s+2} \neq 0$, the data in w_s up to w_{s+3} are substituted by the sequence v_p, consisting of non-zero numbers $v=w_{s+3}$ of length $P=w_{s+2}+1$, and then $(w_s, . . . , w_{s+3}) \Rightarrow (v_1=v , v_2=v ,.., v_P=v)$.

- For $\delta_s = 0$, $w_{s+1} = 1$, $w_{s+2} = 0,..,w_{s+m} = 0$, $w_{s+m+1} \neq 0,..,w_{s+2m} \neq 0$, the data in w_s up to w_{s+2m+1} are substituted by the sequence v_p, consisting of non-zero numbers $v=w_{s+2m+1}$ of length $P=(w_{s+m+1},..,w_{s+2m})+1$ (w_{s+m+1} is the MSW, and w_{s+2m} is the LSW), and then $(w_s, . . . , w_{s+2m-1}) \Rightarrow (v_1=v , v_2=v ,.., v_P=v)$.

At the end is obtained the decoded sequence v_k.

Step 2. Inverse data modification is performed, which transforms every word v_k into y_k adding $r(y)$ to its value without carry, in accordance with the relations:

$$y_k = \begin{cases} [v_k+r(y)] & if & [v_k+r(y)]\in [-2^{n-1}, 2^{n-1}-1]; \\ [v_k+r(y)-2^n] & if & [v_k+r(y)] > 2^{n-1}-1 ; \ for \ k = 1,2,..,N; \\ [v_k + r(y)+2^n] & if & [v_k+r(y)] < -2^{n-1}. \end{cases} \quad (14)$$

Step 3. The flag F_2 is analyzed when the operations, defined in accordance with flags F_1 and F_3, are finished. If $F_2 = 0$, this indicates that the sequence x_k had not been SSP coded and then $u_k = x_k = y_k$ for $k = 1, 2,..., N$. In case, that $F_2=1$, is necessary to transform y_k into u_k performing the SSP decoding:

$$u_k = \begin{cases} (y_k + u_{k-1}), & \text{if } 2^{n-1} \le (y_k + u_{k-1}) \le (2^{n-1}-1), \\ -(y_k + 1) & - \qquad \text{in all other cases.} \end{cases} \text{ for } k=1,2,..,N \text{ and } u_0=0. \quad (15)$$

In result $u_k = x_k$ and the w_s decoding is finished.

3 Evaluation of the Lossless Coding Method Efficiency

The compression ratio is calculated as a relation between the original and the compressed data, i.e. $CR = N/(S+W)$, where W is the number of the words in the header. The minimum value $CR_{min} = N/(N+1) < 1$ is obtained for the case, when the data x_k are not compressed. In this case $S = N$ and $W = 1$, because the header consists of one word only (the control word). The maximum value of CR is obtained for $x_k = x$ ($k = 1, 2,..., N$). In this case $P = N$, $S = 2m$, $W = 4$ and $CR_{max} \le N/[(2/n)lg_2N+4]$. From this follows that the compression ratio is in the range:

$$\frac{N}{N+1} \le CR \le \frac{N}{lg_2\sqrt[n]{N^2}+4} \text{ ; For } N >> 1 \text{ is obtained } 1 \le CR \le N/lg2\sqrt[n]{N^2}. \quad (16)$$

The analysis shows that the compression ratio can achieve significant values when there are long sequences of same numbers in v_k.

The presented method for lossless image compression is suitable for coding of images, whose brightness or color histograms have free values, or they have large areas of same or continuously changing brightness/color. Such histograms are typical for scanned documents, graphics, fingerprints, medical signals, etc. The processing of each of the color components is performed in the way, described above for grayscale images.

4 Experimental Results

The experiments were performed with the still image database of the Laboratory for Image and Sound Processing of the Technical University of Sofia, Bulgaria. For the experiments were used more than 800 grayscale and color still images of fingerprints, signatures, texts and biomedical signals (ECGs, EEGs, etc.) of various sizes,(.bmp files). The results were compared with those obtained for JPEG2000LS-based software. The test software implementation of the method is in C++, Windows environment. On Fig.1 are shown example test images from the image database. The test image "Phong" is color (24 bpp, RGB) and the remaining images are grayscale (8 bpp).

In Table 1 are shown the results obtained for test images from Fig.1. Similar results were obtained for the remaining images from the corresponding image classes. Best results were obtained for images with many free brightness levels in their histograms. For illustration only, the histogram of the image ECG is shown on Fig. 2.

ECG   Signature

Phong   Fingerprint   Text
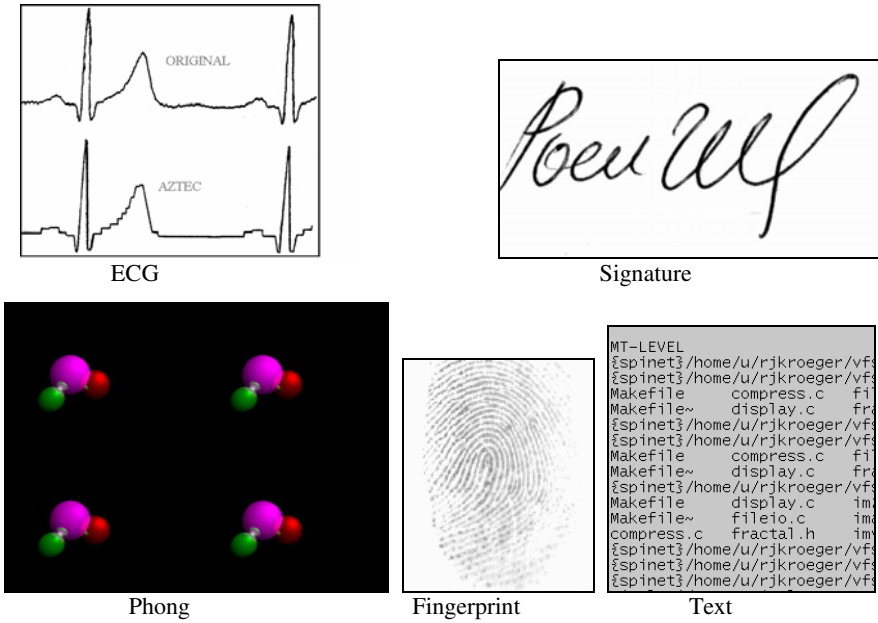
Fig. 1. Test images of various kind: ECG, signature, graphic, fingerprint, text.

Table 1. Results obtained for the losslessly compressed test images

Image	Size	CR_{ARLC}	$CR_{JP2000LS}$
ECG	512×384 pixels, 8bpp	22,80	7,0
Signature	596×312 pixels, 8bpp	23,35	6,2
Fingerprint	128×156 pixels, 8bpp	5,02	4,9
Phong	800×600 pixels, 24bpp	32,97	23,7
Text	256×256 pixels, 8bpp	16,95	1,9
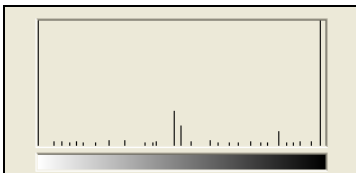
Fig. 2. Brightness histogram of the image "ECG"
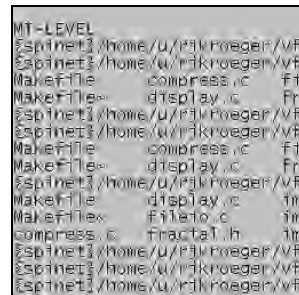
Fig. 3. Image "Text" compressed 16 times with JPEG2000.

Extremely good results were obtained for images, which contain relatively large areas of same or continuously changing brightness or color. Example test image and its histogram are shown on Fig. 4.a,b. The compression ratio, obtained for this image with the ARL method was 89.5, while for JPEG2000LS it was 22. The compression ratio obtained for the color test image "Green" of size 512x512 pixels, 24 bpp (uniform green color), with the ARL method was 52428 (the compressed file was of size 11 bytes only), while the compression ratio for the JPEG2000LS was 84 (the compressed file was of size 3120 B).

a. Test image "Slope"   b. Brightness histogram of the image "Slope"
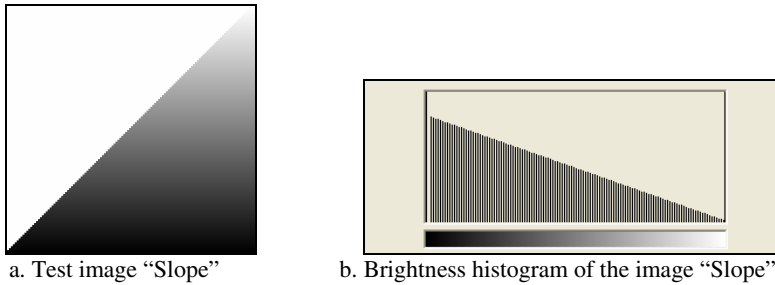
Fig. 4. Test image with regular brightness change

5 Conclusions

Compared to other contemporary methods for lossless still image compression, the ARL coding/decoding has relatively low computational complexity. For example, JPEG 2000LS is based on the wavelet transform (which has very high computational complexity) and arithmetic coding. Specific for the method, presented above, is that it comprises operations "sorting" and "adding", while the operation "multiplication" or some kind of complicated transforms are not used. The method ensures very high compression ratios for cases, when the original data contains long sequences of same values or regularly changing values. These basic characteristics distinguish the new method from the famous methods for RL data coding. The experimental results confirmed the theoretical conclusions. Additional research in the area of documents archiving and processing of biomedical signals proved the wide application area of the method [16, 17]. The ARL method could be further improved, combining it with methods for image pre-processing (adaptive histogram modification), and methods for lossless compression (Huffman or arithmetic coding, etc.). In accordance with the method were developed a special format.

On the basis of the method evaluation follows that the main applications of the ARL coding method and its future development are in the areas:

• Development of new Internet-based systems for e-trade, distance learning, computer games, and many others;

• Creation of algorithm for image compression, which to be integrated in the application software for scanners and faxes, because the method is equally efficient for different alphabets, fonts, etc.

• Development of new application software for mobile phones, medical equipment, biomedical signals storage and transfer, smart cards, documents archiving, etc.

References

1. Ziv, J., Lempel, A.: Compression of individual sequences via variable rate coding. IEEE Trans. on Information Theory IT-24(5), 530–535 (1978)
2. Burrows, M., Wheeler, D.: A block-sorting Lossless Data Compression Algorithm. Digital Systems Research Center. SRC Report 124 (1994)
3. Fano, R.: Transmission of Information. MIT Press, Cambridge (1949)
4. Huffman, D.: A Method for the Construction of Minimum-Redundancy Codes. Proceedings of the IRE 40(9), 1098–1101 (1952)
5. Golomb, W.S.: Run-Length Encodings. IEEE Trans. on Information Theory IT-12(3), 399–401 (1966)
6. Witten, I., Neal, R., Cleary, J.: Arithmetic Coding for Data Compression. Communications of the ACM 30(6) (1987)
7. Raita, T., Teuhola, J.: Predictive text compression by hashing. In: ACM Conf. on Information Retrieval (1987)
8. Karam, L.: Lossless Image Compression. In: Al Bovik (ed.) The Essential Guide to Image Processing, ch. 16, pp. 385–420. Academic Press, London (2009)
9. Salomon, D.: Data compression. Springer, New York (2004)
10. Bell, T., Witten, I., Cleary, J.: Text Compression. Prentice-Hall, Englewood Cliffs (1990)
11. Cappellini, V.: Data Compression and Error Control Techniques with Applications. Academic Press, New York (1985)
12. Sayood, K.: Lossless Compression Handbook. Academic Press, London (2002)
13. Storer, A., Helfgott, H.: Lossless Image Compression by Block Matching. The Computer Journal 40(2/3), 137–145 (1997)
14. Taubman, D., Marcellin, M.: JPEG 2000 Image Compression Fundamentals, Standards and Practice. Kluwer Academic, Norwell (2002)
15. Weinberger, M., Seroussi, G., Sapiro, G.: LOCO-I: A Low Complexity, Context-Based, Lossless Image Compression Algorithm. In: Storer, J. (ed.) Proceedings of Data Compression Conference, pp. 140–149. IEEE Computer Society Press, Los Alamitos (1996)
16. Milanova, M., Kountchev, R.: New Method for Lossless Compression of Medical Records. In: 8th IEEE Intern. Symp. on Signal Processing and In-formation Technology (ISSPIT 2008), Bosnia and Herzegovina, pp. 23–28 (2008)
17. Kountchev, R., Todorov, V., Kountcheva, R.: Efficient archiving of documents with protection of their authentic contents. Int. J. of Reasoning-based Intelligent Systems (IJRIS) 1(1/2), 43–55 (2009)

Scene Categorization with Class Extendibility and Effective Discriminative Ability

Zongyu Lan[1,2], Songzhi Su[1,2], Shu-Yuan Chen[3,*], and Shaozi Li[1,2,*]

[1] Cognitive Science Department, Xiamen University, Xiamen, China
[2] Fujian Key Laboratory of the Brain-like Intelligent Systems, Xiamen University, China
`szlig@xmu.edu.cn`
[3] Department of Computer Science and Engineering, Yuan Ze University, Taiwan
Tel.: 886-3-463-8800, Ext. 3005; Fax: 886-3-463-8850
`cschen@saturn.yzu.edu.tw and`

Abstract. Most of the numerous studies of scene categorization assume a fixed number of classes, and none categorize images with efficient class extendibility while preserving discriminative ability. This capability is crucial for an effective image categorization system. The proposed scene categorization method provides category-specific visual-word construction and image representation. The proposed method is effective for several reasons. First, since the visual-word construction and image representation are category-specific, image features related to the original classes need not be recreated when new classes are added, which minimizes reconstruction overhead. Second, since the visual-word construction and image representation are category-specific, the corresponding learning model for classification has substantial discriminating power. Experimental results confirm that the accuracy of the proposed method is superior to existing methods when using single-type and single-scale features.

Keywords: Scene categorization, classification, category-specific, class extendibility, image retrieval, visual words, codebook.

1 Introduction

The problem of scene categorization is to categorize a scene image where each image in a given set is considered one category. Among the many approaches to scene categorization, Bag-of-Visual-Words approach [1-16] has achieved great success in recent years. In this approach, each image is represented by a non-ordered collection of local features, which are then quantized into a set of visual words to form a codebook. Each image is then represented by the distribution of visual words.

Quelhas *et al.* [1] proposed Bags-of-Visterms to represent invariant local features and Probabilistic Latent Semantic Analysis (pLSA) to model scenes. Fei-Fei and Perona [2] represented a scene image by a collection of local regions, denoted as

---

* Coressponding authors.

codewords obtained by unsupervised learning. They also modified the Latent Dirichlet Allocation to model scenes. Lazebnik *et al.* [3] proposed a spatial pyramid matching based on multi-spatial resolutions for recognizing natural scene categories. Zhou *et al.* [4] used Gaussianized vector representation rather than the distribution of visual words to represent scene images for categorization. A comparative study was performed in Ref. [5].

Some researchers [6-9] have considered contextual information to improve categorization accuracy. A detailed survey can be found in Ref. [6]. Some researchers [10-12] have focused on compact codebook construction. Some researchers [13-16] have proposed novel features or learning models to improve categorization accuracy.

However, most approaches assume a fixed number of classes, and none categorize images with efficient class extendibility while preserving discriminative ability. This capability is crucial for an effective image categorization system. The proposed scene categorization method provides category-specific visual-word construction and image representation. The proposed method is effective for several reasons. First, since the visual-word construction and image representation are category-specific, image features related to the original classes need not be recreated when new classes are added, which minimizes reconstruction overhead. Second, since the visual-word construction and image representation are category-specific, the corresponding learning model for classification has substantial discriminating power. Experimental results confirm that the accuracy of the proposed method is superior to existing methods when using single-type and single-scale features.

Section 2 introduces three strategies for codebook construction and image representation, including the proposed method. Section 3 gives the experimental results. Section 4 concludes the study and proposes future works.

2   Category-Specific Approach

The scene categorization problem can be formulated as follows. Given an image I, a set of C scene categories $\{1,2,\cdots,C\}$, and a training set of labeled images $T = \{I_s^c | s = 1,\cdots,S, c = 1,\cdots,C\}$, identify the category of image I. Here, S is the number of labeled images for each category, *i.e.,* the number in the training set T is $C \times S$. To solve the problem, this section describes three possible strategies. The former two are existing methods, and the latter is the novel proposed method for achieving class extendibility.

2.1   Whole-Construction/Whole-Representation Strategy

The conventional Bag-of-Visual-Words approach adopts the whole-construction/whole-representation (W-C/W-R) strategy for solving the problem. In the training phase of this strategy, each training image I_s^c is divided into N local image patches of fixed size P_n. The training set contains $C \times S \times N$ patches. For each patch, local features $\mathbf{x}_n$ of dimension d, such as SIFT of dimension 128 [17], are extracted. Accordingly, the $C \times S \times N$ features are then fed into an unsupervised learning model such as K-Means clustering [18] to construct a size M

codebook, which is denoted $\left\{\mathbf{w}_k \middle| k = 1, \cdots, M\right\}$ based on the similarity measure. Suppose that training image I_s^c has N patch features, $\mathbf{x}_s^c(n), c = 1, \cdots, 15, s = 1, \cdots, S, \ n = 1, \cdots, N$. Each patch with features $\mathbf{x}_s^c(n)$ is then labeled as a codeword $\mathbf{w}_{L\left(\mathbf{x}_s^c(n)\right)}$ from the codebook $\left\{\mathbf{w}_k \middle| k = 1, \cdots, M\right\}$ according to the Euclidean distance $\left\| \mathbf{x}_s^c(n) - \mathbf{w}_k \right\|$.

For all patches of the image I_s^c, the codeword distribution (histogram $\mathbf{h}_s^c$) can then be calculated by the following equation to represent I_s^c.

$$\mathbf{h}_s^c = \left\langle h_s^c(1), \cdots, h_s^c(M) \right\rangle$$

$$h_s^c(k) = \frac{\left| \left\{ \mathbf{x}_s^c(n) \middle| L\left(\mathbf{x}_s^c(n)\right) = k, n = 1, \cdots, N \right\} \right|}{N}, k = 1, \cdots, M \tag{1}$$

where $\left| - \right|$ denotes cardinality function. The $C \times S$ histograms $\mathbf{h}_s^c$ can then be fed into a learning model such as a support vector machine (SVM) to obtain a C-to-1 classifier. Note that the dimension of the SVM is M. In the testing phase, the unknown image I is also divided into N local image patches of fixed size. Similarly, each patch of I can be labeled as a codeword, and I can be represented by a histogram of codewords $\mathbf{h}$, which can then be fed into the learned C-to-1 classifier to determine its category.

However, the proposed approach to representing a scene image is vector-based [4] rather than histogram-based. Specifically, a scene image is represented by the feature vector of the codeword rather than by the histogram of the codeword count. Restated, each training image I_s^c with patch features $\mathbf{x}_s^c(n)$ is represented by the mean feature vector $\mathbf{F}_s^c = \left\langle \mathbf{f}_s^c(1), \cdots, \mathbf{f}_s^c(M) \right\rangle$ with $\mathbf{f}_s^c(k)$ as defined by the following equation.

$$\mathbf{f}_s^c(k) = \frac{\sum\limits_{L\left(\mathbf{x}_s^c(n)\right) = k, n = 1, \cdots, N} \mathbf{x}_s^c(n)}{\left| \left\{ \mathbf{x}_s^c(n) \middle| L\left(\mathbf{x}_s^c(n)\right) = k, n = 1, \cdots, N \right\} \right|}, c = 1, \cdots, C, s = 1, \cdots, S, k = 1, \cdots, M \tag{2}$$

Note that the dimensions of the C-to-1 SVM classifiers are M and $d \times M$ when using the histogram-based approach and the vector-based approach, respectively. The experiments in Ref. [4] showed that the latter approach generally achieves a higher dimension and thus a higher categorization accuracy rate. This study therefore applied the vector-based approach.

2.2 Category-Specific-Construction/Whole-Representation Strategy

A category-specific-construction/whole-representation (C-C/W-R) strategy (Figure 1) was recently proposed in Ref. [9]. Experimental results in that study confirmed that this strategy is more accurate compared to the conventional W-C/W-R strategy [9, 10]. In the training phase of this approach, each training image I_s^c is divided into N local image patches of fixed size with local features $\mathbf{x}_s^c(n)$. However, the codebook is constructed individually for each category c. Restated, only the $S \times N$ features of images belonging to a specific category i are fed into K-Means clustering to construct a size M codebook $\left\{\mathbf{w}_k^i \middle| k = 1, \cdots, M\right\}$. The C codebooks, one for each category, are then combined into one codebook of size $C \times M$, denoted as $\left\{\mathbf{w}_k^i \middle| i = 1, \cdots, C, k = 1, \cdots, M\right\}$. All subsequent processes in the training phase and in the testing phase are identical to those in the
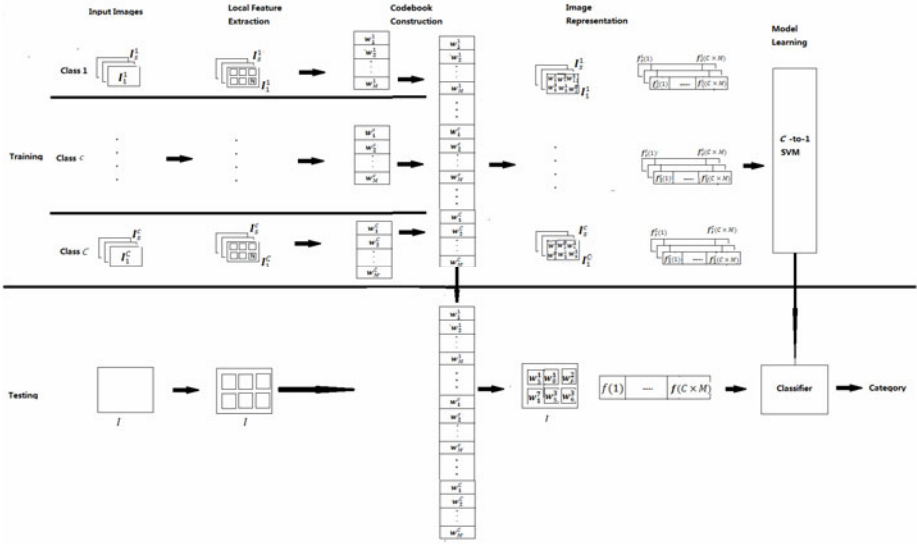
Fig. 1. Category-specific-construction/whole-representation strategy

conventional   W-C/W-R   approach   except   that   the   codebook   is $\left\{\mathbf{w}_k^i \middle| i = 1,\cdots,C, k = 1,\cdots,M\right\}$ rather than $\left\{\mathbf{w}_k \middle| k = 1,\cdots,M\right\}$.

Specifically, each patch with features $\mathbf{x}_s^c(n)$ is then labeled a codeword $\mathbf{w}_k^i {}_{L\left(\mathbf{x}_s^c(n)\right)(k)}^{L\left(\mathbf{x}_s^c(n)\right)(i)}$ from the codebook $\left\{\mathbf{w}_k^i \middle| i = 1,\cdots,C, k = 1,\cdots,M\right\}$ according to the Euclidean distance $\left\|\mathbf{x}_s^c(n) - \mathbf{w}_k^i\right\|$. Each training image I_s^c with patch features $\mathbf{x}_s^c(n)$ is represented by the feature vector $\mathbf{F}_s^c = \left\langle \mathbf{f}_s^c(1),\cdots,\mathbf{f}_s^c(M),\mathbf{f}_s^c(M+1),\cdots,\mathbf{f}_s^c(C \times M)\right\rangle$ with $\mathbf{f}_s^c(j)$ as defined by the following equation.

$$\mathbf{f}_s^c(j) = \frac{\sum\limits_{L\left(\mathbf{x}_s^c(n)\right)(i)=\mathrm{mod}(j),L\left(\mathbf{x}_s^c(n)\right)(k)=rem(j),n=1,\cdots,N} \mathbf{x}_s^c(n)}{\left|\left\{\mathbf{x}_s^c(n)\middle| L\left(\mathbf{x}_s^c(n)\right)(i) = \mathrm{mod}(j), L\left(\mathbf{x}_s^c(n)\right)(k) = rem(j), n = 1,\cdots,N\right\}\right|}, \qquad (3)$$
$$c = 1,\cdots,C, s = 1,\cdots,S, j = 1,\cdots,C \times M$$

where $\mathrm{mod}(j)$ and $rem(j)$ are the quotient and remainder, respectively, after dividing j by M. The above two strategies use the same codebook to represent whole scene images regardless of category. However, the two approaches have different codebook sizes. In the histogram-based approach, each scene image is represented by histograms with dimensions M and $C \times M$ for strategies W-C/W-R and C-C/W-R, respectively, whereas each scene in the vector-based approach is represented by vectors of dimensions $d \times M$ and $d \times C \times M$ for strategies W-C/W-R and C-C/W-R, respectively.

2.3 Category-Specific-Construction/ Category-Specific-Representation Strategy

Nevertheless, the above two strategies assume a fixed number of classes and are not concerned with categorizing images with efficient class extendibility while preserving discriminative ability. However, this issue is crucial for an effective scene categorization system. The category-specific-construction /category-specific-representation (C-C/C-R) strategy proposed in this study obtains a scene categorization with efficient class extendibility and effective discriminative ability. Figure 2 is a flowchart of the detailed steps of the proposed method.

Codebook construction in the proposed method is the same as that in C-C/W-R strategy. Although C codebooks $\left\{\mathbf{w}_k^i \middle| k = 1, \cdots, M\right\}, i = 1, \cdots, C$, are also obtained, they are not combined into a single codebook. Instead, each category-specific codebook is used to represent scene images individually. That is, each patch with features $\mathbf{x}_s^c(n)$ is then labeled for each category i as a codeword $\mathbf{w}_{L^i(\mathbf{x}_s^c(n))}^i$ from the codebook $\left\{\mathbf{w}_k^i \middle| k = 1, \cdots, M\right\}\left\{\mathbf{w}_k^i \middle| k = 1, \cdots, M\right\}$ according to the Euclidean distance $\left\|\mathbf{x}_s^c(n) - \mathbf{w}_k^i\right\|$. Each training image I_s^c with patch features $\mathbf{x}_s^c(n)$ is then represented by the mean feature vector $\mathbf{F}_s^c = \left\langle\mathbf{f}_s^c(1,1), \cdots, \mathbf{f}_s^c(1, M), \cdots, \mathbf{f}_s^c(C,1), \cdots, \mathbf{f}_s^c(C, M)\right\rangle$ with $\mathbf{f}_s^c(i,k)$ as defined by the following equation.

$$\mathbf{f}_s^c(i,k) = \frac{\sum_{L^i(\mathbf{x}_s^c(n)) = k, n = 1, \cdots, N} \mathbf{x}_s^c(n)}{\left\|\left\{\mathbf{x}_s^c(n) \middle| L^i(\mathbf{x}_s^c(n)) = k, n = 1, \cdots, N\right\}\right\|}, c = 1, \cdots, C, s = 1, \cdots, S, i = 1, \cdots, C, k = 1, \cdots, M \quad (4)$$

Notably, each scene image is represented by the same dimension used in the C-C/W-R strategy, regardless of whether the approach is histogram-based or vector-based, *i.e.*, $C \times M$ and $d \times C \times M$, respectively.

The proposed method clearly achieves the goal of efficient class extendibility and effective discriminative ability. First, since the visual-word dictionary and image representation are category-specific, adding new classes does not require reconstruction of all image features related to those in the original class so as to minimize reconstruction overhead. Second, since the visual-word construction and image representation are category-specific, the corresponding learning model for classification should have sufficient discriminatory power.

Notably, Expectation-Maximization (EM) in Gaussian mixtures [18] rather than K-means were used to construct codebooks and to represent images in Ref. [4], which improved performance [4, 18]. Thus, EM is also used in this study, and Eq. (4) should be updated as follows. Each visual word of the codebook $\left\{\mathbf{w}_k^i \middle| k = 1, \cdots, M\right\}, i = 1, \cdots, C$, is first updated as one of M uni-model Gaussian components by the following equation.

$$\mathbf{w}_k^i = N\left(\mathbf{x}; \boldsymbol{\mu}_k^i, \boldsymbol{\Sigma}_k^i\right), i = 1, \cdots, C, k = 1, \cdots, M \quad (5)$$

where $\boldsymbol{\mu}_k^i$ and $\boldsymbol{\Sigma}_k^i$ denote the mean vector and covariance matrix of the kth Gaussian component and where $\mathbf{x}$ denotes the local patch feature. Each patch with features $\mathbf{x}_s^c(n)$ is then labeled for each category i as a codeword $\mathbf{w}_{L^i(\mathbf{x}_s^c(n))}^i$ from the codebook $\left\{\mathbf{w}_k^i \middle| k = 1, \cdots, M\right\}$ according to the probability $N\left(\mathbf{x}_s^c(n); \boldsymbol{\mu}_k^i, \boldsymbol{\Sigma}_k^i\right), i = 1, \cdots, C, k = 1, \cdots, M, n = 1, \cdots, N$. Each training image I_s^c with patch features $\mathbf{x}_s^c(n)$ is

represented by the feature vector $\mathbf{F}_s^c = \left\langle \mathbf{f}_s^c(1,1),\cdots,\mathbf{f}_s^c(1,M),\cdots,\mathbf{f}_s^c(C,1),\cdots,\mathbf{f}_s^c(C,M)\right\rangle$ with $\mathbf{f}_s^c(i,k)$ as defined by the following equation.

$$\mathbf{f}_s^c(i,k)=\left(\frac{\displaystyle\sum_{L^i\left(\mathbf{x}_s^c(n)\right)=k,n=1,\cdots,N}\mathbf{x}_s^c(n)}{\left|\left\{\mathbf{x}_s^c(n)\middle|L^i\left(\mathbf{x}_s^c(n)\right)=k,n=1,\cdots,N\right\}\right|}-\boldsymbol{\mu}_k^i\right)\middle/\sqrt{\left|\boldsymbol{\Sigma}_k^i\right|} \tag{6}$$

$$c=1,\cdots,C, s=1,\cdots,S, i=1,\cdots,C, k=1,\cdots,M$$

Similarly, for the C-C/W-R strategy, Eq. (3) can be updated accordingly if EM in Gaussian mixtures is used instead of K-means to construct codebooks and to represent images as in Ref. [4].
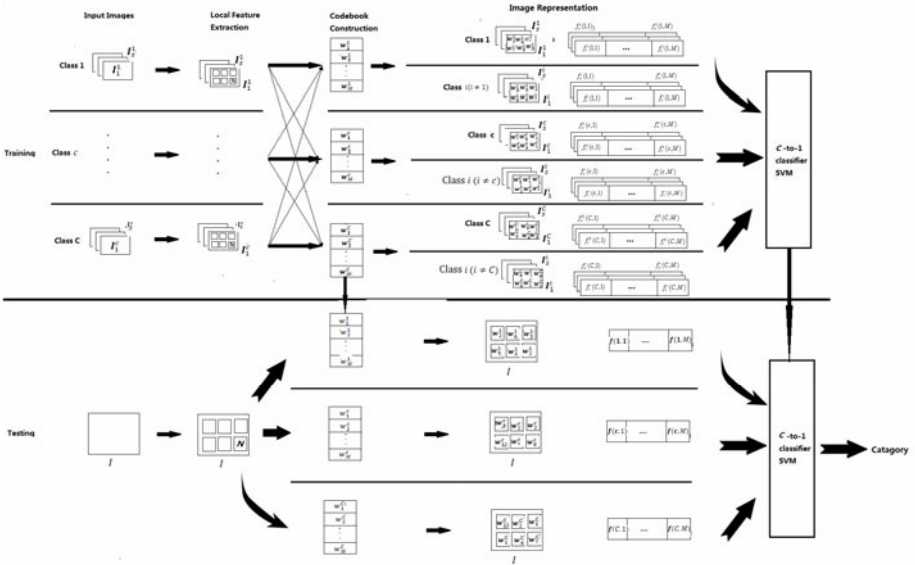
Fig. 2. Category-specific-construction/category-specific-representation strategy

3   Experimental Results

Performance evaluations are performed using two data sets. One is Scene-13 dataset [2], and the other is Scene-15 dataset [3]. Each category consists of 200 to 400 images with average size of 384×256. The SVM classifier adopted in this study is the LIBSVM implemented by Chang and Lin with radial-basis function kernel [19].

Each scene image is tiled with dense overlapping patches of fixed size 16×16 and 8 overlapping pixels. Thus, there are 47×31=1457 patches for each image. The sizes of codebook M are 35, 50 and 80. Two local patch features are adopted: Histogram of Oriented Gradient (HOG) [20] and Scale-Invariant Feature Transform (SIFT) [17]

which have dimensions of 16 and 128, respectively. Principal component analysis (PCA) [18] is applied to SIFT to reduce the SIFT dimension from 128 to 64. The HOG dimension is reduced from the 36 bins in the original version [20] to 16 bins. The histogram channels from 0 to 360 degrees are divided into 16 sectors, each of which is 22.5 degrees. Thus, the HOG in this study has 16 bins.

Experiments are performed to test the performance of the Scene-13 and Scene-15 datasets and the class extendibility of the Scene-15 dataset. For each issue, experiments are repeated ten times with 100 randomly selected images per category for training and the rest for testing.

Tables 1 and 2 show the categorization results for the Scene-13 and Scene-15 datasets for different features and codebook sizes and for different strategies, respectively. The tables show that the proposed method is more accurate compared to the C-C/W-R strategy. Moreover, for the Scene-13 dataset, the highest accuracies obtained in Ref. [4] are 83.6% with 512 visual words and 84.1% with 1024 visual words; in the current study, the accuracies are 85.42% with 13×35=455 visual words and 86.65% with 13×80=1040 visual words. For the Scene-15 dataset, the highest accuracy in Ref. [4] is 83.5% with 1024 visual words; that in the current study is 84.06% with 15×80=1200 visual words. However, Ref [4] adopted SIFT feature for the Scene-13 dataset but SIFT and coordinate information for the Scene-15 dataset. The proposed method adopted only SIFT feature for the Scene-13 and Scene-15 datasets consistently. On the other hand, although Ref. [9] adopted C-C/W-R strategy, it obtains 87.63% and 85.16% accuracies for the Scene-13 and Scene-15 datasets, respectively. However, Ref. [9] adopted multi-scale features rather than single-scale feature as in the proposed method. Thus, the proposed method has higher accuracy compared to all existing methods when single-type and single-scale features are used.

Table 1. Categorization results for Scene-13 dataset.

Features	HOG			SIFT		
M	35	50	80	35	50	80
C-C/W-R	79.88±0.12	80.31±0.12	81.45±0.06	83.44±0.08	85.25±0.12	85.47±0.14
C-C/C-R	83.25±0.06	83.58±0.06	**84.03±0.09**	85.42±0.04	86.35±0.03	**86.65±0.48**

Table 2. Categorization results for Scene-15 dataset.

Features	HOG			SIFT		
M	35	50	80	35	50	80
C-C/W-R	78.49±0.11	79.45±0.19	79.81±0.13	79.45±0.16	79.95±0.14	80.21±0.06
C-C/C-R	81.45±0.08	81.99±0.09	**82.63±0.10**	83.02±0.09	83.42±0.08	**84.06±0.07**

Class extendibility is verified by simulating an increase in the number of classes from the original 2 and 5 categories to 15 categories. The experiments were conducted only on Scene-15 dataset. The assumptions are 2 and 5 original categories and codebook construction based only on 2 and 5 categories. Suppose that the number of categories is increased to 15. In this case, if image representations of the scene images in the original 2 and 5 categories are not reconstructed, the codebook categories are still $M \times 2$ and $M \times 5$, respectively, rather than $M \times 15$ in the original C-C/W-R strategy. Categorization accuracy is clearly degraded (Table 3). However, the proposed strategy can achieve efficient class extendibility easily since image representation is also category-specific.

Table 3. Codebook construction based on different numbers of categories of Scene-15 datasets when applying C-C/W-R strategy.

Features	HOG			SIFT		
M	35	50	80	35	50	80
15	78.49±0.11	79.45±0.19	79.81±0.13	79.45±0.16	79.95±0.14	80.21±0.06
5	76.43±1.24	76.94±1.01	77.45±1.21	77.12±0.86	77.67±0.90	77.86±0.67
2	71.45±1.42	72.06±1.72	72.34±1.85	72.47±1.20	72.84±1.31	73.50±0.92

4   Conclusions

This study developed a scene categorization method with category-specific visual-word construction and image representation to enable scene categorization with efficient class extendibility and effective discriminative ability. Experimental results confirm that the proposed strategy has higher accuracy compared to existing methods when single-type and single-scale features are used. Future work can be directed to subtle codebook construction, contextual information incorporation, and hierarchical scene categorization.

Acknowledgment

References

[1] Quelhas, P., Monay, F., Odobez, J.M., Gatica-Perez, D., Tuytelaars, T., Van Gool, L.: Modeling scenes with local descriptors and latent aspects. In: Proc. IEEE Int. Conf. Computer Vision, pp. 883–890 (2005)

[2] Li, F.-F., Perona, P.: A Bayesian hierarchical model for learning natural scene categories. In: Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition, pp. 524–531 (2005)

[3] Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: Proc. IEEE Int. Conf. Computer Vision, pp. 2169–2178 (2006)

[4] Zhou, X., Zhuang, X., Tang, H., Hasegawa-Johnson, M., Huang, T.S.: Novel Gaussianized vector representation for improved natural scene categorization. Pattern Recognition Letters 31(8), 702–708 (2010)

[5] Bosch, A., Munoz, X., Marti, R.: Which is the best way to organize/classify images by content? Image Vision Computing 25(6), 778–791 (2007)

[6] Galleguillos, C., Belongie, S.: Context-based object categorization: A critical survey. Computer Vision and Image Understanding 114(1), 712–722 (2010)

[7] Hoang, N.V., Gouet-Brunet, V., Rukoz, M., Manouvrier, M.: Embedding spatial information into image content description for scene retrieval. Pattern Recognition 43(9), 3013–3024 (2008)

[8] Lu, Z., Ip, H.H.S.: Combining context, consistency, and diversity cues for interactive image categorization. IEEE Trans. Multimedia 12(3), 194–203 (2010)

[9] Qin, J., Yung, N.H.C.: Scene categorization via contextual visual words. Pattern Recognition 43(5), 1874–1888 (2010)

[10] Van Gemert, J.C., Snoek, C.G.M., Veenman, C.J., Smeulders, A.W.M., Geusebroek, J.-M.: Comparing compact codebooks for visual categorization. Computer Vision and Image Understanding 114(4), 450–462 (2010)

[11] Van Gemert, J.C., Veenman, C.J., Smeulders, A.W.M., Geusebroek, J.-M.: Visual word ambiguity. IEEE Trans. Pattern Recognition and Machine Intelligence 32(7), 1271–71283 (2010)

[12] Wu, L., Hoi, S.C.H., Yu, N.: Semantics preserving bag-of-words models and applications. IEEE Trans. Image Processing 19(7), 1908–1920 (2010)

[13] Sun, Z.-L., Rajan, D., Chia, L.-T.: Scene Classification using multiple features in a two-stage probabilistic classification framework. Neurocomputing 73(16-18), 2971–2979 (2010)

[14] Abdullah, A., Veltkamp, R.C., Wiering, M.A.: Fixed partitioning salient points with MPEG-7 cluster correlograms for image categorization. Pattern Recognition 43(3), 650–662 (2010)

[15] Bosch, A., Zisserman, A., Munoz, X.: Scene classification using a hybrid generative/discriminative approach. IEEE Trans. Pattern Recognition and Machine Intelligence 30(4), 712–727 (2008)

[16] Cheng, H., Wang, R.: Semantic modeling of natural scenes based on contextual Bayesian networks. Pattern Recognition 43(12), 4042–4054 (2010)

[17] Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Computer Vision 60(2), 91–110 (2004)

[18] Alpaydin, E.: Introduction to Machine Learning. MIT Press, Cambridge (2010)

[19] Chang, C.-C., Lin, C.-J.: LIBSVM: A binary for support vector machine. Software available at, http://www.csie.ntu.edu.tw/cjlin/libsvm

[20] Dalal, N., Triggs, B.: Histogram of oriented gradient for human detection. In: Proc. Computer Vision and Pattern Recognition, pp. 886–893 (2005)

Adaptive Navigation in a Web Educational System Using Fuzzy Techniques

Chrysafiadi Konstantina and Virvou Maria

Department of Informatics, University of Piraeus
kchrysafiadi@yahoo.gr, mvirvou@unipi.gr

Abstract. In this paper we develop a technique for creating web-educational systems which offer adaptive navigation support. Our technique is based on fuzzy logic. We represent the dependencies between the domain concepts using fuzzy cognitive maps. Also, we choose fuzzy sets to describe student's knowledge level and cognitive abilities and we use a mechanism of rules over the fuzzy sets, which is triggered after any change of the student's knowledge level of a domain concept, and update the student's knowledge level of all related with this concept, concepts.

1   Introduction

The rapid development of information and communication technologies have changed the ways of teaching and learning giving rise to the field of web-based education. Web-based educational systems offer easy access from everywhere for anyone and whenever to knowledge domains and learning processes. Due to the very large audience, which consists of learners with different characteristics and needs, that use an e-learning system, and the absence of teacher, high adaptivity is required. Adaptive web applications have the ability to deal with different users' needs for enhancing usability and comprehension and for dealing with large repositories [1].

Adaptation is the key to reach the goal of offering each individual student the knowledge they need. Adaptive Hypermedia is a research domain that helps to achieve adaptivity in web-based educational systems [2]. An adaptation technology of adaptive hypermedia is the adaptive navigation support technology that helps students to find an "optimal path" through the learning material. Adaptive navigation support is a specific group of technologies that support user navigation in hyperspace, by adapting to the goals, preferences and knowledge of the individual user [3]. Thus, in order to apply adaptive navigation support to a web-based educational system we have to model the knowledge, to diagnose the needs, misconceptions and cognitive abilities of each individual student.

Student diagnosis is fraught with uncertainty,. One possible approach to deal with this is fuzzy student modelling. Fuzzy sets theory was introduced by Zadeh [4], according to who, the main contribution of Fuzzy logic is a methodology for computing with words, which cannot be done equally well with other methods [5]. Thus, Fuzzy logic techniques can be used to improve the performance of a web-based educational application. Several researchers found fuzzy student modelling as adequate for carrying out the system's assessment and pedagogical functions [6, 7, 8, 9, 10, 11, 12].

In this paper we describe a fuzzy student modelling. A description of the domain knowledge using fuzzy cognitive maps is given first. The description of a student modelling using fuzzy techniques follows. The student model is mainly based on the student cognitive model. The data from the fuzzy user model is the basis for the system adaptation.

2   Adaptive Navigation Support

A web-based educational system is used by students with different needs and cognitive abilities. Therefore, it is not effective all the learners follow the same instruction model. That is, all learners should not read the same material and in the same order. Some learners do not need to read some chapters because these are known to them, some others have to revise some domain concepts, and others should give more attention to some knowledge issues.

Adaptive hypermedia is the solution for offering adaptive navigation. Adaptive navigation support technology support the student in hyperspace orientation and navigation by adaptively sorting, annotating or partly hide the links that constitute the domain knowledge material, to make easier the choice of the next link to proceed [13]. ISIS-Tutor [14], ELM-ART[15], InterBook [16], WEST-KBNS [17] and CALL [18] are some educational system which apply adaptive navigation support. However, which are the criteria, that are the base for adaptive navigation support? Mainly, the adaptive navigation support is applied regarding each individual learner's knowledge and cognitive needs, as well as the domain concepts of the domain knowledge and how they are related with each other. In other words, adaptive navigation support is achieved obtaining information about the domain knowledge and about the learner for her/his student model.

3   The Domain Knowledge

One of the most important components of an adaptive educational application is the representation of knowledge. To enable communication between system and learner at content level, the domain model of the system has to be adequate with respect to inferences and relations of domain entities with the mental domain of a human expert [19]. Taking this into account, the domain knowledge of a web application can be organized into an hierarchical tree, with domain topics as intermediate nodes and learning objectives as leaf nodes [20] as it is depicted in figure 1. The full collection of domain topics constitutes the whole domain knowledge. Each topic has its learning objectives that are classified according to Bloom's taxonomy [21]. Learning objectives determine the concepts that must be understood and learned in each chapter. The hierarchy of the tree depicts the difficulty level of the domain topics and the order in which each topic has to be taught.
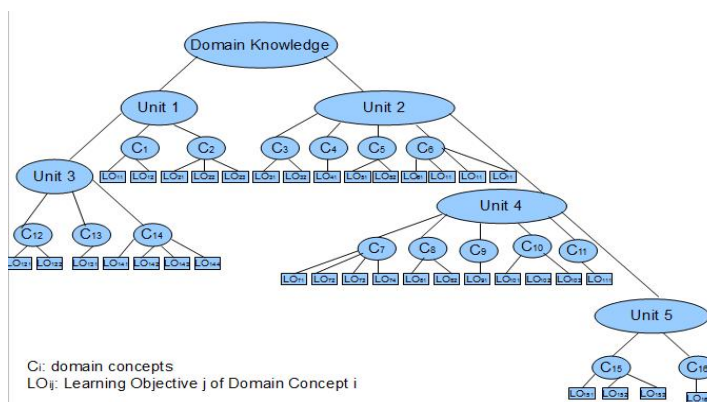
Fig. 1. The tree og the Domain Knowledge

Furthermore, there are some dependencies among the domain concepts. Lets take two depended concepts C_i and C_j. Concept C_i can be a prerequisite of concept C_j, or it can follow concept C_j, or the two concepts can belong to the same level of hierarchy. In all these cases the knowledge of the one concept affects the knowledge of the other domain concept. We have to make explicit that the knowledge of a domain concept does not only affect the knowledge of domain concepts that follow in the hierarchy model of the knowledge domain, but also it affects its prerequisite concepts, as well as the knowledge of the concepts of the same hierarchy level. That happens because the order in which the concepts of the same hierarchy level are read, is not determined but it is selected by the learner. Moreover, the knowledge of a concept affects the knowledge of a prerequisite domain concept, since their knowledge is related.

The learning of an issue helps to better understanding a depended chapter in some degree. Also, the knowledge of a domain concept could not only increase, but also it can decrease in some degree, when the knowledge of a depended topic is not satisfactory. This degree is determined by the experts of the domain knowledge. This kind of dependencies among the concepts of the domain knowledge can be depicted using fuzzy cognitive maps (figure 2). Fuzzy Cognitive Maps constitute a way to represent real-world dynamic systems and they are considered a combination of fuzzy logic and artificial neural networks. They are introduced by Kosko [22,23] and have been applied to many areas [24, 25, 26, 27], with education and student modelling to be among them [28, 29, 30, 31, 32, 33]. A Fuzzy Cognitive Map illustrates the whole system as a combination of concepts and the various causal relations that exist between their concepts [34, 35].The nodes of the map depict the domain concepts, the directed arcs which connect the depended domain concepts represent the causal relationships and the numbers show the degree in which the knowledge level of a domain concept is affected regarding the knowledge of its depended domain concept.
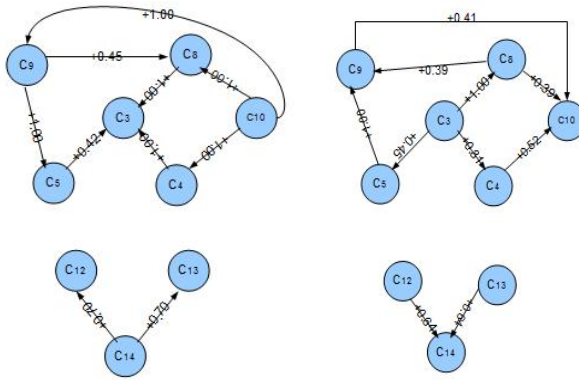
Fig. 2. Fuzzy Cognitive Maps – increasing knowledge

The arcs in a Fuzzy Cognitive Map can take values in the interval [-1, 1]. A positive value indicates a positive causality between concepts Ci and Cj. In other words, the increase on the value of Ci leads to the increase on the value of Cj, or the decrease on the value of Ci leads to the decrease on the value of Cj. On the other hand, a negative value indicates a negative causality between concepts Ci and Cj. That is, the increase on the value of Ci leads to the decrease on the value of Cj, or the decrease on the value of Ci leads to the increase on the value of Cj. In our domain knowledge model the increase on the knowledge of a domain concept Ci leads to the increase on the knowledge of its depended domain concept Cj, or the decrease on the knowledge of a domain concept Ci leads to the decrease on the knowledge of its depended domain concept Cj. Thus, the values of the arcs in our Fuzzy Cognitive Map can be in the interval (0, 1].

4   Student Modeling

Student model is mainly based on the student cognitive model. Therefore, we focus on which parts of the domain knowledge the student knows and how well. Phrases such as "He is very good at domain concept A", "She has a lack of knowledge at domain concept A", "He is moderate at the domain concept B" are vague and imprecise. Moreover, statements such as "She knows at 70% the chapter 1", "He succeeded 85% in the questions about the chapter 3" do not explicitly state that s/he has assimilated the corresponding knowledge issue or s/he has to revise it. Consequently, student's knowledge and cognitive abilities representation is imprecise and one possible approach to deal with this is fuzzy set techniques, with their ability to naturally represent human conceptualization.

We define the following four fuzzy sets for describing student knowledge of a domain concept:

- **Unknown (Un)**: the degree of success in the domain concept is from 0% to 60%.
- **Unsatisfactory Known(UK)**: the degree of success in the domain concept is from 55% to 75%.

- **Known (K)**: the degree of success in the domain concept is from 70% to 90%.
- **Learned (L)**: the degree of success in the domain concept is from 85% to 100%.

The membership functions for the four fuzzy sets are depicted in diagram 1, and are the following:

$$\mu_{Un}(x) = \begin{cases} 1, & x \le 55 \\ 1-(x-55)/5, & 55 < x < 60 \\ 0, & x \ge 60 \end{cases}$$

$$\mu_{K}(x) = \begin{cases} (x-70)/5, & 70 < x < 75 \\ 1, & 75 \le x \le 85 \\ 1-(x-85)/5, & 85 < x < 90 \\ 0, & x \le 70 \text{ or } x \ge 90 \end{cases}$$

$$\mu_{UK}(x) = \begin{cases} (x-55)/5, & 55 < x < 60 \\ 1, & 60 \le x \le 70 \\ 1-(x-70)/5, & 70 < x < 75 \\ 0, & x \le 55 \text{ or } x \ge 75 \end{cases}$$

$$\mu_{L}(x) = \begin{cases} (x-85)/5, & 85 < x < 90 \\ 1, & 90 \le x \le 100 \\ 0, & x \le 85 \end{cases}$$

where x is the student's degree of success in a domain concept.
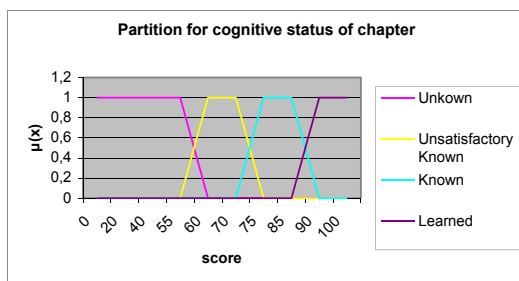
Fig. 3. Partition for cognitive status of chapter

The following expressions stand:

$\mu_{Un}, \mu_{UK}, \mu_{K}, \mu_{L} \in [0,1]$
$\mu_{Un}+\mu_{UK}+\mu_{K}+\mu_{L}=1$
if $\mu_{Un}>0 \Rightarrow \mu_{K}=\mu_{L}=0$
if $\mu_{UK}>0 \Rightarrow \mu_{L}=0$
if $\mu_{K}>0 \Rightarrow \mu_{Un}=0$
if $\mu_{L}>0 \Rightarrow \mu_{Un}=\mu_{UK}=0$

Thus, a quadruplet ($\mu_{Un}, \mu_{UK}, \mu_{K}, \mu_{L}$) is used to express the student knowledge of a domain concept. For example, if a student is succeeding 87% at the domain concept C_1, then her/his knowledge state of this domain concept is described by the quadruplet (0, 0, 0.6, 0.4), which means that the domain concept of C_1 is 60% Known and 40% Learned for the student. To determine the knowledge level of a student, the student takes a test at the end of each chapter.

A characteristic of many domains knowledge is that the experience and the continuous learning process lead to better understanding and assimilation of domain concepts that have been taught a previous time. Furthermore, if a student performs poorly in a chapter, then her/his knowledge level of previous associated chapter has to be reduced. In addition, the learning of some domain concepts helps to better understanding of chapters that follow. In other words, let's symbolize with Ci $\rightarrow$ Cj that there is a dependency between the chapters Ci and Cj, and more concretely that chapter Ci precedes to chapter Cj, then the following four facts can be happened:

f1. Considering the knowledge level of C_i, the knowledge level of C_j increases.
f2. Considering the knowledge level of C_i, the knowledge level of C_j decreases.
f3. Considering the knowledge level of C_j, the knowledge level of C_i increases.
f4. Considering the knowledge level of C_j, the knowledge level of C_i decreases.

When f1 and f3 are happened the student model expands. On the contrary, when f2 and f4 are happened the student model reduces. In other words, after any change of the value of concept knowledge of a domain concept, an inferring mechanism is triggered that updates the values of all related with this concept, concepts.

Let's define D the fuzzy set describing the dependencies between the domain concepts and $\mu_D(Ci, Cj)$ the membership function of the dependency relation of Cj on Ci and $\mu_D(Ci, Cj)$ the membership function of the dependency relation of Ci on Cj. The values of $\mu_D(Ci, Cj)$ $\mu_D(Ci, Cj)$ are the values of the arcs in the Fuzzy Cognitive Maps of the domain knowledge (Figure 1).

Concerning the precedence relation Ci$\rightarrow$Cj the knowledge level of the chapters can change according the following rules:

- Concerning the knowledge level of C_i (C_j) the knowledge level of C_j (C_i) increases according to (S1, S2 are knowledge levels with S1<S2):
  R1:   If C_j (C_i) is S1 and C_i (C_j) is S1, then C_j (C_i) remains S1 with
        $\mu_{S1}(C_j)=\max[\mu_{S1}(C_j),\mu_{S1}(C_i)*\mu_D(C_i,C_j)]$
        $(\mu_{S1}(C_i)=\max[\mu_{S1}(C_i),\mu_{S1}(C_i)*\mu_D(C_j,C_i)])$.
  R2:   If C_j (C_i) is S1 and C_i (C_j) is S2, then C_j (C_i) becomes S2 with
        $\mu_{S2}(C_j)=\mu_{S2}(C_i)*\mu_D(C_i,C_j)$ $(\mu_{S2}(C_i)=\mu_{S2}(C_j)*\mu_D(C_j,C_i))$.
- Concerning the knowledge level of C_i the knowledge level of C_j reduces according to:
  R3:   If C_j is 100% Learned, then it does not change.
  R4:   If C_j is S and C_i is Unknown, then C_j becomes Unknown with
        $\mu_{Un}(C_j)=\mu_{Un}(C_i)*\mu_D(C_i,C_j)$. S is knowledge level with S>Unknown.
  R5:   If C_j is S and C_i is Unsatisfactory Known, then C_j becomes Unsatisfactory Known if $\mu_D(C_i,C_j)=1$ or Known with $\mu_K(C_j)=1-\mu_{UK}(C_j)=1-\mu_{UK}(C_i)*\mu_D(C_i,C_j)$. S is knowledge level with S>Unsatisfacory Known.
  R6:   If C_j is Partially Learned and C_i is Known, then C_j remains Known with
        $\mu_K(C_j)=\mu_K(C_i)*\mu_D(C_i,C_j)$.

▪ Concerning the knowledge level of C_j the knowledge level of C_i reduces according to:

R7:   We use the formula $p_i=min[p_i, (1-\mu_D(C_i, C_j))*p_i+p_j]$, where p_i and p_j are the student's degree of success in C_i and C_j respectively, and then using the new p_i we determine the new quadruplet $(\mu_{Un}, \mu_{UK}, \mu_K, \mu_L)$ for C_i.

Let's see some examples between the domain concept C_3 and the domain concept C_4. The prerequisite relation is $C_3 \rightarrow C_4$ and $\mu_D(C_3,C_4)=0.81$, $\mu_D(C_4,C_3)=1$.

i.Let's consider that C_4 is 40% Learned and the student is being examined at C_3 and it is being concluded that C_3 is 100% Learned, so the knowledge level of C_4 will increase according to the rule R1 and it will become Learned with $\mu_L(C_4)=max[\mu_L(C_4),\mu_L(C_3)*\mu_D(C_3,C_4)]=max[0.4,1*0.81]=0.81$. So, C_4 will become 81% Learned.

ii.Let's consider that C_4 is 100%Unsatisfactory Known and the student is being examined at C_3 and it is being concluded that C_3 is 30% Unknown, so the knowledge level of C_4 will decrease according to the rule R4 and it will become Unknown with $\mu_{Un}(C_4)=\mu_{Un}(C_3)*\mu_D(C_3,C_4)=0.3*0.81=0.243$. So, C_4 will become 24.3% Unknown.

iii.Let's consider that C_3 is 60%Unknown and the student is being examined at C_4 and it is being concluded that C_4 is 100% Known, so the knowledge level of C_3 will increase according to the rule R2 and it will become Known with $\mu_K(C_3)=\mu_K(C_4)*\mu_D(C_4,C_3)=1*1=1$. So, C_3 will become 100% Known.

iv.Let's consider that C_3 is 20%Learned and the student is being examined at C_4 and it is being concluded that C_4 is 100% Unsatisfactory Known, so the knowledge level of C_3 will decrease according to the rule R7. $p_3=min[p_3,(1-\mu_D(C_3,C_4))*p_3+p_4]= min[86, (1-0.81)*68]=min[86, 84.34]=84.34$. So the new quadruplet $(\mu_{Un}, \mu_{UK}, \mu_K, \mu_L)$ for C_3 is (0,0,1,0). In other words, C_3 will become 100% Known.

5 Discussion on the Fuzzy Cognitive Maps and Fuzzy User Modeling Used

Many web-based educational systems use a network of concepts for representing the domain knowledge and the technique of stereotypes in user modelling, in order to provide adaptation to students. Such system is CASTLE [36] in which knowledge is organized in a network of concepts with nodes representing concepts and arcs representing relations between concepts (e.g. whether some concepts must be taught before another) and each student, according to her/his progress in each domain concept, is characterised as novice, beginner, intermediate or advanced. Another system is Web-PTV [37] in which knowledge is organised in a network of concepts with nodes representing a category of concepts and arcs representing relations between concepts, that can be "part-of", "is-a" and prerequisite. Based on the student action while interacting with the system, it distinguishes between three possible states of concepts, namely "not read", "read and not known" and "read and known". Web_Tutor_Pas [38] is another web-based educational system that teaches the programming language Pascal. In this system the domain knowledge is organised into an hierarchical tree, which depicts the difficulty level of each domain topic and the order in which each topic has to be taught. According to

which topics the student knows and at what degree, the system categorised her/him to a stereotype which range from "novice" to "expert".

All these systems use the representation of the domain knowledge in order to describe what should be taught and in which hierarchy. Furthermore, the set of the concepts that are known to a student determine her/his stereotype. The student takes a test that concerns each domain concept and her/his answers determine her/his knowledge level and thus her/his stereotype. However, the relations between concepts are restricted to "part-of", "is-a" and prerequisite relations. This has as a result the knowledge of a concept not to affect the knowledge of another related concept. In other words, the knowledge of a domain concept A may means that a percentage of a related domain concept B is already known. So, it is possible that the student should not read and be examined in domain concept B. Moreover, if the knowledge of domain concept B reduces, it is possible that the student has to revise the domain concept A. Therefore, the representation of the domain knowledge should depict this kind of relations between concepts, except of the hierarchy level. That's why we use Fuzzy Cognitive Maps.

But, how we use the values of the arcs between 2 concepts in order to determine how the knowledge level of a related concept changes? An answer to this constitutes the fuzzy logic mechanism that has been introduced by Alenka Kavčič[39]. She uses six rules which determine the new knowledge level of an essential prerequisite concept C_i based on the demonstrated knowledge of the concept C_j. However, she does not determine how the knowledge level of the concept C_j is affected by the knowledge level of its prerequisite concept C_i. Furthermore, the user knowledge in her model always expands. Even when the user performs poorly, it never reduces. So, we introduce our fuzzy logic model in which the user knowledge increases or decreases, according to the user performing, and also the interaction between the knowledge level of two related domain concepts is bidirectional.

6 Conclusion

Our target in this paper was to develop a technique for creating adaptive web-based educational applications. The personalized navigation support is realized due to the representation of knowledge domain with fuzzy cognitive maps and to user model which adopts fuzzy logic techniques. Fuzzy Cognitive Maps are used to illustrate the dependencies that exist between the domain concepts. Due to the fact that student's actions, misconceptions and needs are imprecise information, we choose fuzzy logic to manage the uncertainty and to describe human descriptions of knowledge and of student's cognitive abilities. We use fuzzy sets in order to describe for each domain concept how well is known and learned and a mechanism of rules which is triggered after any change of the value of concept knowledge of a domain concept, and updates the values of concept knowledge of all related with this concept, concepts.

References

[1] Garlatti, S., Iksal, S.: Declarative specifications for adaptive hypermedia based on a semantic web approach. In: Brusilovsky, P., Corbett, A.T., de Rosis, F. (eds.) UM 2003. LNCS, vol. 2702, pp. 81–85. Springer, Heidelberg (2003)

[2] Brusilovsky, P.: Methods and techniques of adaptive hypermedia. User Modeling and User-Adapted Interaction 6(2-3), 87–129 (1996)

[3] Brusilovsky, P.: Adaptive Navigation Support. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) Adaptive Web 2007. LNCS, vol. 4321, pp. 263–290. Springer, Heidelberg (2007)

[4] Zadeh, L.: Fuzzy sets. Information and Control 8, 338–353 (1965)

[5] Zadeh, L.A.: Fuzzy logic=Computing with words. IEEE Transactions on Fuzzy System 4(2), 103–111 (1996)

[6] Katz, S., Lesgold, A., Eggan, G., Gordin, M.: Modelling the Student in SHERLOCK II. In: Greer, J., McCalla, G. (eds.) Student Modelling: The Key to Individualized Knowledge-based Istruction, pp. 99–125. Springer, Berlin (1994)

[7] Kosba, E., Dimitrova, V., Boyle, R.: Using Fuzzy Techniques to Model Students in Web-Based Learning Environments. In: Palade, V., Howlett, R.J., Jain, L. (eds.) KES 2003. Lecture Notes in Computer Science (LNAI), vol. 2774, pp. 222–229. Springer, Heidelberg (2003)

[8] Nykänen, O.: Inducing Fuzzy Models for Student Classification. Educational Technology & Society 9(2), 223–234 (2006)

[9] Panagiotou, M., Grigoriadou, M.: An Application of Fuzzy Logic to Student Modeling. In: Proceedings of the IFIP World conference on Computer in Education, WCCE 1995, Birmigham (1995)

[10] Suarez-Cansino, J., Hernandez-Gomez, A.: Adaptive Testing System Modeled Through Fuzzy Logic. In: 2nd WSEAS Int. Conf on Computer Engineering and Applications (CEA 2008), Acapulco, Mexico, pp. 85–89 (2008)

[11] Warendorf, K., Tsao, S.J.: Application of Fuzzy Logic Techniques in the BSS1 Tutoring System. Journal of Artificial Intelligence in Education 8(1), 113–146 (1997)

[12] Xu, D., Wang, H., Su, K.: Intelligent Student with Fuzzy Models. In: Proceedings of the 35th Hawaii International Conference on System Sciences (2002)

[13] Brusilovsky, P.: Adaptive Educational Systems on the World Wide Web:A review of available technologies. In: Proceedings of the Workshop "WWW-Based Tutoring" at the 4th International Conference on Intelligent Tutoring Systems (1998)

[14] Brusilovsky, P., Pesin, L.: Adaptive navigation support in educational hypermedia: An evaluation of the ISIS-Tutor. Journal of Computing and Information Technology 6(1), 27–38 (1998)

[15] Brusilovsky, P., Schwarz, E., Weber, G.: ELM-ART: An Intelligent Tutoring System on World Wide Web. In: Proceedings of the Third International Conference on Intelligent Tutoring Systems, pp. 261–269 (1996)

[16] Brusilovsky, P., Schwarz, E.: User as student: Towards an adaptive interface for advanced Web-based applications. In: Jameson, A., Paris, C., Tasso, C. (eds.), pp. 177–188 (1997)

[17] Brusilovsky, P., Eklund, J., Schwarz, E.: Adaptive Navigation Support in Educational Hypermedia on the World Wide Web. In: Proceedings of the IFIP TC13 International Conference on Human-Computer Interaction, pp. 278–285 (1997)

[18] Tsiriga, V., Virvou, M.: Evaluation of an Intelligent Web-Based Language Tutor. In: Palade, V., Howlett, R.J., Jain, L. (eds.) KES 2003. LNCS, vol. 2774, pp. 275–281. Springer, Heidelberg (2003)

[19] Peylo, C., Teiken, W., Rollinger, C., Gust, H.: An ontology as domain model in a web-based educational system for prolog. In: Etheredge, J., Manaris, B. (eds.) Proceedings of the 13th International Florida Artificial Intelligence Research Society Conference, pp. 55–59. AAAI Press, Menlo Park (2000)

[20] Kumar, A.: Rule-Based Adaptive Problem Generation in Programming Tutors and its Evaluation. In: Proceedings of the 12th International Conference on Artificial Intelligence in Education, Amsterdam, pp. 35–43 (2005)

[21] Bloom, B.S.: Taxonomy of Educational Objectives. Handbook I: The Cognitive Domain. David McKay Co Inc., New York (1956)

[22] Kosko, B.: Fuzzy cognitive maps. International Journal of Man-Machine Studies 24(1), 65–75 (1986)

[23] Kosko, B.: Neural Networks and Fuzzy Systems. Prentice-Hall, Englewood Cliffs (1992)

[24] Kosko, B.: Fuzzy Engineering. Prentice Hall, New Jersey (1999)

[25] Miao, Y., Liu, Z.Q.: On causal inference in fuzzy cognitive maps. IEEE Trans Fuzzy Syst 8(1), 107–119 (2000)

[26] Craiger, J.P., Goodman, D.F., Weiss, R.J., Butler, A.: Modeling organizational behavior with fuzzy cognitive maps. Journal of Computational Intelligence and Organizations 1, 120–123 (1996)

[27] Stylios, C.D., Groumpos, P.P.: Modeling complex systems using fuzzy cognitive maps. IEEE Transactions on Systems, Man and Cybernetics Part A 34(1), 155–162 (2004)

[28] Cole, J.R., Persichitte, K.A.: Fuzzy Cognitive Mapping: Applications in Education. International Journal of Intelligent Systems 15, 1–25 (2000)

[29] Peña, A., Sossa, H.: Cognitive Maps : an overview and their application for student modeling. Computación y Sistemas 10(3), 230–250 (2007)

[30] Peña, A., Sossa, H., Gutierrez, F.: Negotiated learning by fuzzy cognitive maps. In: WBE05, 4th International Conference Web-based Education IASTED, Switzerland, pp. 590–595 (2005)

[31] Calais, G.: Fuzzy cognitive maps theory: implications for interdisciplinary reading: National implications. FOCUS On Colleges, Universities, and Schools, 2 (1) (2008)

[32] Georgiou, D., Makry, D.: A Learner's Style and Profile Recognition via Fuzzy Cognitive Map. In: Proceedings of the IEEE International Conference on Advanced Learning Technologies (ICALT 2004), pp. 36–40 (2004)

[33] Hossain, S., Brooks, L.: Fuzzy cognitive map modelling educational software adoption. Journal Computers & Education 51(4) (2008)

[34] Aguilar, J.: A survey about fuzzy cognitive maps papers (Invited Paper). International Journal of Computational Cognition 3(2), 27–33 (2005)

[35] Tsadiras, A. K., Margaritis, K.G.: Cognitive Mapping and Certainty Neuron Fuzzy Cognitive Maps. Information Sciences 101, 109–130 (1997)

[36] Murphy, M., McTear, M.: Learner Modelling for Intelligent CALL. In: Jameson, A., Paris, C., Tasso, C. (eds.) Proceedings of the 6th International Conference on User Modeling, pp. 301–312. Springer, New York (1997)

[37] Tsiriga, V., Virvou, M.: Modelling the Student to Individualise Tutoring in a Web-Based ICALL. International Journal of Continuing Engineering Education and Lifelong Learning 13(3-4), 350–365 (2003)

[38] C.K., Virvou, M.: Personalized Teaching of a Programming language over the web: Stereotypes and rule-based mechanisms. In: Proceedings of the 8th Joint Conference on Knowledge-Based Software Engineering, vol. 180, pp. 484–492 (2008)

[39] Kavčič, A.: Fuzzy User Modeling for Adaptation in Educational Hypermedia. IEEE Transactions on Systems, Man and Cybernetics. Part C: Applications and Reviews 34(4), 439–449 (2004)

Visualization Tool for Scientific Gateway

Eva Pajorová and Ladislav Hluchý

Institute of Informatics, Slovak Academy of Sciences, Slovakia
`utrrepaj@savba.sk`

Abstract. The science gateway is important component of many large-scale Earth, astronomical, environmental and natural disasters science projects. Developing the sciences portals and the science gateways is coverage of requirements of large scale sciences such as Earth science, astronomy and all sciences which are using grid, cloud or cluster computing and high-performance computing infrastructure. The paper shows the main position of *visualization* in Science Gateway and describes architecture of the Visualization Tool (VT), for Earth and astrophysics simulations and shows some examples. VT is integrated in the web portal, as is e-science gateway for astronomy and astrophysics.

Keywords: Scientific gateway, Earth science, Grid.

1 Introduction

Since 2004 numerous scientific gateways have been developed lot of scientific gateways funded by the Tera-Grid Science Gateways program [1]. The gateway paradigm requires gateway developers to compile and install scientific applications on a variety of HPC clusters available from the resource providers in TeraGrid, to build service middleware for the management of the applications, and to develop web interfaces for delivering the applications to a user's web browser. Consequently, lot of web-service frameworks [2], [3], [4] have been designed and applied in building domain-specific science gateways. Some of them enable workflow based on the web services [4], but they commonly don't provide solutions to support web interface generation. Developers were usually hindered. Usually they need to spend a lot of time learning web programming, especially JavaScript and AJAX Technologies to implement a user-friendly and interactive web interface to these services.

Scientific gateways are able to provide a community-centric view, workflow/dataflow services and a strong support in accessing the cyber infrastructure including grid and cloud based resources. In each of science contexts, scientific gateways play a key role since they allow scientists to transparently access distributed data repositories (across several domains and institutions) and metadata sources to carry out search & discovery activities, as well as *visualization* and analysis ones, etc. Finally, scientific gateways can play an important role in training students (at academic level) in different scientific disciplines, attract new users and represent a relevant centralized knowledge repository in the sciences context. It is also a collaborative cyber-environment on which researchers working the same or similar domains can easily team up to perform computational thinking on challenging scientific problems by sharing their computational software tools and

elevating experimental datasets to scientific knowledge and innovative theories. Our paper deals with the position of *visualization* as one of the main components of scientific gateway. The scientific web portal - gateway cumulate all types of *visualization*. This paper describes VT for Earth science and for astrophysics, in which is cumulate all types of *visualization*.

Visualization tool is a part of gateway and proposes a new web based application framework for Earth science and astrophysics environment. The framework including the can import the astronomy specific workflow scripts easily can generate web appliance for running astronomical application workflows and *visualization* the outputs results directly from workflow execution, online visualization through their web browsers.

2   Visual Representation of Datasets

Simulation and execution with a huge data usually spend long execution time. Good solution for execution is represented by grid and actually on cloud computing. In both infrastructures *visualization* has the main position as a way to control the execution process. Visual control has in all infrastructure very useful position. The modal parametric studies applications include, for example, astronomical simulations. The simulation was realized as a sequence of parameter studies, where each sub-simulation was submitted to the grid as a separate parameter study. The job management was rather time consuming due to the analysis of failed jobs and to their re-submission.

Visualization is included as a visual control process. Client asks for visualization is a "visualization client". Output data on the storage element are the inputs data for visualization jobs. Visualization workers are to modify data to the formats, which can be visualized, but also to prepare the typical visualization scenes. Client can render such scenes on the browser, can make the visual control and modify executions. For example, to immediately understand the evolution of the investigated proto-planetary disc we have developed a Visualization Tool (VT). The VT is composed of several modules, which are responsible for creating scenes and converting data to, the "visualize": format. The VT is designed as a plug-in module. The components generating rendering scenes are easy to exchange, according to the requirements of the given application. In case of our gridified application the output data of the simulation located on the SE can be used directly as the input for the VT. The final product of the VT includes a set of files containing data in the VRML (Virtual Reality Modeling Language) format. These output files can be rendered by many available VRML web-browsers. The whole visualization process is maintained through a visualization script, whose basic function is invoking the individual VT components in successive steps, transferring data, and handling error events. The script is written using the Bourne shell scripts and all VT modules are implemented in the C++ language. The VT can be embedded into the framework described above, or can be used separately as a stand-alone program. By using the on-line VT the client can stop the execution process, change the input parameters and restart the execution process again. In grid environment, such architecture can be used for all applications from different science spheres which have the character of a parametric study.

Actually, the research community needs not only "traditional" batch computations of huge bunches of data but also the ability to perform complex data processing; this requires capabilities like on-line access to databases, interactivity, fine real-time job control, sophisticated *visualization* and data management tools (also in real time), remote control and monitoring. The user can completely control the job during execution and change the input parameters, while the execution is still running. Both tools, the tool for submission designed before and continued sequential visualization tool, provide complete solution of the specific main problem in Grid environment. The position of the visualization tool as a visual control process is shown in figure 1. Astrophysics scientists are able to run scientific simulations, data analysis, and visualization through web browsers.

Through Earth and astronomical science gateway scientists are able to import they sophisticated scripts by which the VT can be activated as well, as the output from workflow executions without writing any web related code [5].

2.1   VT as a New Discovery for Presenting Academic Research Results

Advance in sciences and engineering results in high demand of tools for high-performance large-scale visual data exploration and analysis. For example, astronomical scientists can now study evolution of all Solar systems on numerous astronomical simulations. These simulations can generate large amount of data, possibly with high resolution (in three-dimensional space), and long time series. Single-system visualization software running on commodity machines cannot scale up to the large amount of data generated by these simulations. To address this problem, a lot of different grid-based visualization frameworks have been developed for time-critical, interactively controlled file-set transfer for visual browsing of spatially and temporally large datasets in a grid environment. To address the problem, many frameworks for grid and cloud based visualization are used. We can go through evolution of sophisticated grid-based visualization frameworks with actualized functionality, for example, Reality Grid, UniGrid and TerraGrid.

All of the frameworks have been included in the *visualization*. Frameworks were created during grid-based projects and create new features for presentations of the academic research results in visualization. Visualization resources enabled by the astronomical science gateway the top of research experiences.

Multiple visualizations generated from a common model will improve the process of creation, reviewing and understanding of requirements. Visual representations, when effective, provide cognitive support by highlighting the most relevant interactions and aspects of a specification for a particular use. The goal of scientific visualization is to help scientists view and better understand their data. This data can come from experiments or from numerical simulations. Often the size and complexity of the data makes them difficult to understand by direct inspection. Also, the data may be generated at several times during an experiment or simulation and understanding how the data varies with time may be difficult. Scientific visualization can help with these difficulties by representing the data so that it may be viewed in its entirety. In the case of time data varying in time, animations can be created that show this variation in a natural way. Using virtual reality techniques, the data can be viewed and handled naturally in a true three-dimensional environment (e.g. depth is explicitly

perceived and not just implied). All these techniques can allow scientists to better understand their data. Viewing the data in this way can quickly draw the scientist's attention to interesting and/or anomalous portions of the data. Because of this, we encourage scientists to use scientific visualization from the beginning of their experiments and simulations and not just when they think they have everything operating correctly. This also allows scientists to develop a set of visualization tools and techniques that will help them understand their data as their research matures. For example, depending on of our astronomical example, in order to understand immediately the evolution of the investigated proto-planetary disc we have developed a Visualization Tool (VT) for astronomers.

2.2   Architecture of Visualization Tool

3D visualization service for animation of natural disasters applications, astrophysical applications and all complicated applications based on HPC (High Performance Computing), grid and Cloud computing should integrate visualization requests. Many applications from this area are using different kinds of simulation tools, which produce output data for displaying the computation results. The purpose of the visualization service is to model and display the results of various simulations. Such service requires unified standards such as integration of input data formats and especially creation of unified visualization tools.

When running parametric simulation with a large number of jobs (such astrophysical simulations), the main problem was in the grid infrastructure reliability. The job management was rather time consuming due to the analysis of failed jobs and to their re-submission. Moreover, the jobs, waiting in a queue for a long time, were blocking the simulation. To overcome these problems, we developed an easy-to-use framework based on pilot jobs concept that uses only services and technologies available in EGEE (Enabling grids for E-science) infrastructure, grid middleware gLite and Bourne Shell scripting language. The framework consists of pilot jobs – workers, and of automatic job management script. Workers are running the application code in cycle with input datasets downloaded from a storage element using remote file access. The storage element contains the input, working and output areas (as subdirectories of the directory created by the user for each parameter study). The user prepares input datasets on user interface and transfers them into the input area before starting the simulation. The working area is used by workers to store static information about computing nodes (name of the computing element and computing node, CPU type and available memory), and to monitor information updated in regular intervals, datasets, that are currently processed, and statistics about processed datasets. Output data is stored into the output area, where the user can see the progress of simulation. To check the progress, the user only needs to list the contents of the output folder. The storage element is accessible also for grid FTP clients, therefore grid portals can also be used to watch the progress. To identify hanging jobs or jobs that perform too slowly, workers are periodically sending monitoring information to the storage element. To avoid termination of workers by the queuing system, workers are running only for a limited time. The main function of the job management script is to maintain the defined number of active workers with detection of failed submissions, finished and waiting workers. The script uses job collections to speed up

the start up and automatic blacklisting of full and erroneous sites. In case of our application the output data of the simulation located on the storage element can be directly used as the input for the visualization tool. The whole process is shown in Fig. 1. The architecture of the submission process is shown in figure 2 (left). The architecture of the visualization process is shown in figure 2 (right).
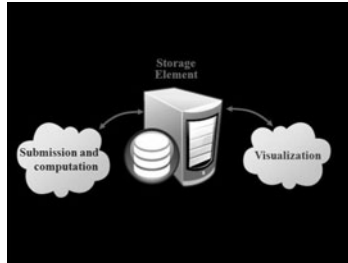
Fig. 1. Process of submission and on-line visualization

Fig. 2. Process of submission application to the grid (left), on-line visualization process (right)

In this paper, we present a visualization tool which has been used on parametric astrophysical simulations and natural disaster simulations. The first simulation project is a collaboration among Astronomical Institute (Slovakia), Catania Observatory (Italy) and Adam Mickiewicz University in Poznan (Poland). The second project performed natural disasters simulations computed at our institute. The applications were ported to EGEE grid infrastructure by the Institute of Informatics Slovak Academy of Sciences (Slovakia) [6], [7].

Natural disasters simulation is a very complicated, challenging problem sensitive to the input data required. Therefore, intense research and development o sophisticated software systems and tools is extremely important for natural disasters fighting management purposes. For example for Slovak forests, original methodology for forest vegetation classification and new fuel models have been developed and proper forest fire simulations related to the locality Krompla (National Park Slovak Paradise), where the large destructive and its reconstruction have been analyzed. These efforts induced the need of better auxiliary tools for 3D visualization o obtained simulation results and for animation of the forest fire spread. The importance is increasingly expanded for environmental problems.

VT tool for Earth Science provides pictures from simulations big Fire in Crompla region and from flood in river Vah. Both simulations was developed in our institut see Figure 3.

Fig. 3. Visualization outputs from research results in Earth science natural disasters simulations.

VT tool for astronomical applications provides pictures from simulation of the evolution of proto-planetary disc from 1Myr to 1000 Myr. An unsolved question of the Solar System cosmogony is the origin of comets and minor bodies with respect to the Solar System evolution [7]. In the past, authors predicted the existence of reservoirs of the objects and tried an effort to explain the origin and subsequent evolution of these reservoirs. Several partial theories have been developed to clarify the problem. Recently, the researchers try to present a unified theory of the formation of small-body reservoirs in the Solar System (the Kuiper Belt, the Scattered Disc), situated beyond the orbit of Neptune. In our application we developed a new improved model for explaining the formation of the Oort Cloud. One has to assume dynamical evolution of a high number of particles under gravitational influence of the giant planets: Jupiter, Saturn, Uranus, Neptune, the Galactic Tide and nearby passing alien stars. Before our work, only two similar simulations have been performed by Duncan et al. in 1987 and by Dones et al. in 2005[8]. In our application we assumed 10038 test particles. It is several times more than in the previous simulations. Our extensive simulations required very large computing capacity. To complete our model on a single 2.8GHz CPU would last about 21 years. Using the grid infrastructure, the whole computation lasted 5 months; thus, it was more than 40 times faster. The result of our simulation is dynamical evolution of orbits of test particles during the first giga year of the Solar System lifetime. Detailed study of this first giga year evolution results in a general agreement with the results of previously mentioned models as well as in new facts and questions. Having used the mentioned visualization tool we obtain Specifically, Figure 4 shows the evolution of proto-planetary disc in the time of 1 Myr. We can see that during the 1000 Myr time that the particles were replaced from inside to outside of the spheres. Pictures show the result of dynamical evolution of Oort-cloud as a part of proto-planetary disk after its evolutionary stage which was the first Gyr (giga year) [6].
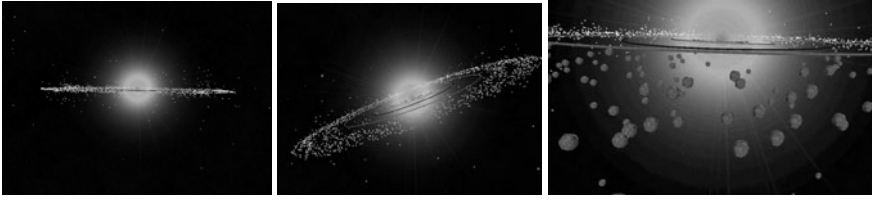
Fig. 4. Visualization outputs from research results in Astrophysics science simulations

2.3   Directly Visual Education Form

Educational visualization uses a simulation normally created on a computer to develop an image of something so it can be taught about. This is very useful when teaching a topic which is difficult to see otherwise, for example, proto-planetary disk, its evolution or evolution in Solar system. It can also be used to view past events, such as looking at the Solar system during its evolution stage, or look at things that are difficult. For astronomers, the VT has in education roles well.

3   Conclusion

The goal of the paper was to describe the VT architecture and to support the *visualization* as essential component in new portals - gateways technologies and to show some examples. For the future we want to extend the use of the VT for other scientific disciplines in addition to astronomy, but also for Earth Sciences with all *visualization* aspects. Now we are preparing the proposal for a new project of a new astronomical sciences gateway. For the future we plan to participate in a project in which the main activity will be to create and operate a pan-European e-Science

Acknowledgement

References

1. Wilkins-Diehr, N., Gannon, D., Klimeck, G., Oster, S., Pamidighantam, S.: TeraGrid Science Gateways and Their Impact on Science. IEEE Computer 41(11), 32–41 (2008)
2. Kandaswamy, G., Fang, L., Huang, Y., Shirasuna, S., Marru, S., Gannon, D.: Building Web Services for Scientific Grid Applications. IBM Journal of Research and Development 50(2-3) (2006)
3. Krishnan, L., Stearn, B., et al.: Opal: Simple Web Services Wrappers for Scientific Applications. In: IEEE International Conference on Web Services (ICWS 2006), Chicago, September 18-22 (2006)

4. Oinn, T., Addis, M., et al.: Taverna: A tool for the composition and enactment of bioinformatics workflows. Bioinformatics Journal 20(17), 3045–3054 (2004)

5. Paulech, T., Jakubík, M., Neslušan, L.: Extended modeling of the Oort cloud formation from the initial protoplanetary disc. In: On 4th International Workshopon Grid Computing for Complex Problems, October 27-29, pp. 142–150 (2008)

6. Jan, A.: Experiences from porting the astrophysical simulation The unified theory of Kuiper-belt and Oort-cloud formationi to EGEE grid. The 3rd EGEE UF

7. Neslušan, L., Leto, G., Jakubík, M., Paulech, T.: The model of the current stellar perturbations on the Oort Cloud. In: Second International Workshop on grid computing for Complex Problems, nstitute of Informatics, Slovak Acad. Sci., Bratislava, November 27-29, pp. 52–59 (2006)

8. Duncan, M., Quinn, T., Tremaine, S.: The formation and extent of the Solar System comet cloud. Astronomical Journal 94, 1330–1338 (1987); cloud. I. The reference model. Icarus

Digital Text Based Activity: Teaching Geometrical Entities at the Kindergarten

Mahmoud Huleihil[1] and Huriya Huleihil[2]

[1] Academic Institute for training arab teachers – AITAT
BeitBerlCollege, Doar Beit Berl, 44905,
Tel.: 09-7476333; Fax.: 09-7454104
[2] Kindergarten Alhuzayyel A – Rahat, South Israel

Abstract. Plane and solid geometry are important disciplines for mathematics, computer graphics and physics. In this article, different issues concerning geometry concepts are raised and discussed. First, it was observed that there is a lack of knowledge of geometrical concepts and entities among math teachers at the level of middle school, among college students and among middle school students. Second, in order to quantify the level of knowledge about basic geometrical entities, a questionnaire made of nine items was distributed and data was collected from 15 math teachers, 90 college students and 180 elementary (6 grade) and middle school students (7 and 8 grades). Third, an action plan which aims to teach some geometrical entities is suggested. It is recommended in this article to start teaching geometry at the kindergarten using different types of real objects and tools. Among the many possible tools, digital text (multimedia) is considered to be very important due to its several natural inputs to the human sensory system.

I Introduction

The fundamental rules of geometry go all the way back to the time of the ancient Egyptians and Greeks, who used geometry to calculate the diameter of the earth and the distance to the moon. They employed the laws of Euclidean geometry (named after Euclid, a Greek mathematician who lived in the 3rd century B.C.). Euclidean plane geometry involves points and lines on perfectly flat surfaces. In plane geometry, certain starting concepts aren't defined formally, but are considered intuitively obvious. The point and the line are examples. A point can be envisioned as an infinitely tiny sphere, having height, width, and depth all equal to zero, but nevertheless possessing a specific location. A line can be thought of as an infinitely thin, perfectly straight, infinitely long wire (Gibilisco, 2003, Hartshorne, 2000).

One might ask, "Why should we study geometry?" One possible answer to this question is given by Bursill-Hall, who claims that at least at one level, the answer is quite surprisingly simple. Over most of the last two and a half thousand years in the European or Western tradition, geometry has been studied because it has been held to be the most exquisite, perfect, paradigmatic truth available to us outside divine revelation. It is the surest, clearest way of thinking available to us. Studying geometry

reveals – in some way – the deepest true essence of the physical world. And teaching geometry trains the mind in clear and rigorous thinking (Bursill-Hall, 2002).

Geometry is a basic discipline, it has many applications, it is found in most of the other disciplines. Einstein initiated and stressed the role of geometry in fundamental physics. Fifty years after his death the links between geometry and physics (General relativity, electro-magnetism, quantum mechanics, nuclear forces, string theory, M-theory, topology) have been significantly extended with benefits to both sides (Atiyah, 2005).

Geometric concepts not only are used by sciences, but rather one can find geometric shapes in the fields of arts and buildings. Geometric patterns occur in rich profusion throughout Islamic cultures. They are found on diversity of materials – tiles, bricks, wood, brass, paper, plaster, glass and on many types of objects. They occur on carpets, windows, doors, screens, railings, bowls, furniture-specially pulpits in mosques, and on other surfaces. They can be seen in abundance in Persian miniatures in the illumination of the Holy Koran and on architectural surfaces of mosques, palaces, madersas (centres of education) and tombs (Abas, 2001).

Teachers, while exploring how individuals develop "different paths on their own cognitive journey of personal mathematical growth", Tall discusses about three worlds of mathematics: conceptual, symbolic, and formal (Singera and Voicab, 2008).

One of the main pedagogical tasks that mathematics teachers engage on daily basis when teaching geometry is choosing appropriate activities considering: the content they are teaching, and their students' background. There are many different frameworks and instruments that have been developed (classical and computer based) in recent years to evaluate the quality of various educational technology applications. In their paper, Manizade and Mason propose a framework for teachers' decision making when assessing appropriateness of GeoGebra applications available on-line for teaching a specific geometric content (Manizade and Mason, 2010).

The attractiveness of using technological tools (computers) relies on their multimedia capabilities and networking which could foster students' abilities, revolutionize the way they work and think, and give them new access to the world (Peck and Dorricott, 1994). More than that, Dynamic geometry systems have been described as computational environments that embody some sub domain of mathematics or science, generally using linked symbolic and graphical representations. Through computers' environment and dynamic geometry environments especially can allow students to explore the various solution paths individually and in small groups in which they make decisions and receive feedback on their ideas and strategies. Creating visual mathematical representation in the software can make it easier for students to grasp the relevant concepts by allowing the continuous manipulation of mathematical objects in real time. Teachers can thus improve their students' knowledge by eliciting mental schemas from them, which is to say the students can be guided to reach conclusions which form a step-by-step visual proof (Patsiomitou, 2008).

Another example of using computers to teach geometry concepts is given by Benacka, who developed examples to show an unusual way of using spreadsheets as a 3D computer graphics tool. The applications can serve as a simple introduction to the general principles of computer graphics, to the graphics with spreadsheets, and as a tool for exercising stereoscopic vision. The presented approach is usable at

visualizing 3D scenes within some topics of secondary school curricula as solid geometry (angles and distances of lines and planes within simple bodies) or analytic geometry in space (angles and distances of lines and planes in E3), and even at university level within calculus at visualising graphs of z = f(x,y) functions. Examples are pictured (Benacka, 2008).

A persistent dilemma for teachers of mathematics concerns how to help children understand abstract concepts, such as addition and multiplication, and the symbols that are used to represent these concepts. Teachers face a double challenge. Symbols may be difficult to teach to children who have not yet grasped the concepts that they represent. At the same time, the concepts may be difficult to teach to children who have not yet mastered the symbols. Not surprisingly, both teachers and mathematics researchers have called for better techniques to help children learn mathematical concepts and symbols (Uttal et. Al., 1997).

Hanna addressed the situation in the Israel, noting that the Israeli educational system has suffered substantial deterioration from the early 60ies until now. Furthermore; the level of the Israeli matriculation certificate has gradually declined and its value has reached a minimal stage (Hanna, 2010). From personal experience, while teaching for several years: mathematics, physics to mechanical engineering students and introduction to computer graphics, it was observed that the level of knowledge about basic geometric concept is very weak. Furthermore, it was observed that understanding and visualizing points and line are very difficult for most of the students.

In section 2, the geometry standards for young students are reviewed. The method – questionnaire and data collection are addressed in section 3. A model of action – teaching geometry basics using digital text in the kindergarten is described in section 4 and finally summary and conclusions are given in section 5.

2   Review Standards

The following review is based on the paper of Clements and Sarama, who addressed the NCTM standards of teaching geometry and early stages (Clements and Sarama, 2000) with suggestions of teaching.

From prekindergarten to grade 12, the Geometry Standard addresses four main areas:properties of shapes, location and spatial relationships, transformations and symmetry, and visualization.

Properties of shapes: The NCTM's Standards document states that instructional programs for grades pre- K–12 should enable all children to "analyze characteristics and properties of two- and three-dimensional geometric shapes and develop mathematical arguments about geometric relationships". The expectations for the early years are that children should—

- recognize, name, build, draw, compare, andsort two- and three-dimensional shapes;
- describe attributes and parts of two- and threedimensionalshapes; and
- investigate and predict the results of putting shapes together and taking them apart.

Do teachers need to address these ideas with very young children? Yes! Young children's ideas about shapes—including limited ideas—stabilize as early as six years old. If a six-year-old believes, for example, that non-isosceles triangles are not triangles, he or she will likely continue to believe that misconception for years to come—regardless of what the teacher or a textbook says.

Location and spatial relationships: The Standards document states that pre-K–12 students should be able to "specify locations and describe spatial relationships using coordinate geometry and other representational systems". Young children should—

- describe, name, interpret, and apply ideas of relative position in space;
- describe, name, interpret, and apply ideas of direction and distance in navigating space; and
- find and name locations with simple relationships, such as "near to," and use coordinate systems, such as those in maps.

Moving and mapping on computers Computer activities can facilitate children's learning of navigational and map skills. Young children can abstract and generalize directions and measurement by working with navigation environments that require them to direct a car or an animal around the screen. One kindergartner abstracted the geometric notion of path by saying, "A path is like the trail a bug leaves after it walks through purple paint." Simple coordinate games on computers can help children learn location ideas. For example, the on-screen version of Battleship requires players to guess a location by given coordinates, such as "B, 5." When children enter a coordinate to move an object but it goes to a location that is different from what they had planned, the feedback is natural and meaningful.

Transformations and Symmetry: The Standards document states that pre-K–12 students should "apply transformations and use symmetry to analyze mathematical situations". Young children should—

- "recognize and apply slides, flips, and turns";and
- "recognize and create shapes that have symmetry."

Geometric motions: Children use geometric motions intuitively when they solve puzzles. They turn the pieces, flip them over, and slide them into place. If they use computer programs to solve puzzles, they must choose each motion deliberately. Such activities help students become aware of the motions and the result of each one. They also learn that changing an object's position or orientation does not change its size or shape.

Symmetry: Many activities help children explore symmetry. Children's unit-block buildings frequently display symmetry. Teachers can help the children make symmetric shapes using mirrors and by folding and cutting paper. Children can also explore symmetry by using computers.Activities that ask children to complete the other "half" of a symmetric design or explore pattern-block designs with built-in mirrors allow children to explore symmetry dynamically.The design explored in such activities is always symmetric.

Visualization: The Standards document recommends that pre- K–12 students "<u>use visualization, spatial reasoning, and geometric modeling to solve problems</u>". Young children should—

- form mental images of geometric shapes by using spatial memory and spatialvisualization;
- recognize and represent objects from different points of view;
- relate geometric ideas to number and measurement ideas; and
- recognize and locate geometric shapes and structures in the environment.

Activities based on exploring real objects and playing yard games, could enhance visualization. Computer programs (e.g. MsPaint, Excel, and PowerPoint) are essential for developing imagination through pictures, sounds and animation.

3 Method – Data Collection and Observations

In order to check the level of knowledge about basic geometric entities, a questionnaire made of nine items were prepared. The questionnaire were distributed among elementary and middle school math teachers (15 teachers who work in 6 different schools), among college students (90) and among elementary and secondary school students (180, 6th-8th grades). The questionnaire included open items, which allowed the questioners to express their knowledge in free written text. The questionnaire and the collected data were arranged in the following table.

Table 1. Questionnaire items

Item #	Item value
1	Define point
2	Define straight line
3	Define ray
4	Define angle
5	Define parallel lines
6	Define plane
7	Define space
8	Define axis
9	What is the value of the resulted angle after dividing the plane into four equal parts? What its value in the space? Explain

Table 2. Responses from teachers (15 teachers from 6 different elementary and middle schools)

Item	Response #	#% of responses
1 – point	Two coordinates in two axis system	13
	It is used to specify a location, it has no dimensions	27
	It is a point with no size	13
	It has no definition	20
	I don't know	27
2 – straight line	A trajectory of a point	7
	Set of points which produce a straight line	20
	Number of infinite points which produce straight line	13
	A line connecting two points	20
	It has no definition	20
	I don't know	20
3 – ray	A straight line which has a starting point	80
	I don't know	20
4 – angle	It is produced from intersection of two rays	27
	It is the intersection of two straight lines	20
	The enclosed area between two rays	20
	The meeting point of two rays	33
5 – parallel lines	Two opposite lines which don't intersect	40
	Lines which don't intersect, there is a constant distance between them	60
6 – plane	Straight surface	13
	Two axes which represent length and width	27
	It is a space with length, width and height	20
	It is an axiom	13
	I don't know	27
7 – space	It is three axes (length, width, height)	20
	Space produced by a geometric shape	33
	It is a part of space	13
	I don't know	34
8 – axis	It is a straight line which includes many equidistance points	27
	A line	27
	Axes system	13
	A straight line with arrow points to the axis direction	13
	I don't know	20
9 – value of angle	45 degrees	27
	90 degrees	33
	Body centre	13
	I don't know	27

Observations from the teachers' responses: 1) approximately 20-30% of the teachers claimed that they don't know; 2) most of the definitions either incomplete or inexact.

Table 3. Responses from college students' (sample size=90)

Item	Response #	#% of responses
1 – point	A location in the space	29
	It is a point in the plane	4
	It has no definition	8
	It is an axiom	11
	It is coordinates of two points (x,y)	18
	I don't know	30
2 – straight line	Set of points falls in some direction	24
	It is a line which connect between two points	21
	It is a trajectory of a point	13
	It is a set of points on the same plane	17
	I don't know	25
3 – ray	It is a straight line which has a starting point	71
	It is infinite straight line which has no ends	23
	I don't know	6
4 – angle	It is the area between two rays	18
	It is the intersection of two rays	31
	It is the intersection of two lines	16
	It is a point between two rays	6
	It is the point of intersection of two rays	21
	I don't know	8
5 – parallel lines	Two nonintersecting lines	10
	Straight lies which does not meet	23
	Parallel lines does not meet	15
	Two lines with the same slope, never intersect	14
	Two lines which does not intersect and the distance between them is fixed	21
	Opposite lines which never intersect	10
	I don't know	7
6 – plane	It is a two dimensional surface	18
	I don't know	82
7 – space	It is empty space, nothing	16
	It is three dimensional shape	7
	I don't know	77
8 – axis	It is the intersection of axes x and y	18
	It is center point	9
	It is a straight line with numbers	16
	I don't know	57
9 – value of angle	90 degrees in the plane	43
	Other angles	17
	I don't know	40

Observations from the college students' responses: 1) there is some knowledge about items 1 to 5, but; 2) there is al little knowledge about items 6 to 10. Although plane geometry is taught to all students at the middle age level, it is interesting to note that the concept of a plane is not clear to the them.

Table 4. Responses from elementary and secondary students (sample size=180)

Item	Response #	# of responses
1 – point	It is a point	27
	Set of points	4
	It is a point taken from the edge of an angle	3
	It is similar to the decimal point	5
	It is a basis for shapes	3
	When the angle is 90 degrees we find the point	4
	It is the diagonal of the ray	2
	It is the end of a sentence	2
	It is similar to anything	2
	It is intersection of lines	15
	It is a point which we find in fractions	5
	It is the intersection of lines to produce angles	2
	I don't know	26
2 – straight line	It is a set of points	23
	It is a line used to produce geometric shapes	36
	I don't know	41
3 – ray	It has a start but not an end	29
	It is parallel lines	3
	It starts at an angle	3
	It extends from angle to angle	5
	It is a diameter like in the square	9
	It is a not ending straight line	32
	Other definitions: side of a square, side of an angle	15
	…	4
	I don't know	
4 – angle	It is an open and broken line	7
	It is two rays which starts from the same point	15
	It is a ray and angles of different types	39
	It is intersection of two lines	9
	Other definitions: angle of a triangle, angle of a square angle of the classroom …	19
	I don't know	11

Table 4. (*continued*)

5 – parallel lines	Two opposite lines	11
	Two not meeting lines	31
	Two equal lines	17
	Two side by side lines	4
	Nonintersecting perpendicular lines	7
	Intersecting perpendicular lines	4
	I don't know	26
6 – plane	It is a direction	18
	It is angles of different kinds	13
	It is a triangle, a square, etc…	21
	It is the surface of something	16
	I don't know	32
7 – space	There is nothing	38
	It is a kind of empty zone	18
	It is similar to an empty circle, empty square …	11
	I don't know	33
8 – axis	It is a straight line	19
	Axis of numbers, line of numbers …	49
	I don't know	32
9 – value of angle	45 degrees=180:4	18
	180 degrees	8
	90 degrees=360:4	19
	From 90 to 180 degrees	4
	A square of four sides	6
	I don't know	45

<u>**Observations from the elementary and middle age school students' responses**</u>: elementary and middle school students showed a little knowledge about geometry concepts.

A general result which is obvious from the answers of all the questioners' responses is the lack of knowledge about solid geometry.

The observed lack of knowledge of basic geometric concepts, probably starts at early ages and continues towards university level. One of the main reasons for this phenomenon might be explained by the fact that teachers adopt Piaget theory of cognitive development. Piaget identified four stages in cognitive development: sensory-motor, pre-operational, concrete, and formal (Balke and Pope, 2008).According to Piaget theory, teachers cannot teach concepts if the students did not reach appropriate mental level. Thus, teachers while adopting Piaget theory, with the difficulty of identifying the proper age, they limit themselves from introducing abstract concepts at early stages. This limitation propagates through the years as one might observe while teaching physics and introduction to computer graphics at the university level.

A solution to this problem, might stem from adopting Vygotsky's theory of cognitive development. Vygotsky developed concepts of cognitive learning zones. The Zone of Actual Development (ZAD) occurs when students can complete tasks on

their own.On the other hand, Vygotsky suggests that students are able to grasp abstract concepts at the age of eight Balke and Pope, 2008).

According to our observations regarding the level of knowledge of the basic geometric concepts, we suggest that these concepts should be taught starting at early childhood.

In the following section we propose computer based activities for teaching points and lines in the kindergarten.

4 Digital Based Activities at the Kindergarten

When young students are asked about points and lines, usually they draw a dot and a straight segment on a piece of paper, pointing to either one of each and claiming: "this is a point, this is a line". The difficulty to describe a point by words is raised due to the fact that a point is an abstract concept with no size. Thus, it is difficult to visualize it. So, the first mission of a teacher is to help in constructing an image in the student's memory. This could be done by variety of methods, including real objects, stories, games, and computer graphical programs. Based on its definition, i.e., a point, actually, is an expression of a location one might reach or specify, it should be easy to link the point concept to some location in the neighboring environment (trees or any object could be found in the neighborhood) or in the distant space (the moon, the sun, and the stars). At the kindergarten, there are many hours of watching educational movies, which could be used to address many locations in the outer space, on the surface of our planet and in the near neighborhood. Insects, e.g., Ants or Bees could be viewed as points (a misconception that should be relaxed in later stages of development). Afterwards, Ants or Bees could be traced to describe lines (straight or curved). These real activities are helpful and are used as complementary and pre (and / or simultaneously) to computer activities which are described next.

4.1 Using Graphical Programs (Mspaint)

Microsoft paint is an example of a graphical program which emulates the computer screen as an array of pixels (which could be seen as models for points). Then young students could draw dots with different sizes (by choosing different brush sizes). The students brainstorm with the teacher the concept of a point by relating to the size of the dot been drawn on the screen. The discussion is accompanied with drawing smaller and smaller dots. Along with this activity, the young students classify fruits and arrange them in a row from biggest to smallest. The occupied position of each dot or each fruit is highlighted and pointed to.

The next concept is a Line. A line is defined as an infinite set of points which spans a specified direction. Thus, the graphical program facilitates visualizing the concept. The real world activity could be of arranging lentils in a line. The concept of infinity could be easily discussed with young students due to the practical limitation of the amount they could bring from home (mass and money. In a summer day, while visiting the fields, Ants could be traced and photographed via a digital camera, and in a later time, the pictures should be viewed at the computer screen. The dual activity is important to establish concepts in the memory of the learners.

Graphical programs are ideal for demonstrating and clarifying concepts like symmetry, rotation, axis of rotation, drawing shapes (in 2D and 3d). The interactive feature of such activities increases the effectiveness of the learning process.

4.2   Using Slide Shows (PowerPoint)

Slide shows are perfect for animating locations (points), paths (straight lines) and shapes. These concepts (and other concepts) are visualized as moving graphical objects and entities. Building educational slideshow movies with the young students is powerful for grasping new concepts, including abstract ides. Again, acting in a dual fashion is essential to the process of learning. Some ideas of slideshow movies which are important for the young children are: crossing traffic lights with adults (highlighting the proper behavior expected from them) highlighting dangers following wrong behaviors, an activity which is a perfect match to a slideshow movie; a race between rabbits; the eagle and the rabbit (up-down, sky-earth).

4.3   Using Digital Cameras

Digital cameras linked to distant computers are used in some kindergartens in Israel, as a necessity for the parents. Recently, there were objection to use digital cameras due to exposing the children and the staff continuously. For the current interest, digital cameras are excellent choice to facilitate the process of learning. Digital cameras, with the ability of capturing movies enable performing the dual activities simultaneously. The special location of the giraffe kindergarten (which is located in south of Israel-Rahat city), in the desert with green neighborhood (trees and vegetables planted by the teacher and young students) enables simultaneous dual activities (e.g. capturing the plants growth, each plant in its specified location (Fig. 1), capturing insects (Fig. 2) and other available things (e.g. the moon Fig. 3).

Fig. 1. Picture of the bee occupying a specific location on the

Fig. 2. Picture of the plants showing their specific location

Fig. 3. Picture of the moon representing a location in the space

4.4   Using Spreadsheets (EXCEL)

Spread sheets are good candidates for visualizing points. By formatting the cell width and height, it is possible to test different cell sizes, starting from bigger cells to smaller cells. Coloring activity enables practicing mouse movements and reaching

different locations. Painting cells in some organized manner makes it possible to make different shapes (horizontal lines, vertical lines, slanted lines, squares, triangles, rectangles, circles, trees, and any possible shape. Excel is ideal of creating XY charts. With the ability of macros, drawing points, lines and shapes could be simulated and animated by applying the built-in DoEvents macro.

5   Summary and Conclusions

In this article, the basic geometric concepts (point, straight line, ray, angle, parallel lines, plane, space and axis) were discussed. Three main points were raised: 1) the level of knowledge of basic geometric concepts were observed to be weak among teachers at elementary and middle age Arab schools, among college students and among elementary and middle age students (6th -8th); 2) the aforementioned level of knowledge were quantified and data were collected via an open questionnaire which included 9 geometric concepts (table #1). The sample size included 15 teachers from 6 schools, 90 college students and 180 elementary and middle age students; 3) and finally an action plan in the form of digital text based activities were suggested to teach geometric concepts starting at the kindergarten.

The results were arranged in tables (2-4). The observations that one can read in the tables are as follows: 1) the teachers showed some fair knowledge about concepts 1-5; 2) the college students showed some knowledge about concepts 1-5; 3) the elementary and middle age students showed weak knowledge about all the concepts; 4) all the groups showed very weak knowledge about solid geometry.

Due to the fact that these basic concepts are abstract in their nature, it is difficult for the students to visualize them. Based on this understanding, the first mission of the math teacher is to construct these concepts in the memory of the learners via appropriate activities (which include real objects and computer based visualization).

Following Vygotsky's theory for cognitive development (opposing Piaget's theory for cognitive development, which accounts for four stages of development) assuming that young students are able to grasp abstract concepts, the following computer software types were applied in the giraffe kindergarten which located at south of Israel, in the Rahat city: 1) mspaint as a graphical tool; 2) Microsoft PowerPoint for preparing slideshow movies; 3) digital cameras as a tool for documentation and capturing movies; and 4) Microsoft Excel (spreadsheet) as a simulation and animation tool. The computer based activities accompanied with real world activities empowers the process of learning.

This type of implementation, just started in the giraffe kindergarten, showed positive and promising results expressed in the observed motivation among the young learners. Accordingly, we suggest that this type of dual action plan should be implemented for teaching geometric concepts starting from the kindergarten to university level. After all , it is difficult to specify the exact moment when the students are ready to accept abstract concepts and when they reach the level of abstract thinking.

References

1. Abas, S.J.: Islamic geometrical patterns for the teaching of mathematics of symmetry. Symmetry in Ethnomathematics 12(1-2), 53–65 (2001); Special issue of Symmetry: Culture and Science. Budapest, Hungary: nternational Symmetry Foundation
2. Atiyah, M.: Einstein and geometry. Current Science 89(12), 25, 2041–2044 (2005)
3. Blake, B., Pope, T.: Journal of Cross-Disciplinary Perspectives in Education 1(1), 59–67 (2008)
4. Bursill-Hall, P.: Why do we study geometry? Answers through the ages. In: DPMMS Centre for Mathematical Sciences Wilberforce Road, Cambridge, pp. 1–31 (2002), `http://www.dpmms.cam.ac.uk/~piers/F-I-G_opening_ppr.pdf` (last visited February 11, 2011)
5. Jan, B.: 3D Graphics with Spreadsheets. Spreadsheets in Education (eJSiE) 3(1), Article 7 (2008)
6. Clements, D.H., Sarama, J.: The earliest geometry, Early childhood corner, NCTM, 1906 Association drive, Reston VA 20191-9988, pp. 82-86 (2000)
7. Gibilisco, S.: Geometry Demystified, p. 3. McGraw-Hill, New York (2003)
8. Hanna, D.: Ten Lies, Half-Truths and Prejudices of the Israeli Education System. International Online Journal of Educational Sciences 2(2), 319–338 (2010)
9. Hartshorne, R.: Geometry: Euclid and beyond. Springer, New-York (2000)
10. Manizade, A.G., Mason, M.: Choosing geogebra applications most appropriate for teacher's current geometry classroom: pedagogical prospective. In: GeoGebra NA 2010, Ithaca College, Ithaca, NY, USA (2010)
11. Patsiomitou, S.: The Development of Students Geometrical Thinking through Transformational Processes and Interaction Techniques in a Dynamic Geometry Environment. Issues in Informing Science and Information Technology 5 (2008)
12. Peck, K.L., Dorricott, D.: Realizing the Promise of Technology Pages 11-14 Why Use Technology? Educational Leadership 51(7), 11–14 (1994)
13. Singara, F.M., Mihaela, F., Voicab, C.: Between perception and intuition: Learning about infinity. Journal of Mathematical Behavior 27, 188–205 (2008)
14. Uttal, D.H., Scudder, K.V., Deloache, J.S.: Manipulatives as Symbols: A New Perspective on the Use of Concrete Objects to Teach Mathematics. J. App. Dev. Psych. 18, 37–54 (1997)
15. Geometry for elementary school, Office of Elementary Instructional Services Virginia Department of Education P.O. Box 2120 Richmond, Virginia 23218-2120 (2003), `http://www.math.vcu.edu/Statewide_Webpages/Course_Descriptions/geometry/session_3/Geometry_elementary.pdf` (last visited February 11, 2011)

Cross Format Embedding of Metadata in Images Using QR Codes

Athanasios Zigomitros and Constantinos Patsakis

Department of Informatics, University of Piraeus

Abstract. Image steganography has various usages due to the recent advances in technology, specially in the field of data communication. Even though steganography has been used so far mostly to secretly embed data, we make use of it for embedding cleartext data, which everyone can has access. This work proposes a new method for cross format metadata embedding in images that is based on QR codes, that seems to have several significant advantages over current method of storing metadata.

Keywords: steganography, QR codes, LSB, image metadata.

1 Introduction

Steganography has been used for centuries and it is only recently that it has become a standardized, well studied and formalized part of computer science. The main reason for its adoption is its core nature, the fact that it enables us to embed messages in other mediums. The fact that current technology gives us the privilege to communicate using several medias, compared to the past and that it enables others to have access to communication via covert channels, triggers a need for privacy and secrecy.

Of course steganography can have malicious applications, depending on the nature of the content of the message that is embedded in the medium, yet as a knife can be used for good or malicious objectives, steganography has more to offer. In the past years, steganography has been widely used for watermarking and applying DRM in images, audio and video assets, as it doesn't alter the mediums significantly and the quality remains the same.

Image steganography has several techniques, the most famous is LSB embedding, where the data are being stored on the least semantic bits, so that they are not detectable by the eye. Moreover, if the pixels that are used are picked using a good and secure pseudo-random function, then the embedded information remains secure. Other techniques use the frequency domain like DCT and DWT. A good introduction to steganographic techniques is given in [1,2].

This work tries to make use of QR codes in order to embed image metadata inside them. The reason for doing so is that when changing image format metadata, which in many cases are very useful, are lost. Our approach towards this problem is to alter the image and embed the metadata in the image, so that if lossless image formats are used, then the metadata remain untouched. The

proposed method provides several advantages that are going to be discussed in the following sections.

2 QR Codes

Barcodes are being used for more than three decades in order to help us extract data from databases about the objects to which they are attached. From product prices to blood rhesus and phenotype, barcodes have proved to be very beneficial. Yet, they have an inherent problem due to their structure. Barcodes are not suitable to store information, but only tags about the information that can later be used to identify it using a proper database.

QR codes come to fill in this gap, as they are a generalization of barcodes. These codes can be regarded as the two dimensional version of barcode and have started gaining wider acceptance as an industry standard. The name is from the initials of "Quick Response" and depending on the standard that will be used may contain several thousands bytes of information. Semacode [4], which is very popular as well, is another form of two dimensional barcodes and has been developed to store mainly URLs, thus their structure does not store much information. Both of these codes resemble a crossword puzzle but follow special patterns in order to store the information and be easily read from mobile devices.

Due to the development of smart-phones, QRs have found their way on the market, being used by many applications on iOS and Android mobile platforms. An example of a QR and its anatomy is presented in Figure ??. For extensive information on QR codes and its standards the reader may refer to [3].
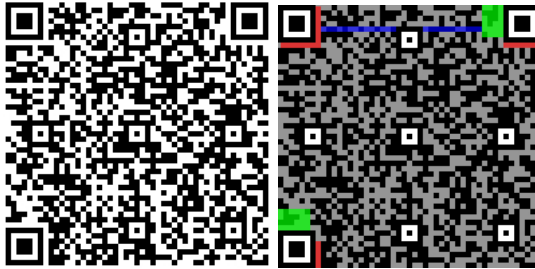
Fig. 1. An example QR code containing as data the text "Cross format embedding of metadata in images using QR codes. Athanasios Zigomitros, Constantinos Patsakis. Department of Informatics, University of Piraeus.". On the left its anatomy is presented, timestamps, version etc.

3 Our Proposal

The use of QR in steganography has already been made [5,6,7], but the applications of the proposed method are wider and more generic.

The proposed algorithm takes as input the data to be stored in QR form and a random seed R. The random seed is stored in a 32 by 32 square which is stored in the upper left corner of the picture using LSB. The result is that the image is divided as in Figure ??, as an alternative, the first row of the pixels of the image can be used as the random seed. We use the upper square in order to create to more rectangles which will store information like, what type of content is stored, what is the size of the QRs how they are arranged or even the hash of the embedded QRs.

The random seed R is used in order to initialize a pseudo random permutation P which will scramble the QR squares in the image. After the scramble, the QRs are embedded in the image by altering the least significant bits of the image. The extraction process is exactly the reverse. On receiving the image, we extract R and we have the random permutation, thus by using P^{-1} we can find the bits which that are embedded in the QRs.

Both the embedding and extraction processes can be seen in Figure 3.
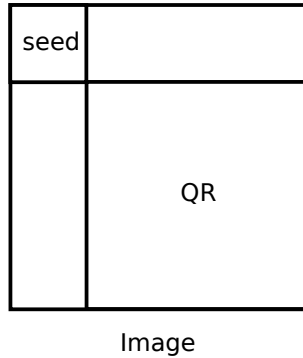
Fig. 2. QR embedding

4   Results

The tests that are presented in this work were made with grayscale pictures, but could easily be applied to full color ones. Figure ?? shows several well known "free" images and the image after QR embedding. For the images that we used in our experiments, we made some image measurements and the results can be seen in Table 1. The next two figures, Figure ?? and Figure ??, show two distorted images and the recovered QR. Due to the structure of the QR and the scrambled embedding of it at the image a noise removal filter can improve the quality of the retrieved QR when the watermarked image was distorted. The QR can also be used as a fragile watermark and in his scrabbled form is possible to show a pattern of distortion in the altered image Figure ?? (b).
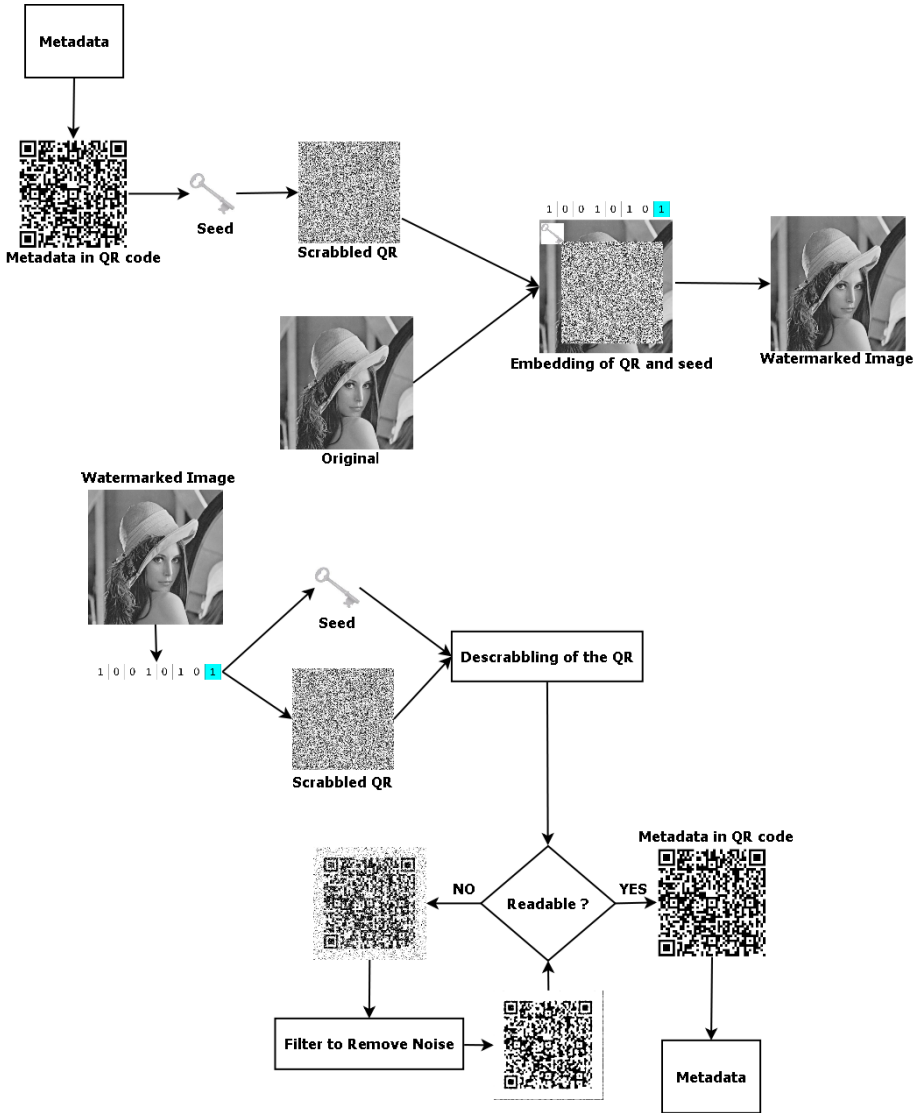
Fig. 3. Embedding and extraction process

(a) Lena picture and on the right the watermarked version.
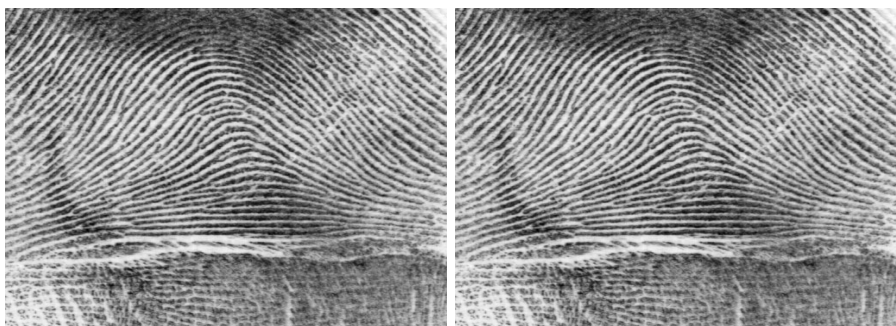
(b) Fingerprint picture and on the right the watermarked version.
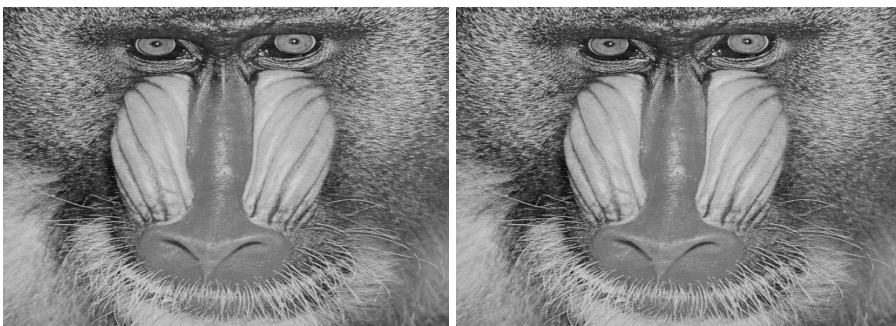
(c) Baboon picture and on the right the watermarked version.

Fig. 4. Image results for some well known pictures

(a) Lena watermarked picture and on the right the distorted version.
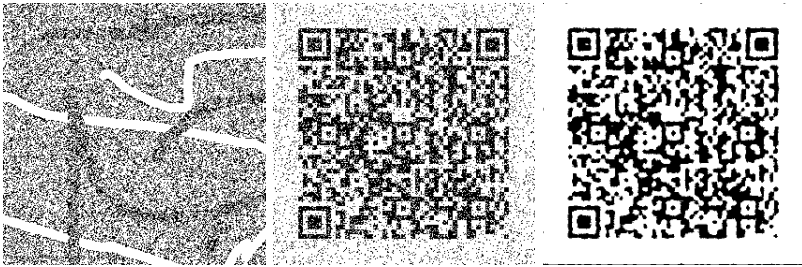
(b) Retrieved Scrambled QR, Retrieved QR data before and after filtering.

Fig. 5. Image results for some well known pictures after embedding and distortion

Table 1. Image measurements

Measurement of 51 test images	Value
Mean of Mean Square Error	0.1513
Mean of Peak Signal to Noise Ratio	56.3397
Mean of Normalized Cross-Correlation	0.9995
Mean of Structural Content	1.0010
Mean of Average Difference	0.0563
Mean of Maximum Difference	1
Mean of Normalized Absolute Error	0.0016

The measurements were made using the Matlab package Image quality measures, developed by Athi Narayanan.

(a) Lena watermarked picture and on the right the distorted version.
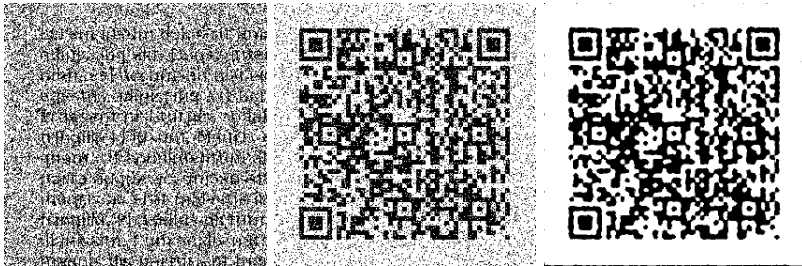
(b) Retrieved Scrambled QR, Retrieved QR data before and after filtering..

Fig. 6. Image results for some well known pictures after embedding and distortion

5   Applications

The proposed method has several advantages to offer with its usage. Firstly it enables independent cross format metadata over lossless image formats, thus important data like creator, dates etc can be easily embedded and transmitted without any losses if the image is saved in another format. Moreover, a standardized metadata embedding can prove to be an ideal method for speeding up image searches on the Internet, through search engines.

Secondly, the method can be used to compress usual HTML code. Since the anchor (<a>) tag in common HTML code is very usual, it can be embedded inside the picture compressing usual HTML code significantly, given the fact that current web pages have re-directions when clicking images. Moreover, the QR could have javascript code embedded, offering fancy image effects when interacting with the photos. Of course this technique will demand more processing from the browser, yet this cost may be overridden by the decrease of traffic load on central servers and routers.

The proposed method can retrieve data after image cropping and distortion. QRs have already embedded in their structure error detecting and correcting codes that can be used to retrieve image contents if the distortion is not severe. In the case of image cropping, depending on what parts of the image have been left out, embedded data can still be retrieved, as not all of the embedded QRs will have been removed.

By having access to image metadata, it is very easy to make cross site checks for digital rights infringement. In most cases the images have small distortion or they are saved on different image format. Using the proposed metadata embedding techniques, one could detect easier these cases and prove their true origin, yet it is obvious that the method alters any already embedded hidden watermark.

The proposed method could be used in medical images as well, where metadata is an essential need and the alteration of the images must be minimum. For applications of steganography and watermarking and in medical images the reader may refer to [8].

6   Conclusions

Currently, the proposed method enables browsers to have access to the image metadata independent of the format, while enabling images to have extra code embedded. On one hand this compresses the data that are transmitted, while on the other hand it makes images "alive".

A possible extension of the presented method would be to enable high definition cameras on mobile device to be able to extract stored information of images, printed or not. The human eye can't read barcodes or QRs so there is no point of seeing it. Of couse the trade off would be the quality loss of the image because the noise that interferes between the image and the camera won't make it readable by the device with the proposed method but a possibly solution to this is to altering bits of higher value or by applying certain filters over specific areas of an image. This would enable cameras to take photos of already content-aware objects and suppling the embedded information.

This method easily can be extended to video files, where one could embed metadata to every frame of the video e.g. subtitles.

References

1. Wayner, P.: Disappearing Cryptography: Information Hiding: Steganography & Watermarking, 3rd edn. Morgan Kaufmann, San Francisco (2008)
2. Cox, I., Miller, M., Bloom, J., Fridrich, J., Kalker, T.: Digital Watermarking and Steganography, 2nd edn. Morgan Kaufmann, San Francisco (2007)
3. QR codes, http://www.denso-wave.com/qrcode/
4. Semacode, http://www.semacode.com/
5. Zhang, S., Yoshino, K.: DWT-Based Watermarking Using QR Code. Science Journal of Kanagawa University 19, 3–6 (2008)

6. Chen, W.-Y., Wang, J.-W.: Nested image steganography scheme using QR-barcode technique. Optical Engineering 48(5), 057004 (2009)
7. Chung, C.-H., Chen, W.-Y., Tu, C.-M.: Image Hidden Technique Using QR-Barcode. In: Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pp. 522–525 (2009)
8. Coatrieux, C.G., Lecornu, L., Roux, C., Sankur, B.: A Review of Image Watermarking Applications in Healthcare. In: EMBC 2006: IEEE Int. Conference on Engineering in Medicine and Biology, New York (September 2006)

An Empirical Study for Integrating Personality Characteristics in Stereotype-Based Student Modelling in a Collaborative Learning Environment for UML

Kalliopi Tourtoglou and Maria Virvou

Department of Informatics, University of Piraeus, 80 Karaoli & Dimitriou St.,
18534 Piraeus, Greece
`ktourtog@yahoo.gr, mvirvou@unipi.gr`

Abstract. The aim of this paper is to present an empirical study for defining the appropriate triggers for personality related stereotypes. These stereotypes are used for modelling the students in a Computer Supported Collaborative Learning (CSCL) system for UML. The system builds individual student models to provide adaptive and intelligent advice concerning their knowledge and the most adequate colleagues for collaboration. Most of existing CSCL systems and Intelligent Learning Environments (ILE's) also include student models. However, the vast majority of them use the student models to describe the students from the perspectives of knowledge and/or participation in collaborative activities. In our approach, the student models additionally to the knowledge describe the students regarding specific personality characteristics related to their learning and collaboration attitudes. The student models are built using the stereotype-based method, which entails the definition of the stereotypes, their facets and the triggers (the triggering conditions for a student to belong to a stereotype). As the definition of the triggers is a task of high importance for the effectiveness and accuracy of the student models, we conducted an empirical study among experienced trainers of software engineering.

Keywords: Collaboration, collaborative learning, CSCL, stereotypes, student modelling, UML, intelligent learning environment, triggers.

1 Introduction

The concept of collaborative learning, the grouping and pairing of students for the purpose of achieving an academic goal, has been widely researched and advocated throughout the professional literature (Gokhale 1995). Over the last ten years, cooperative learning has become accepted as one of the "best practices" in education (Cohen et al. 2004). As computer technology developed, Computer-Supported Collaborative Learning (CSCL) systems were established. CSCL systems are learning software environments that allow distant users to collaborate with each other promoting effective collaborative learning processes. Computer-supported collaborative learning (CSCL) is one of the most promising innovations to improve teaching and learning with the help of modern information and communication technology (Lehtinen et al. 1999).

In this paper, we present our work aiming to provide an effective CSCL environment for teaching UML, which is called AUTO-COLLEAGUE (AUTOmated COL-LaborativE leArning Uml Environment). Trainees working in this environment can collaborate with each other as they learn UML. The system provides adaptive and intelligent help to the trainees in order to facilitate collaborative learning. This help is adapted to the special needs and personality characteristics of each trainee. It concerns (a) indications on which help topics to study and (b) suggestions on the most appropriate colleagues to collaborate with.

Many effective Computer-Supported Collaborative Learning (CSCL) systems have been developed during the last decade, such as (Rosatelli and Self 2004), (Baghaei and Mitrovic 2005), (Chen et al. 2006), (Martínez Carreras et al. 2004), (Tamura and Furukawa 2008)**,** (Khandaker et al. 2006), (Lukosch et al. 2006)**,** (Rick and Guzdial 2006)**,** (Edwards et al. 2001). The main purpose of these systems is to allow remote users to collaborate with each other while working in the same environment at the same time. Some of them are platforms where users can share data in various formats (e.g. documents), but there is no advice mechanism and no common goal/problem to solve as a team. Not all of the systems offer advice to users. In the rest of the systems, the content of the advice is generated considering the level of expertise and the participation of the users in social activities (chat, whiteboard etc). AUTO-COLLEAGUE is, also, a CSCL system. Unlike the aforementioned CSCL systems, AUTO-COLLEAGUE offers adaptive advice to the trainees according not only to their state of knowledge but, also, to personality characteristics related to their learning and collaboration processes.

The essential functionality for every adaptive learning environment is to maintain individual student models. It has long been recognized that in order to build a good system in which a person and a machine cooperate to perform a task it is important to take into account some significant characteristics of people. These characteristics are used to build some kind of a "user model" (Rich 1983). Student modelling is a special type of user modelling which is relevant to the adaptivity of intelligent tutoring systems (Elsom-Cook 1993). There are many methods for building user models, such as the stereotype theory, the overlay model (Virvou and Tsiriga 2001), the perturbation (or buggy) model (Faraco et al. 2004), fuzzy logic (Kavcic 2004), and Bayesian networks (Reye 2004).

In order to achieve adaptivity, AUTO-COLLEAGUE builds individual student models that describe the trainees regarding their level of expertise and personality. These student models are built following the stereotype-based theory proposed by Rich (1979). This theory simulates the way people make assumptions on others based on relevant information about them**.** Stereotypes constitute a very powerful technique in providing information based on few observations (Tsiriga and Virvou 2002). It has been used effectively in adaptive systems (Ardissono et al. 2004), (Goren-Bar 2001) to build user models and adapt their content according to the user profile. The stereotype-based theory has, also, been widely used in intelligent learning environments in order to serve the educational and learning needs of the individual student (Jeremic et al. 2009), (Hatzilygeroudis and Prentzas 2004), (Kabassi et al. 2006), (Wei et al. 2005), (Virvou and Moundridou 2001), (Surjono and Maltby 2003), (Tsiriga and Virvou 2002). The facets used in the stereotypes of these intelligent learning environments are related to the errors, performance, knowledge and level of expertise of the

learners. Some of these systems include other characteristics such as interests, preferred exercise types and multimedia type preference (Jeremic et al. 2009)**,** carefulness while solving exercises (Virvou and Moundridou 2001), (Tsiriga and Virvou 2002) and concentration level (Hatzilygeroudis and Prentzas 2004). Especially, in (Surjono and Maltby 2003) specific learning styles are also included in the student model to be used (it has not yet been implemented, only analyzed).

Similarly to the aforementioned intelligent learning environments, the stereotype-based student-modelling component of AUTO-COLLEAGUE includes characteristics related to the level of expertise of the students based on their performance during solving tests. The difference between them and our system concerns the nature of other stereotypes used related to the personality of the students. As far as we have surveyed, there is no other system to include user models with personality characteristics, except for an approach to user modelling for recommender systems (González et al. 2002). In this approach, values to specific personality characteristics are assigned to users after having them to answer to personality tests/questionnaires and are not evaluated in a computerized way during their stay in the system. The kind of the stereotypes used is the outcome of an empirical study among experienced trainers on the software engineering course and a study on learning styles based on the Five-Factor Model. Another substantial difference is the fact that in AUTO-COLLEAGUE the stereotypes of the students are evaluated in combination with those of their colleagues in order to suggest appropriate collaboration schemes.

Modelling the student using the stereotype-based method means defining the stereotypes, the facets that describe these stereotypes and the triggers. Triggers are comprised of facet-value combinations. They denote the satisfying conditions for a student to belong or not belong to a stereotype. The purpose of this paper is to describe how we defined the triggers used in our system, evaluating the results of a relative empirical study among experienced trainers.

2   Personality Related Stereotypes

The level of expertise of the student is the most commonly used characteristic included in intelligent learning and CSCL environments that need to model the student. However, there are much more characteristics affecting the student during the learning process. These are related to the way the student performs, behaves and thinks. The personality of the student affects these processes. Previous studies have shown how personality influences learning strategies and learning outcome (Heinström 2000). Especially, in our case of creating optimum groups of learners, evaluating the personality of the individual students is crucial as it has a significant impact on the performance of the groups. Various studies have indicated that personality has an effect on team effectiveness […] (Gustavsson and Bäccman 2005). The stereotypes we have included in our implementation are related to the level of expertise of the user, as well as the behaviour during the learning process. The stereotypes concern two aspects of the trainee: the Level of Expertise and the Personality.

The level of expertise stereotypes describe the knowledge level of the student on the domain, which is UML. There are four stereotypes in this category: Basics, Junior, Senior and Expert. Each of these stereotypes represents a specific structure of

knowledge and its degree. This degree is a real number that can get values between 0 and 1, indicating the level of knowledge upon each UML concept. The stereotypes are associated with a subset of the expert's model.

The personality stereotypes are related to the characteristics that influence the user behaviour regarding the way of learning and collaborating with others. There are 8 Personality stereotypes: Self-confident, Diligent, Participative, Willing-to-help, Sceptical, Hurried, Unconcentrated and Efficient. The self-confident user believes in him/herself and his/her skills. When a person has self-confidence, s/he maintains a positive attitude even if his/her knowledge and skills are not of a high level or even if probably s/he actually is not highly esteemed by his/her colleagues. The diligent user has earnest and persistent application to the training task and makes steady efforts during the learning process. The participative user seems to like collaborating with others and has an active presence in the task elaboration. The willing to help user demonstrates good disposal to help his/her colleagues. Sceptical is the user that seems to need more than the average time to process the data of the problem. The sceptical user tends to make unreasonable mistakes and (relatively to his/her knowledge) the progress in his/her skills could have been faster. The hurried user usually submits the answers to problems quickly without examining their correctness and effectiveness. This results to frequent errors. The unconcentrated user seems to lose his/her abstraction during the training task and perhaps is engaged with other irrelevant tasks at the same time. This kind of characteristic leads to frequent errors and increase in the average time needed to complete a task. The efficient user appears to successfully fulfill the demands of the training task. This kind of user always submits correct solutions/diagrams after a usual or even lower than the usual average amount of time.

The facets (the attributes used to describe the stereotypes) used in AUTO-COLLEAGUE are: useless mouse movements and clicks frequency, average idle time, number of actions, error frequency, correct frequency, help utilization frequency, advice given frequency, help given to a member/non member of the group, help request from a member/non member of the group, communication frequency and number of upgrades/downgrades in level of expertise.

3   Empirical Study for Defining the Triggers

A system that is going to use stereotypes must also know about a set of *triggers* - those events whose occurrence signals the appropriateness of particular stereotypes (Rich 1979). A trigger is a set of rules/conditions. If these conditions are satisfied/dissatisfied for a user, then the corresponding stereotype will be activated/deactivated. These conditions examine the values of the facets used in the system. In order to define the triggers used in our system we conducted an empirical study with experienced trainers.

We have chosen 22 experienced software engineering trainers to participate in the empirical study. The 9 of these trainers were project managers of 2 software houses that, among other duties, they train software engineers in the Borland Delphi programming language, the C# (in Microsoft Visual Studio) and MS SQL Server. There were, also, 8 teachers in the last grade of greek high schools that teach the software engineering course and the rest 5 trainers were scientific collaborators of the University of Piraeus, experienced in training students in UML and programming languages.

The aim of this empirical study was to decide which are the conditions of the facet values that would trigger each stereotype (concerning the stereotypes of Personality, not the Level of Expertise). Therefore, the trainers were asked to fill in a table with the combinations of facet values for every stereotype. For example, in table 1 the answers of a trainer for the Hurried stereotype are shown. F1 is the useless mouse movements and clicks frequency, F2 the average idle time, F3 the number of actions, F4 the error frequency, F5 the correct frequency, F6 the help utilization frequency, F7 the advice given frequency, F8 the help given to a member of the group, F9 the help given to a non-member of the group, F10 the help request from a member of the group, F11 the help request from a non-member of the group, F12 the communication frequency, F13 the number of upgrades in level of expertise and F14 the number of downgrades in level of expertise. In this way, she states that the conditions that should

Table 1. Example of answers of a trainer about the triggers

VALUE	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14
HURRIED														
No							X	X	X	X	X	X	X	X
Low		X			X	X								
Medium	X													
High			X	X										

Table 2. Results showing the rates of the selected facets per stereotype

FACET	NUMBERS OF SELECTION (%)							
	S1	S2	S3	S4	S5	S6	S7	S8
F1	100	100	100	5	9	14		
F2	100	100	100	100	100	100		
F3	100	95	100	100		100		14
F4	86	91	100	100				
F5	86	95	100	100				
F6	91	86	77	9	14	5		
F7					100			100
F8					100		100	100
F9					100		100	100
F10		9	18		100		100	
F11		9	18		100		100	
F12	95		95		5	5	100	9
F13			64	100		100		
F14			14	100		100		

be satisfied for a user to belong to this stereotype would be: the value of the useless mouse movements and clicks frequency is medium, the value of the average idle time is low, the value of the number of actions is high, the value of the error frequency is high, the value of the correct frequency is low and the value of the help utilization frequency is low.

The results of this study were evaluated in two levels. At the first level we concluded to the facets we should include in the triggering conditions for each stereotype. We calculated the rates of the selected facets per stereotype and decided to include only those that had a percentage of selection over 15%. These results are shown in table 2.

At the second level, we calculated the average values of the selected facets with a percentage over 15%. These results are shown in table 3 (L stands for low, M stands for medium and H stands for high) and constitute the final results of the evaluation of the empirical study. These are the conditions of facets values per stereotype that form the main triggers of the stereotypes.

Table 3. Results showing the average facet values per stereotype

FACET	AVERAGE FACET VALUES							
	S1	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**
F1	L	M	H					
F2	M	L	H	L	L	L		
F3	M	H	L	M		H		
F4	M	H	H	L				
F5	M	L	L	H				
F6	M	L	L					
F7					M			M
F8					M		M	M
F9					M		M	M
F10			L		L		M	
F11			L		L		M	
F12	L		H				M	
F13			L	H		M		
F14				L		L		

4   Implementation of Triggers

The Triggers in AUTO-COLLEAGUE are implemented in a Rule-Based subsystem (Ligeza 2006). We found it appropriate to use this technique as the triggering conditions do not constitute a large problem area and can be written in the if-then structure. Moreover, the rule-based technique meets our need to use a structure that could be easily updated even at runtime by the trainer of the system.

The basic modules of a Rule-Based system are the rule-base, the working memory and the inference engine. The rule-base stores all the triggering conditions cited in table

3 in if-then structure. The condition is written in the "If" section and the stereotype to be triggered is written in the "then" section. The working memory is the student data derived from the system's database. This data concerns the student's facet values. The inference engine examines the facet values of the student at a specific moment and checks if these match with any of the conditions of the rules stored in the rule base. If it finds any matching rule, the inference engine fires/triggers the stereotype of the corresponding condition. The inference method used is forward chaining. This means that the inference engine will continue to search for matching conditions, taking into consideration the new data derived from previous triggered rules-stereotypes.

The use of the forward chaining inference method was necessary for the Triggering subsystem. The main triggers are those derived from the empirical study explained in the previous section. However, there are other triggers too that do not concern only the facet values, but also the causality of the activation of another stereotype. For example, there are 2 triggers for the *Efficient* stereotype. The first trigger is the one described in table 3 and related to facet values. The second trigger is related to the inference that a student may belong to Efficient stereotype because s/he already belongs to the *Expert* stereotype.

The triggers for the Level of Expertise stereotypes are implemented in the same subsystem in the same way. The only difference is the parameters of the condition sections. They do not include the facets discussed in the previous section, but the knowledge level on the domain that is derived from the level of expertise student model.

5   Conclusion

One of the most determinative factors in designing an effective adaptive and intelligent learning environment is building accurate student models that would describe the students in an extended range, as extended is the human nature. Therefore, our system is not limited to representing only the obvious aspects of the student, such as knowledge and participation in collaborative activities. In addition to them, the system infers personality characteristics that influence the students' behaviour during learning and collaboration. The stereotype-based student modelling method used for building the personality-related student models involves defining appropriate triggers for each of the defined stereotypes. Aiming at maximizing the accuracy and effectiveness of the extracted student models, we conducted an empirical study for defining these triggers. This empirical study is presented in this paper, along with some details on how the personality-related stereotypes were implemented.

References

Ardissono, L., Gena, C., Torasso, P., Bellifemine, F., Difino, A., Negro, B.: User modeling and recommendation techniques for electronic program guides. In: Ardissono, L., Kobsa, A., Maybury, M. (eds.) Personalized Digital Television. Targeting Programs to Individual Users. Kluwer Academic Publishers, Dordrecht (2004)

Baghaei, N., Mitrović, A.: COLLECT-UML: Supporting Individual and Collaborative Learning of UML Class Diagrams in a Constraint-Based Intelligent Tutoring System. In: Khosla, R., Howlett, R.J., Jain, L.C. (eds.) KES 2005. LNCS (LNAI), vol. 3684, pp. 458–464. Springer, Heidelberg (2005)

Chen, W., Pedersen, R.H., Pettersen, Ø.: CoLeMo: A collaborative learning environment for UML modelling. Interactive Learning Environments 14(3), 233–249 (2006)

Cohen, E.G., Celeste, M., Brody, S.-S.M.: Teaching Cooperative Learning: The Challenge for Teacher Education. State University of New York Press (2004)

Edwards, E., Elliott, J., Bruckman, A.: AquaMOOSE 3D: math learning in a 3D multi-user virtual world. In: CHI 2001 Extended Abstracts on Human Factors in Computing Systems, Seattle, Washington, March 31 - April 05, pp. 259–260. ACM (2001)

Elsom-Cook, M.: Student Modelling in Intelligent Tutoring Systems. Artificial Intelligence Review 7, 227–240 (1993)

Faraco, R.A., Rosatelli, M.C., Gauthier, F.A.O.: An Approach of Student Modelling in a Learning Companion System. In: Lemaître, C., Reyes, C.A., González, J.A. (eds.) IBERAMIA 2004. LNCS (LNAI), vol. 3315, pp. 891–900. Springer, Heidelberg (2004)

Gokhale, A.: Collaborative Learning Enhances Critical Thinking. Journal of Technology Education 7(7) (1995)

González, G., López, B., de la Rosa, J.L.: The Emotional Factor: An Innovative Approach to User Modelling for Recommender Systems. In: Workshop on Recommendation and Personalization in e-Commerce, Málaga, Spain, pp. 90–99 (May 2002)

Goren-Bar, D.: Designing model-based intelligent dialogue systems. In: Information Modeling in the New Millennium, pp. 268–284. IGI Publishing, Hershey (2001)

Gustavsson, B., Bäccman, C.: eam-personality: How to use relevant measurements to predict team-performance. Paper presented at the 47th International Military Testing Association, Singapore, November 8-10 (2005)

Hatzilygeroudis, I., Prentzas, J.: Using a Hybrid Rule-Based Approach in Developing an Intelligent Tutoring System with Knowledge Acquisition and Update Capabilities. Journal of Expert Systems with Applications 26(4), 477–492 (2004)

Heinström, J.: The impact of personality and approaches to learning on information behaviour. Information Research 5(3) (2000)

Jeremic, Z., Jovanovic, J., Gasevic, D.: Evaluating an Intelligent Tutoring System for Design Patterns: the DEPTHS Experience. Educational Technology & Society 12(2), 111–130 (2009)

Kabassi, K., Virvou, M., Tsihrintzis, G.: Requirements Capture for a Personalised Medical Tutor. In: The International Special Topic Conference on Information Technology in Biomedicine (2006)

Kavcic, A.: Fuzzy User Modeling for Adaptation in Educational Hypermedia. IEEE Transactions on Systems, Man, and Cybernetics - part C: Applications and Reviews 34(4), 439–449 (2004)

Khandaker, N., Soh, L.-K., Jiang, H.: Student Learning and Team Formation in a Structured CSCL Environment. In: Proc. ICCE 2006, Beijing, China, pp. 185–192 (2006)

Lehtinen, E., Hakkarainen, K., Lipponen, L., Rahikainen, M., Muukkonen, H.: Computer supported collaborative learning: A review. The J.H.G.I. Giesbers Reports on Education, Number 10. Department of Educational Sciences. University on Nijmegen (1999)

Ligeza, A.: Logical Foundations for Rule-Based Systems. Springer, Heidelberg (2006)

Lukosch, S., Hellweg, M., Rasel, M.: CSCL, Anywhere and Anytime. In: Dimitriadis, Y.A., Zigurs, I., Gómez-Sánchez, E. (eds.) CRIWG 2006. LNCS, vol. 4154, pp. 326–340. Springer, Heidelberg (2006)

Martínez Carreras, M.A., Gómez-Skarmeta, A.F., Martínez Graciá, E., Mora Gónzalez, M.: COLAB: A platform design for collaborative learning in virtual laboratories. In: Workshop held on the 18th IFIP World Computer Congress (2004)

Reye, J.: Student Modelling Based on Belief Networks. Int. J. Artif. Intell. Ed. 14(1), 63–96 (2004)

Rich, E.: User Modeling via Stereotypes. International Journal of Cognitive Science 3, 329–354 (1979)

Rich, E.: Users are individuals: Individualizing user models. Journal of Man-machine Studies 18(3), 199–214 (1983)

Rick, J., Guzdial, M.: Situating CoWeb: a scholarship of application. International Journal of Computer-Supported Collaborative Learning 1, 89–115 (2006)

Rosatelli, M.C., Self, J.: A Collaborative Case Study System For Distance Learning. International Journal of Artificial Intelligence in Education 14, 1–29 (2004)

Surjono, H.D., Maltby, J.R.: Adaptive educational hypermedia based on multiple student characteristics. In: Zhou, W., Nicholson, P., Corbitt, B., Fong, J. (eds.) ICWL 2003. LNCS, vol. 2783, pp. 442–449. Springer, Heidelberg (2003)

Tamura, Y., Furukawa, S.: CSCL Environment for "Six Thinking Hats" Discussion. In: Knowledge-Based Intelligent Information and Engineering Systems, pp. 583–589 (2008)

Tsiriga, V., Virvou, M.: Dynamically Initializing the Student Model in a Web-based Language Tutor. In: Proceedings of the 2002 First International IEEE Symposium "Intelligent Systems", pp. 138–143. IEEE Computer Society, Los Alamitos (2002)

Virvou, M., Moundridou, M.: Student and instructor models: Two kinds of user model and their interaction in an ITS authoring tool. In: Bauer, M., Gmytrasiewicz, P.J., Vassileva, J. (eds.) UM 2001. LNCS (LNAI), vol. 2109, p. 158. Springer, Heidelberg (2001)

Virvou, M., Tsiriga, V.: Student Modelling in a Web-based Algebra Tutor. In: Proceedings of TELEMATICA 2001 International Conference on Telematics and Web-Based Education, pp. 43–44 (2001)

Wei, F., Moritz, S.H., Parvez, S.M., Blank, G.D.: A student model for object-oriented design and programming. J. Comput. Small Coll. 20(5), 260–273 (2005)

An Efficient Parallel Architecture for H.264/AVC Fractional Motion Estimation

Zhuo Zhao[1] and Ping Liang[2]

[1] University of California, Riverside
zhaozhuo@ee.ucr.edu
[2] University of California, Riverside
liang@ee.ucr.edu

Abstract. This paper presents a new VLSI architecture for fractional motion estimation (FME) in H.264/AVC. Statistical characteristics of the motion vectors of different inter-prediction modes are analyzed. The FME architecture explored block-level parallelism and can process multiple blocks with the same prediction mode simultaneously, external memory data accesses are dramatically reduced. Simulation results show that the proposed architecture can support 1080P (1960x1088) at 30fps with a frequency of 80MHz.

1 Introduction

As the newest international video coding standard, H.264/AVC, provides much better image quality and compression ratio, compared with previous standards. This mainly comes from many new techniques, such as variable block size motion estimation (VB-SME), pixel fractional motion estimation, multiple reference frame (MRF), in-the-loop deblocking filter and so on.

However, the intensive computation of H.264/AVC is the major obstacle for real-time encoding systems, especially consider that H.264/AVC requires almost 10 times the computation of previous coding standards. In H.264/AVC hardware encoder[1], the ME consists of 2 stages, the integer motion estimation (IME) for 7 block modes, and fractional motion estimation (FME) with 1/2 and 1/4 pixel accuracy.

In the IME stage, all the 7 block modes are processed in parallel with the SAD-reusing technique, and the integer motion vectors (IMV) of the 41 sub-blocks are then transferred to FME. In the FME stage, because these 41 sub-blocks have different IMVs, all the sub-blocks should be separately processed. Moreover, because 4x4 block size is the smallest unit for all the block modes, usually, a 4x4 based PU is adopted and all bigger blocks were divided into 4x4 to fully utilize the hardware, e.g. see [4]. Then the FME compuatation is in direct proportion to the number of block modes. For H.264/AVC with 7 block modes and a single reference frame, there are 7x16 = 112 4x4 blocks. Lack of block level parallelism, blocks like 16x8, 8x16, 8x8, etc, have to be processed sequentially, e.g. [2], [4] and [5]. To overcome this difficulty, some fast algorithms are brought forward, e.g. [3], bilinear filters are used to replace 6-tap FIR filters, and only single-iteration FME is needed. However, this inevitablly introduced prediction error propagation. Considering other factors, like irregular memory accessing, obtaining FME from IME becomes the bottleneck for the whole encoding system.

In order to improve the FME processing speed and keep the same encoding performance, in this paper, we explored block-level parallelism and a hardware architecture for parallel processing of two blocks (equal or greater than 8x8) is proposed. Also, by adopting a shifting register array, intensive access to reference memory is significantly alleviated.

2   H.264/AVC FME Observations

2.1   Encoding with INTER8x8 Mode or above

Our experiments show that for natural images, modes above 8x8 dominate the final MB modes. Moreover, experiments have shown that the image quality even has slightly better without the small block modes. Fig.1 shows a case with high motion intensities and complex content details. Similar observations can also be found in [2]. Therefore, in our realized FME engine, the block modes below 8x8 are all removed, which saves about 40% of the compuatational complexity.
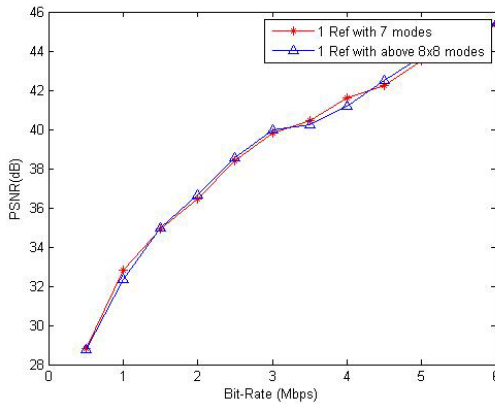
Fig. 1. Mode statistics for STEPHEN.SIF

2.2   Statistic Charactistics of Motion Vectors

For FME, the fact that different blocks may have different IMVs prevents the encoder from parallel processing. The relative position of neighboring blocks is not fixed, which adds difficulties to multi-FME engines data fetching.

However, natural images are likely to have high correlations among neighboring motion vectors (MV). Fig.2 is one of our experiment results. Delta MVs are the absolute difference values between the motion vectors of neighboring blocks. Eqs (1) and (2) are the formula for calculating delta MVs for mode 16x8 and mode 8x16. Delta MVs for mode 8x8 is more complicated. For that mode, neighboring blocks include both horizontal and vertical neighbors. With the reference codec, JM16, we did experiments
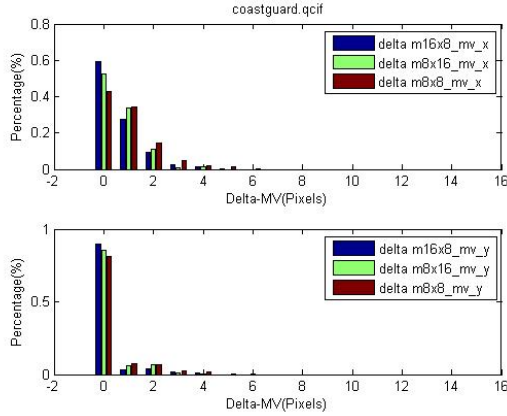
Fig. 2. IMV Statistics for COASTGUARD.QCIF

for most of the test medias in [6], we found more than 90% of the delta MVs are less than 4. This is also the motivation of our FME architecture.

$$dmv\_x = |blk0\_mv\_x - blk1\_mv\_x| \tag{1}$$

$$dmv\_y = |blk0\_mv\_y - blk1\_mv\_y| \tag{2}$$

In the equations above, dmv_x/y stands for motion vector differences for x or y components, $blk_i\_mv\_x/y$ refers to the motion vectors for $block_i$.

3   The Proposed Architecture

For H.264/AVC FME design, three important issues need to be considered. The first is the data fetching, accesses to external memories need to be minimized. The second is the local memory size for FME, it should be minimized because abour 80% of the chip area is occupied by on-chip memories in an ASIC. The third is the data sharing and parallelism, which determines the FME processing speed and system frequency. The proposed architecture improves on all three issues over prior art designs.

Fig. 3 shows our proposed FME system architecture. It consists of a control FSM, a reference pixel array, a sampler line mux, a current MB pixel array, two FME engines for interpolation and cost calculation, a unit for comparator and output buffer for motion compensation (MC). The FME Control FSM gets optimal IMVs and current Ref-Array locations from the IME engine, it controls the shifting directions of both the Ref-Array and the Cur-MB Array, also the cooperation of all the internal FME sub-modules described below.

IMVs for all supported modes,
current position of Ref-Array, etc
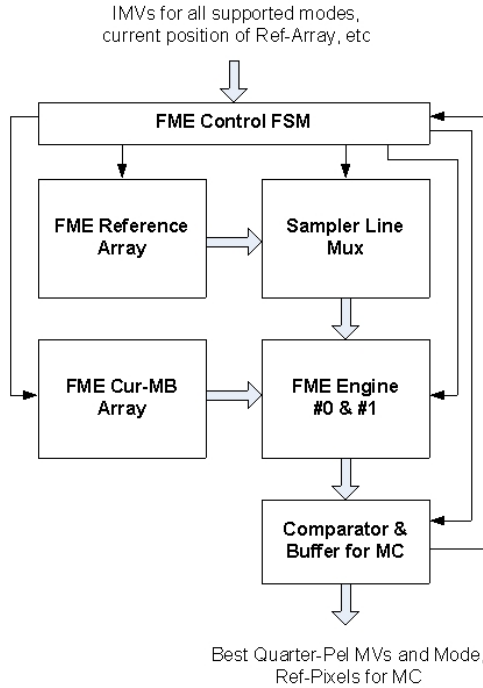
Fig. 3. FME System Architecture

3.1   Reference Pixel Array

In our design, a 32x32 pixels array is used for the reference window. This array was used by IME before it goes to FME. Two such arrays are used to enable parallel ping-pong processing between IME and FME.

As shown in Fig. 4, the reference array can shift in both horizontal and vertical directions. Each cell consists of an 8-bit flip-flop and a 4-to-1 mux. The mux is programmed by the FME control FSM, determining the shifting direction of the reference array.

This reference array is fetched by the IME engine and is handed over to FME engine once IME is done. So, the reference data is only loaded once and used by both IME and FME. This reduces intensive memory access and irregular memory addressing issues.

3.2   Integer Pixel Sampler in Reference Array

In the reference array, we have a horizontal and a vertical sampler lines, both with 26 samplers. The positions of these two sampler lines are fixed in the array. Only pixels in these lines can be sampled and used for FME.

At any time, only one of these two sampler lines is active. That is, when the array is shifting horizontally, the vertical sampler line will be used; and when the array is shifting vertically, the horizontal sampler line will be used.

As shown in Fig. 5, 2-to-1 muxs are used to select either the horizontal sampler line or the vertical sampler line. Muxs are configured by the FME control FSM.
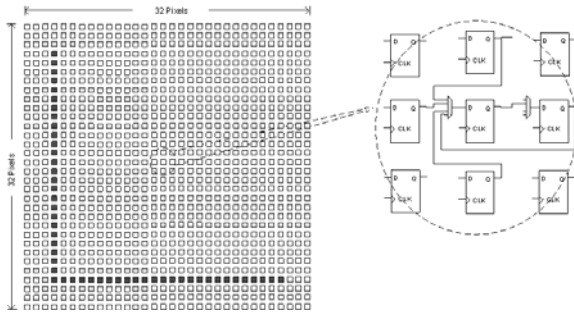
Fig. 4. FME Reference Array
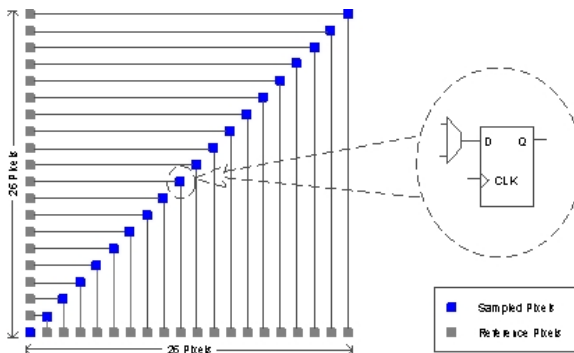
Fig. 5. Sampler Line Mux

3.3   14-Input FME Engine

The two 14-Input (Ref-pixels) FME Engine working in parallel are the key parts of the proposed architecture. As show in Fig.6, these colored input units are from the mux outputs in Fig.5.

The FME engines make use of the statistical characteristics of IMVs which we discussed above. 14 8-to-1 muxs are adopted for the inputs of the second FME engine. So, the delta IMVs between neighboring blocks do not have to be zero, they can be any values from 0 to 4, which counts more than 90% of the cases. For INTER8x8 mode, the situation is a little bit more complicated, but we still can benefit form this architecture.

Inside every FME engine, an interpolation engine shown as Fig.7 is used. It takes 14 integer reference pixels as inputs, and in every clock cycle, it will produce 8 half-pels around every selected 8 integer pixels pointed by IMVs, or 8 quarter-pels around selected 8 half-pels pointed by half-pel motion vectors (bilinear filters were not drawn in Fig.7).

Once we picked a sampler line, the input pixels to FME engine 0 are always from the fixed locations (pixel_0 to pixel_13). On the other side, we used 14 8-to-1 muxs for FME
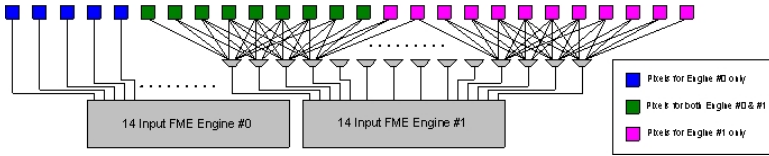
Fig. 6. Two 14-Input FME Engine
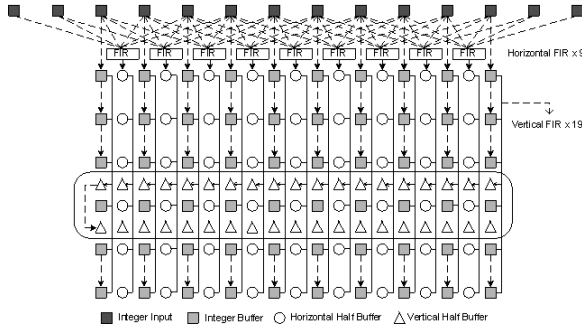
Fig. 7. Interpolation Unit inside FME Engine

engine 1 to select input pixels from 21 locations. The usage of muxs is to accommodate the different motion vectors among neighboring blocks.

Briefly, the tasks for these 2 FME engines include: half-pel interpolation and motion search among 8 half-pel positions, quarter-pel interpolation and motion search among 8 quarter-pel positions, dump reference pixels (integer-pel, half-pel or quarter-pel) to MC stage. It will output SATD values and corresponding motion vectors for different prediction modes, from INTER16x16 to INTER8x8.

3.4   Data Processing Order

The FME will always start from INTER8x8 mode, since it can be processed either horizontally or vertically. As shown in Fig.8 and Fig.9, there are two data processing orders. Depending on the MV distribution of 4 blocks in INTRA8x8 mode, FME control unit will compute which direction will consume less processing cycles before the Ref-Array shifting starts.

In Fig.8 and Fig.9, the blue bars stand for the active sampler lines. Once the FME control FSM decides to shift 4 INTER8x8 reference blocks left, the data processing order will be like Fig.8, followed by INTER16x8 mode, INTER8x16 mode and IN-TER16x16 mode. If the FME control FSM decides to shift 4 INTER8x8 refrence blockd down, the data flow will be like Fig.9, followed by INTER8x16 mode, INTER16x8 mode and INTER16x16 mode.

In the best case, FME engine 0 and FME engine 1 can work simutaneously. It can process block0 and block 1 in INTER16x8 mode, INTER8x16 mode, two half parts in
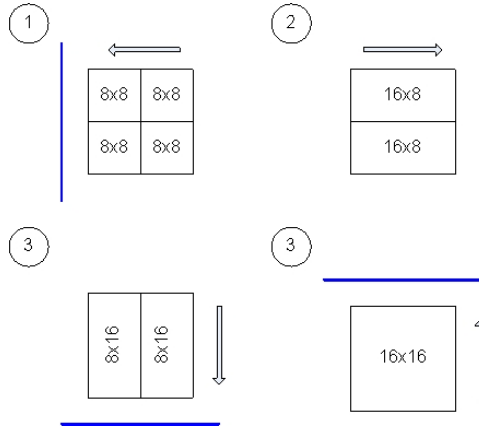
Fig. 8. FME Processing Flow Option 0
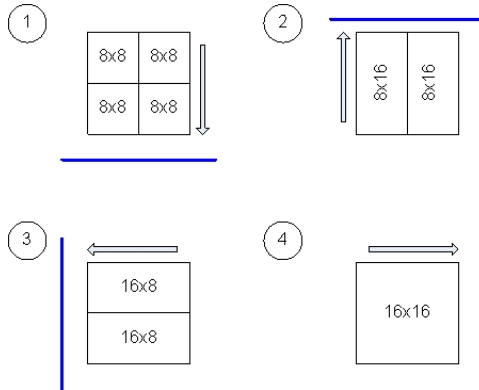
Fig. 9. FME Processing Flow Option 1

INTER16x16 mode or neighboring 8x8 blocks in INTER8x8 modes. However, the parallel processing has to satisfy certain conditions. For INTER8x16 mode, Eq.(3) should be satisfied and for INTER16x8 mode, Eq. (4) should be satisfied, where *MUX_DIST* is 4 in our experiments. INTER8x8 mode has a much more complicated algorithm, which will not be discussed in here.

$$|blk0\_mv\_x - blk1\_mv\_x| \leq MUX\_DIST \tag{3}$$

$$|blk0\_mv\_y - blk1\_mv\_y| \leq MUX\_DIST \tag{4}$$

In the equations above, $blk_i\_mv\_x/y$ refers to the motion vectors for $block_i$, *MUX_DIST* refers to the maximal differences of motion vectors between neighboring blocks.

If the current processing mode does not meet the parallel processing requirement, a *miss* penalty will be applied. For example, if INTER8x16 mode does not satisfy Eq.(3), then after it finishes FME of block 0 in engine 0, it has to scan back for block 1 and shifts it to the desired position for FME processing in engine 1. In this scenario, a certain number of bubble cycles are generated.

3.5   3-Stages Processing

The interpolation structure in Fig. 7 can generate both half-pels and quarter-pels for FME. However, if we do both of them in a single iteration, a lot of memory cells will be used to store these sub-pels. In order to save expensive on-chip memory space, a hierarchical processing flow consisting of 3 stages is adopted in our design.

In the first stage, half-pel FME will be conducted as stated in section D. The best half-pel MVs (HMV) will be generated. In the second stage, quarter-pel FME is conducted and the best quarter-pel MVs (QMV) and the best INTER mode is generated. In the third stage, the reference pixels based on best mode and best QMV is output to MC buffer.

4   Simulation Results

The proposed FME architecture and data processing order have been simulated in a SystemC implementation. Test videos with different content details and motion intensities have been tested. In most of the experimental results, the average FME processing periods are around 200 clock cycles as Table. 1 shows. Fig. 10 shows the possibility distributions of number of processing cycles per MB for several video sequences with different motion intensity and content details. Our simulation also shows that for almost all the test videos in [6] (up to 1080P 30fps), a clock frequency of 80MHz will be sufficient.

Table 1. Part of the Simulation Results

Video Name	amp; Motion Intensity & Content Detail	amp; Avg Cycles per MB
Container	amp; Less	amp; 198
Coastguard	amp; Median	amp; 198
Foreman	amp; Median	amp; 226
Highway	amp; High	amp; 225
Stephen	amp; Very High	amp; 319
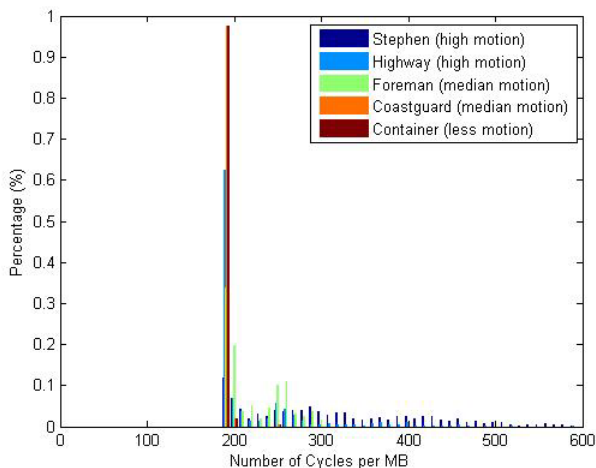
Fig. 10. Processing Cycles Distributions

5   Conclusions

In this paper, we proposed a new simplified FME architecture with a shifting reference array, which explores block-level parallelism for processing of multiple blocks with different IMVs simultaneously. With this architecture, FME processing throughput in a large number of practical applications was dramatically improved and intensive memory accesses were avoided. Simulation shows an 80MHz processing frequency for most video contents up to 1080P 30fps.

References

1. Chen, T.C., Chien, S.Y., et al.: Analysis and architecture design of HDTV720P 30 frams/s H.264/AVC encoder. IEEE Trans. Circuits Systems. Video Technology 16(6), 673–688 (2006)
2. Song, Y., Shao, M., et al.: H.264/AVC Fractional Motion E stimation Engine with ComputationReusing in HDTV1080P Real-Time Encoding Applications. In: 2007 IEEEWorkshop on Signal Processing Systems, October 17-19, pp. 509–514 (2007)
3. Tsung, P.K., Chen, W.Y., Ding, L.F., et al.: Single-iteration full-search fractional motion estimationfor quad full HD H.264/AVC encoding. In: ICME 2009, pp. 9–12 (2009)
4. Chen, T.C., Huang, Y.W., Chen, L.G.: Fully utilized and reusable architecture for fractional motion estimation of H.264/AVC. In: ICASSP 2004, May 17-21, vol. 5 (2004)
5. Zhang, L., Gao, W.: Reusable Architecture and Complexity-Controllable Algorithm for the Int eger/Fractional Motion Estimation of H.264. IEEE Transactions on Consumer Electronics 53(2), 749–756 (2007)
6. Xiph.org Test Media, http://media.xiph.org/video/derf/

Fast Two-Stage Global Motion Estimation:
A Blocks and Pixels Sampling Approach

Adel Ahmadi[1], Farhad Pouladi[2], Hojjat Salehinejad[3], and Siamak Talebi[1,3]

[1] Electrical Engineering Department, Shahid Bahonar University of Kerman, Iran
[2] Scientific-Applied Faculty of Post & Telecommunications Ministry of Information
and Communication Technology, Tehran, Iran
[3] Advanced Communication Research Institute, Sharif University of Technology, Tehran, Iran
adel.ahmadi@gmail.com, pouladi@ictfaculty.ir,
{h.salehi,siamak.talebi}@mail.uk.ac.ir

Abstract. Global motion estimation (GME) is an important technique in image and video processing. Whereas the direct global motion estimation techniques boast reasonable precision they tend to suffer from high computational complexity. As with indirect methods, though presenting lower computational complexity they mostly exhibit lower accuracy than their direct counterparts. In this paper, the authors introduce a robust algorithm for GME with near identical accuracy and almost 50-times faster than MPEG-4 verification model (VM). This approach entails two stages in which, first, motion vector of sampled block is employed to obtain initial GME then Levenberg-Marquardt algorithm is applied to the sub-sampled pixels to optimize the initial GME values. As will be shown, the proposed solution exhibits remarkable accuracy and speed features with simulation results distinctively bearing them out.

1 Introduction

Motion estimation and compensation is one of the most essential techniques in video compression and processing. Motions in video are categorized into local motion (LM) and global motion (GM) [1]. The LMs are resulted from movement, rotation and re-form of objects, while the GMs are due to movement, rotation, and camera zoom [2]. Global motion estimation (GME) has many applications such as video coding, image stabilization, video object segmentation, virtual reality and etc. In MPEG-4 standard, some techniques such as sprite coding and global motion compensation (GMC) are required for GME [3].

The common GME methods are divided into direct and indirect categories. In the direct category, which is pixel-based, prediction error is minimized by using optimization methods such as Levenberg-Marquardt algorithm (LMA) [1],[2],[4]-[7]. The indirect methods consist of two stages. In the first stage, motion vectors of blocks are calculated and by using these vectors, GM of frame is estimated in the second stage [8]-[14].

In MPEG-4 verification model (VM), GME is a direct type scheme where LMA is applied to the whole frame. Since LMA has high computational complexity, some methods have been devised by considering a limited number of pixels in the calculations. One such

technique is called FFRGMET that is used in MPEG-4 optimizing model. This technique just applies LMA to a number of pixels called feature pixels [15]. In [6], pixels are selected using gradient method. In this work, each frame is divided into 100 blocks and then 10% of pixels with the highest gradient are selected from each block. This procedure requires gradient calculations and pixels arrangement based on the gradients. Therefore, this method has a considerable computational complexity. The idea of random pixels selection is introduced in [16]. In spite of the method presented in [6], this technique has much lower computational complexity. However, random pixel selection causes numerical instabilities. In [4] and [5], pixels are selected based on a static pattern. In these papers, authors divide the frame into non-overlapped blocks and then select a few pixels with static pattern from each block. This method has low computational complexity and also does not cause numerical instabilities. In comparison to MPEG-4 VM, this scheme is faster with little accuracy degradation. An indirect GME for the affine model is proposed in [14]. In this study, firstly the amount of translation is estimated by using integral projection algorithm (IPA) and then based on that information a limited block-matching is performed for each sampled block.

In this paper, we have improved the proposed method in [14] and intend to use the perspective model. This is expected to achieve an improvement of peak signal to noise ratio (PSNR) at low complexity.

The reminder of this paper is organized as follows. The motion models are described in section 2 and in section 3, the proposed method including its different steps are discussed in details. The simulation studies are provided in section 4 and finally the paper is concluded in section 5.

2   Motion Models

The most comprehensive GM model in MPEG-4 is the perspective model. This model encompasses simpler models. This model is defined by:

$$x'_i = \frac{m_1 x_i + m_2 y_i + m_3}{m_7 x_i + m_8 y_i + 1} \tag{1}$$

$$y'_i = \frac{m_4 x_i + m_5 y_i + m_6}{m_7 x_i + m_8 y_i + 1} \tag{2}$$

$$\mathbf{m} = \begin{bmatrix} m_1 & m_2 & \cdots & m_8 \end{bmatrix}^T \tag{3}$$

where $\mathbf{m}$ is GM vector from current frame pixels (x_i, y_i) to reference frame pixels (x'_i, y'_i). This vector consists of translation parameters (m_3 and m_6), rotation and zoom parameters (m_1, m_2, m_4, and m_5), and perspective parameters (m_7 and m_8). Simpler models such as affine (with 6 parameters, $m_7 = m_8 = 0$), Translation-Zoom-Rotation (with 4 parameters, $m_1 = m_5$, $m_2 = -m_4$, $m_7 = m_8 = 0$), Translation-Zoom (with 3 parameters, $m_1 = m_5$, $m_2 = m_4 = m_7 = m_8 = 0$) and Translation (with 2 parameters, $m_1 = m_5 = 1$, $m_2 = m_4 = m_7 = m_8 = 0$) are special cases of perspective model.

3   Global Motion Estimation

The proposed algorithm consists of two stages. The first process calls for a rough estimation of GM. When this is obtained second stage takes place in which the initial estimation has to be optimized with greater precision. Structure of the proposed algorithm is as follows.

Stage I

- Estimating translation between two frames using IPA.
- Sampling blocks from the current frame as in Fig.1. Calculating motion vectors of sampled blocks using block matching (with shifted search centre and small searching range). Excluding 30% of blocks with maximum sum of absolute differences (SAD).
- Estimating eight parameters of GM vector using above motion vectors.

Stage II

- Sampling current frame pixels using 1:12×12 model as in Fig.2-d. Applying LMA to sampled pixels to optimize initially estimated GM of the first stage. The LMA iterations are continued until either of the following conditions is satisfied: reaching 10 iterations or updated term be lower than 0.001 for translational components *and* lower than 0.00001 for other components.

3.1   Initial Translation Estimation

In the first stage of GME, translation components must be estimated. In [1]-[5], a three-step search is used for this purpose. IPA is employed instead of a three-step search in our algorithm, because it is more accurate and robust [14].

To estimate translation between two frames, horizontal and vertical projection vectors are calculated as:

$$IP_k^{horiz}(y) = \frac{1}{M} \sum_{x=1}^{M} F_k(x,y) \tag{4}$$

$$IP_k^{vert}(x) = \frac{1}{N} \sum_{y=1}^{N} F_k(x,y) \tag{5}$$

where F_k denotes luminance of frame k and (M, N) are dimensions of frames. IP_k^{vert} and IP_k^{horiz} are integral projection values of F_k in vertical and horizontal directions respectively. By using the correlation between horizontal and vertical integral projection vectors of F_k and F_{k-1}, a translation value is calculated in vertical and in horizontal directions as below:

$$d_x = \min_{t=\{-s,s\}} \left\{ \sum_{x=1}^{M} (IP_k^{vert}(x) - IP_{k-1}^{vert}(x-t))^2 \right\} \tag{6}$$

$$d_y = \min_{t=\{-s,s\}} \left\{ \sum_{y=1}^{N} (IP_k^{horiz}(y) - IP_{k-1}^{horiz}(y-t))^2 \right\} \quad (7)$$

where (d_x, d_y) is translation of the current frame with respect to previous frame and s is maximum search range. The maximum search range is determined based on the size and contents of the video. To give some examples, $s=8$ for QCIF format and $s=16$ for CIF and SIF formats seems reasonable.
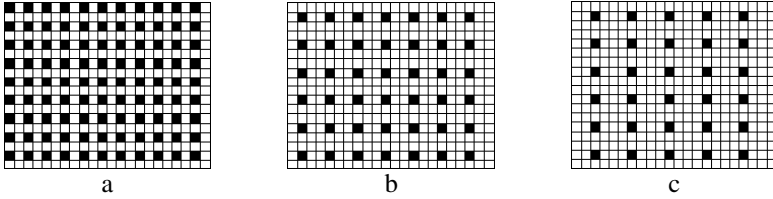
Fig. 1. Blocks sampling pattern [14]; a) 1:4; b) 1:9; c) 30:369

3.2 Block Sampling and Limited Block Matching

After translation estimation, one of the patterns in Fig.1 is employed for blocks sampling. Size of each block for different formats is considered as: 8×8 for QCIF, 16×16 for CIF and SIF and 32×32 for 4CIF. Then for each sampled block, a modified full search block matching algorithm (BMA) is obtained. In this search, the search centre is shifted (d_x, d_y) units and searching range is as small as (-3, +3). This results in less SAD computations and sufficient accuracy for motion vectors of background blocks. Since blocks with high SAD are mostly part of the foreground, 30% of them are excluded. The motion vectors of remaining blocks will be used in the next subsection.

3.3 Initial Estimation of Perspective Model GM Parameters

By considering (x_i, y_i) as central pixel coordinate of the current frame sampled block and (x_i', y_i') as central pixel coordinate of the best matched block, we can have:

$$x_i' = v_{x,i} + x_i \quad (8)$$

$$y_i' = v_{y,i} + y_i \quad (9)$$

where $v_{x,i}$ and $v_{y,i}$ are motion vectors obtained from the previous step.

To find GM between two frames, we must minimize the Euclidean error:

$$E = \sum_{i=1}^{N_b} \left[\left(\frac{m_1 x_i + m_2 y_i + m_3}{m_7 x_i + m_8 y_i + 1} - x_i' \right)^2 + \left(\frac{m_4 x_i + m_5 y_i + m_6}{m_7 x_i + m_8 y_i + 1} - y_i' \right)^2 \right] \quad (10)$$

whrere N_b is number of blocks. Since the perspective model is nonlinear, (10) could be solved by using LMA which results in significant computational complexity. On the other hand, by using algebraic error definition [17], (10) can be modified as:

$$E = \sum_{i=1}^{N_b}\left[\left(\left(\frac{m_1 x_i + m_2 y_i + m_3}{m_7 x_i + m_8 y_i + 1} - x_i'\right)^2 + \left(\frac{m_4 x_i + m_5 y_i + m_6}{m_7 x_i + m_8 y_i + 1} - y_i'\right)^2\right) \times D_i^2\right] \quad (11)$$

where D_i is the denominator of motion model:

$$D_i = (m_7 x_i + m_8 y_i + 1) \quad (12)$$

Therefore, we can simplify (11) as:

$$E = \sum_{i=1}^{N_b}\left[(m_1 x_i + m_2 y_i + m_3 - x_i' D_i)^2 + (m_4 x_i + m_5 y_i + m_6 - y_i' D_i)^2\right] \quad (13)$$

At this stage, we can minimize (11) by solving $\partial E / \partial m_j = 0$ and arriving at:

$$\left(\sum_{i=1}^{N_b} \mathbf{A}_i\right)\mathbf{m} = \sum_{i=1}^{N_b} \mathbf{b}_i \quad (14)$$

where $\mathbf{m}$ is GM vector. The $\mathbf{A}_i$ matrix and $\mathbf{b}_i$ vector are defined as:

$$\mathbf{A}_i = \begin{bmatrix} x_i^2 & x_i y_i & x_i & 0 & 0 & 0 & -x_i^2 x_i' & -x_i y_i x_i' \\ x_i y_i & y_i^2 & y_i & 0 & 0 & 0 & -x_i y_i x_i' & -y_i^2 x_i' \\ x_i & y_i & 1 & 0 & 0 & 0 & -x_i x_i' & -y_i x_i' \\ 0 & 0 & 0 & x_i^2 & x_i y_i & x_i & -x_i^2 y_i' & -x_i y_i y_i' \\ 0 & 0 & 0 & x_i y_i & y_i^2 & y_i & -x_i y_i y_i' & -y_i^2 y_i' \\ 0 & 0 & 0 & x_i & y_i & 1 & -x_i y_i' & -y_i y_i' \\ x_i^2 x_i' & x_i y_i x_i' & x_i x_i' & x_i^2 y_i' & x_i y_i y_i' & x_i y_i' & -x_i^2 s_i' & -x_i y_i s_i' \\ x_i y_i x_i' & y_i^2 x_i' & y_i x_i' & x_i y_i y_i' & y_i^2 y_i' & y_i y_i' & -x_i y_i s_i' & -y_i^2 s_i' \end{bmatrix} \quad (15)$$

$$\mathbf{b}_i = \begin{bmatrix} x_i x_i' & y_i x_i' & x_i' & x_i y_i' & y_i y_i' & y_i' & x_i s_i' & y_i s_i' \end{bmatrix}^T \quad (16)$$

$$s_i' = x_i'^2 + y_i'^2. \quad (17)$$

3.4 Subsampling Pixels and Levenberg-Marquardt Algorithm

In this stage, the estimated GM from the previous stage is optimized with greater accuracy by employing LMA. In this paper, we suggest subsampling from all pixels of current frame with a static pattern as in [4], instead of just selecting feature pixels

a   b   c
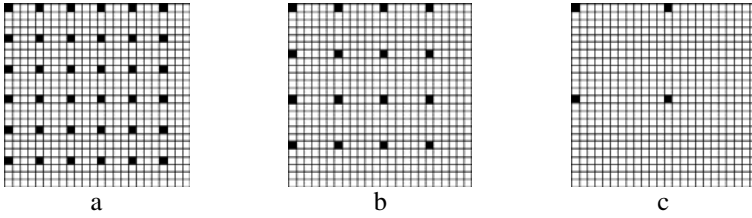
Fig. 2. Pixels subsampling pattern; a) 1:4×4; b) 1:6×6; c) 1:12×12.

among the remaining blocks as in [14]. This selection technique poses less computational complexity than [14] and it is more precise.

In this paper, the 1:12×12 sampling method is used which means that we select one pixel from each 12×12 block. After pixels subsampling, initial GM is optimized by applying LMA to these pixels. To reduce outlier effects, 10% of pixels with the most error are discarded after first iteration [4].

4   Simulation

In this section, the proposed method is examined and compared against MPEG-4 VM, [14] and [4] with a sampling factor 1:9×9.The following sequences with CIF format are considered for simulations: Akiyo (300 frames), Bus (150 frames), Carphone (300 frames), Coastguard (300 frames), Foreman (400 frames), Flower (150 frames), Mobile (300 frames), Stefan (300 frames), Tempete (260 frames), and Waterfall (260 frames). The simulations are run on a desktop computer featuring 2.66GHz Core2Quad CPU, 4GB RAM and MS Windows Vista operating system in MATLAB environment.

The GME time of different sequences are presented in Table 1. Judging from the Table, it is seen that the proposed method's GME time is less than that in [4] for most of the sequences. Furthermore, this is almost the same as the GME time in [14] with affine model.

Table 2 compares speed of the proposed method versus other methods in relation to the MPEG-4 VM method with perspective model. As these results illustrate, the proposed technique is 53 times faster than VM with perspective model. This is while the method in [14] is about 43 times faster than VM with affine model and about 60 times faster than VM with perspective model. The Proposed method as well as [4] both work with perspective model whereas [14] only works with affine model.

The PSNR of sequences is calculated by:

$$PSNR = 10\log_{10}\frac{255^2}{MSE} \tag{18}$$

Table 1. GME Time Comparison of 5 Different Methods (Sec.)

Sequence	VM(Pers.)	VM(Aff.)	[4]	[14]	Proposed
Akiyo	433.11	254.25	7.15	7.18	8.45
Bus	232.66	145.86	5.24	3.38	4.05
Carphone	152.68	99.86	4.47	3.77	4.10
Coastguard	436.75	299.95	6.96	6.67	7.75
Foreman	960.57	640.99	12.61	8.89	10.77
Flower	518.55	279.10	7.03	5.48	6.74
Mobile	354.57	222.69	11.12	6.70	8.10
Stephan	297.07	204.22	8.79	6.57	7.80
Tempete	225.25	153.99	7.00	5.64	7.01
Waterfall	345.65	190.19	6.84	6.15	7.27

Table 2. Speed Comparison of the [4] and MPEG-4 VM Perspective GM

Sequence	VM(Pers.)	VM(Aff.)	[4]	[14]	Proposed
Akiyo	1.00	1.70	60.57	60.32	51.26
Bus	1.00	1.60	44.40	68.83	57.45
Carphone	1.00	1.53	34.16	40.50	37.24
Coastguard	1.00	1.46	62.75	65.48	56.35
Foreman	1.00	1.50	76.18	108.05	89.19
Flower	1.00	1.86	73.76	94.63	76.94
Mobile	1.00	1.59	31.89	52.92	43.77
Stephan	1.00	1.45	33.80	45.22	38.09
Tempete	1.00	1.46	32.18	39.94	32.13
Waterfall	1.00	1.82	50.53	56.20	47.54
Avg.	1.00	1.60	50.02	63.21	53.00

Table 3. PSNR Comparison for Different Sequences (dB)

Sequence	VM(Pers.)	VM(Aff.)	[4]	[14]	Proposed
Akiyo	41.010	41.011	41.101	36.301	41.012
Bus	21.687	21.679	21.623	21.805	21.831
Carphone	30.811	30.739	30.403	28.855	29.729
Coastguard	26.376	26.384	26.358	26.242	26.599
Foreman	25.279	25.256	25.289	23.237	25.085
Flower	28.312	28.160	27.884	27.227	27.716
Mobile	25.538	25.495	25.583	25.206	25.581
Stephan	24.494	24.157	22.753	23.591	23.916
Tempete	27.786	27.778	27.726	27.434	27.715
Waterfall	35.675	35.634	35.573	34.918	35.725

Table 4. PSNR Degradation in Respect of MPEG-4 VM Perspective GM

Sequence	VM(Pers.)	VM(Aff.)	[4]	[14]	Proposed
Akiyo	0.000	0.001	0.090	-4.709	0.002
Bus	0.000	-0.008	-0.064	0.118	0.144
Carphone	0.000	-0.072	-0.408	-1.956	-1.082
Coastguard	0.000	0.008	-0.018	-0.135	0.222
Foreman	0.000	-0.152	-0.428	-1.086	-0.597
Flower	0.000	0.022	0.011	-2.042	-0.193
Mobile	0.000	-0.043	0.045	-0.332	0.043
Stephan	0.000	-0.336	-1.740	-0.903	-0.577
Tempete	0.000	-0.009	-0.060	-0.353	-0.071
Waterfall	0.000	-0.041	-0.103	-0.758	0.050
Avg.	0.000	-0.068	-0.268	-1.215	-0.206

where

$$MSE = \frac{1}{MN} \sum_{x=1}^{M} \sum_{y=1}^{N} (F_k(x, y) - F_{k-1}(x', y'))^2 \tag{19}$$

In Table 3, PSNR of GME for each sequence is presented. Table 4 also displays PSNR degradation in respect of VM with perspective motion model. As the results demonstrate, the proposed method has on average reduced the PSNR by -0.2 dB while [14] method degrades the PSNR by -1.2 dB.

5   Conclusion

In this paper a fast two-stage algorithm for global motion estimation (GME) with perspective model is introduced. In the first stage, eight parameters of global motion (GM) are estimated by using sampled motion vectors of blocks. In the second stage, by subsampling of pixels and using Levenberg-Marquardt algorithm (LMA), the estimated GM of the first stage is estimated more accurately.

As the simulation results demonstrate, one key advantage of the proposed solution in this paper is that it is almost 53 times faster than the MPEG-4 VM method. Another outstanding feature of the innovative technique is its enhanced estimation accuracy which is more than FFRGMET's and [4]'s and almost the same as MPEG-4 VM's. Still, when compared against [14], the algorithm exhibits better precision under the same speed. This is while our method works with the perspective model and [14] estimates the simpler affine model.

References

1. Qi, B., Ghazal, M., Amer, A.: Robust Global Motion Estimation Oriented to Video Object Segmentation. IEEE Trans. Image Process. 17(6), 958–967 (2008)
2. Dufaux, F., Konrad, J.: Efficient, robust and fast global motion estimation for video coding. IEEE Trans. Image Process. 9(3), 497–501 (2000)
3. MPEG-4 video verification model version 18.0. In: ISO/IEC JTC1/SC29/WG11 N3908, Pisa, Italy (2001)
4. Alzoubi, H., Pan, W.D.: Fast and Accurate Global Motion Estimation Algorithm Using Pixel Subsampling. Information Sciences 178(17), 3415–3425 (2008)
5. Alzoubi, H., Pan, W.D.: Reducing the complexity of MPEG-4 global motion estimation using pixel subsampling. IET Electronic letters 44(1), 20–21 (2008)
6. Keller, Y., Averbuch, A.: Fast gradient methods based on global motion estimation for video compression. IEEE Transactions on Circuits and Systems for Video Technology 13(4), 300–309 (2003)
7. Chan, W.C., Au, O.C., Fu, M.F.: Improved global motion estimation using prediction and early termination. Proc. IEEE Int. Conf. Image Processing 2, 285–288 (2002)
8. Moscheni, F., Dufaux, F., Kunt, M.: A new two-stage global/local motion estimation based on a background/foreground segmentation. In: Proc. of Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP 1995), Detroit, MI, vol. 4, pp. 2261–2264 (1995)
9. Rath, G.B., Makur, A.: Iterative least squares and compression based estimation for a four-parameter linear motion model and global motion compensation. IEEE Transactions on Circuits & Systems for Video Technology 9(7), 1075–1099 (1999)
10. Xu, G., Ding, M., Cheng, Y., Tian, Y.: Global motion estimation based on kalman predictor. In: Proc. of 2009 IEEE Int. Workshop on Imaging Systems and Techniques, Shenzhen, China, pp. 395–398 (2009)
11. Chung, Y., He, Z.: Reliability Analysis for Global Motion Estimation. IEEE Signal Processing Letters 11(11), 980–997 (2009)
12. Tarannum, N., Pickering, M.R., Frater, M.R.: An automatic and robust approach for global motion estimation. In: Proc. of IEEE Int. Workshop on Multimedia Signal Processing (MMSP 2008), Australia, pp. 83–88 (2008)
13. Shang, F., Yang, G., Yang, H., Tian, D.: Efficient Global Motion Estimation Using Macroblock Pair Vectors. In: Proc. of Int. Conf. on Information Technology and Computer Science (ITCS 2009), Ukraine, vol. 1(1), pp. 225–228 (2009)
14. Lei, L., Zhiliang, W., Jiwei, L., Zhaohui, C.: Fast Global Motion Estimation. In: Proc. of 2nd IEEE Int. Conf. on Broadband Network & Multimedia Technology (IC-BNMT 2009), Beijing, pp. 220–225 (2009)
15. ISO/IEC JTC1/SC29/WG11 MPEG Video Group. Optimization model. ISO/IECJTC1/SC29/WG11N3675 LaBaule, France (2000)
16. Dellaert, F., Collins, R.: Fast image-based tracking by selective pixel integration. In: ICCV Workshop on Frame-Rate Vision, Corfu., Greece (1999)
17. Farin, D.: Automatic Video Segmentation Employing Object/Camera Modeling Techniques. Ph.D. Thesis, Technical University of Eindhoven, pp. 110-114 (2005)

Frame Extraction Based on Displacement Amount for Automatic Comic Generation from Metaverse Museum Visit Log

Ruck Thawonmas and Akira Fukumoto

Intelligent Computer Entertainment Laboratory
Graduate School of Science and Engineering, Ritsumeikan University
Kusatsu, Shiga, 525-8577, Japan
`ruck@ci.ritsumei.ac.jp`

Abstract. The paper describes a system for automatically generating comics from visit log at a metaverse museum. Metaverse is a 3D virtual world in which users can act freely, such as visiting museums or chatting with others, according to their own purposes. Compared with existing approaches for representing user experiences using snapshots or video clips, the comic approach can 1) allow users to grasp the whole story at a glance, 2) facilitate distinguishing of important frames, and 3) exploit varieties of comic writing techniques. In order to summarize user experience into comic's frames, detection of important experience, interesting exhibits in case of the museums, is an important task. In this paper, we propose a method for frame extraction based on the displacement amount of a user of interest at a metaverse museum. After describing each module of our system, we discuss a user evaluation in which the effectiveness of the proposed frame extraction method is confirmed when it is compared with a typical baseline method.

1 Introduction

Virtual 3D space called metaverse has recently gained interests from not only educators but also researchers as a promising educational and research platform. Second Life (SL) is representative metaverse equipped with a function that enables its users to build objects or architectures. Unlike online games, metaverse, in general, has no specific roles assigned to the users. As a result, user experiences are arguably limitless including, for example, visiting to museums built by other users or chatting to other users. Other usages of metaverse include an experiment (Prendinger 2009) that aims at realization of an eco-friendly society as well as Machinima (Lowood 2006) that uses in-game avatars (user characters), objects, and architectures for filming.

A number of metaverse systems have functions that allow users to take snapshots of their experiences or record them into video clips. Such snapshots or video clips are used by the users to recall their memories or shown to other users via, for example, blogs or SNS sites. However, manually taking a snapshot each time imposes a burden to the user. And video clips require considerable time

to grasp the story and distinguish between important and unimportant events therein. In order to cope with these issues, we focus on the use of comics for representing user experiences.

Using the comic style, one can summarize a given user experience into a limited number of frames, provided that a proper summarization mechanism is used. Hence a whole story can be apprehended at one glance. In addition, a frame layout technique can be applied to emphasize important frames and underemphasize unimportant ones. In this paper, we focus on museum visiting, one of the major usages in metaverse, present a system for generating a comic based on the museum visiting experience of a user of interest, and propose a frame extraction method that uses the information on the amount of displacement at a museum of interest.

2   Comic Generation System

We have developed a comic generating system and related methods (Shuda and Thawonmas 2008; Thawonmas and Shuda 2008; Thawonmas and Oda 2010) for on-line games. Our objectives therein are to summarize players' game experiences into comics so that the players can exchange their comics with others and recall their memories about the game. However, an online game, the research target in our previous work, is essentially different from metaverse, which is the research target of this paper.

In a typical online game, players' actions, such as "attack a monster" or "open a treasure" are prepared and provided in advance by the game developer. Such actions are meaningful, can be interpreted as important episodes, and are thus worth memorizing. However, most of the actions in metaverse, such as "move" or "sit" have not much meaning. It is therefore difficult to extract important episodes from log. As far as the museum visit application is concerned, one might come up with a method for extracting comic frames that cover important exhibits specified in advance by the museum's developers. However, this method lacks the universality because it cannot be used at other museums where such information is not available. In addition, the method does not take into account the preference of a user of interest.

According to a study on a real-world museum (Sparacino 2003), the more interest a user has in a given exhibit, the higher probability is that he or she will spend longer time viewing the exhibit. Based on this finding, we propose a frame extraction method that bases on the displacement amount (cf. (1)) to predict the user's interest at a given virtual museum. Namely, the displacement amount of zero implies that the user is stopping at a particular place viewing an exhibit, and the longer stop period, the more interest the user has in the exhibit. The proposed frame extraction method puts higher priorities to such stops.

Figure 1 shows the architecture of the comic generation system, where "visit log" indicates a file that contains information on the user traces (the coordinate (x, y, z) at each loop-count, the number of times in sampling the coordinate), the user actions, and the object positions in the museum. In the following, we describe each of the four modules shown in Fig. 1.
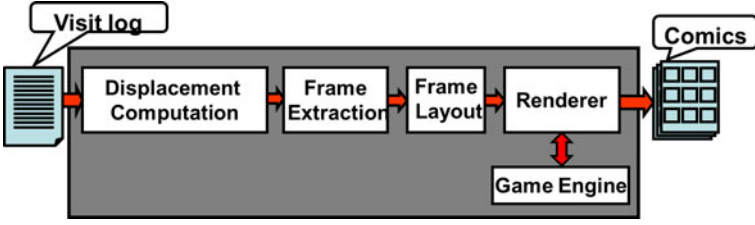
Fig. 1. Architecture of the comic generation system

1. At this module, the amount of displacement for a given user trace at each loop-count l is calculated as follows:

$$\text{displacement}^2 = \big(x(l)-x(l-1)\big)^2 + \big(y(l)-y(l-1)\big)^2 + \big(z(l)-z(l-1)\big)^2 \quad (1)$$

Next, the given user trace are segmented into stop segments or move segments. The stop segments are those whose displacement amount is zero while the remaining segments become the move segments. For segment n, the following pieces of information are stored in Segment(n), i.e., state, length, start, end, sumDisplacement, and count, indicating the type of this segment (stop or move), the length of this segment in terms of the number of loop-counts, the starting loop-count, the ending loop-count, the total amount of displacement, and the number of frames extracted from this segment, respectively. In addition, for each stop-type segment, the information on the stop position, stopPosition, is also stored.

2. At the frame extraction step, the following two conditions are applied to all stop-type segments for extraction of them:

Condition I: Is its stopPosition not adjacent to stopPosition of the previously extracted stop-type segment?

Condition II: Is its length above the average length of all stop-type segments?

The former condition is employed to prevent extraction of frames showing similar user experiences, such as frames showing the user stopping to view the same exhibit. The latter condition is employed to prevent extraction of frames showing spontaneous stops by the user.

For each stop-type segment n satisfying both conditions, its count is incremented and a frame, Frame(f), is added into a frame-candidate list, by which the content of Segment(n) is copied to FrameInfo(f), containing the information about Frame(f), where f represents the frame number. If the number of extracted frames is more than the number of frames specified by the user, frames in the frame-candidate list will be removed in increasing order of FrameInfo(f).$length$ until the specified number of frames, N, is met.

Frame extraction is done for move-type segments only when the number of stop-type frames extracted above is lower than N. In this case, this is done for the move-type segments in decreasing order of sumDisplacement. Namely,

first, a frame, Frame(f), is extracted from the move-type segment with the highest sumDisplacement and added to the frame-candidate list, by which after incrementing count Segment(n) is copied to FrameInfo(f); then the one with the second highest sumDisplacement; and so on. For example, if N is 10, and there exist 7 stop-type segments and 5 move-type segments, then, after extraction of 7 frames from the stop-type segments, 3 frames will be extracted from the move-type segments with the top-three sumDisplacement.

If frame extraction has been done for all move-type segments, but the total number of extracted frames is still lower than N, then another round of frame extraction will be done for the move-type segments. This process is repeated until N is filled. Due to such repetitions, multiple frames might be extracted from a same move-type segment; however, such frames have different snapshot timings as discussed below.

Once N is met, the frame-candidate list will be copied to a frame list, at which the snapshot timing $t(f)$ is determined for being used later in the renderer module. For a stop-type frame, Frame(f), the snapshot timing is as follows:

$$t(f) = \text{FrameInfo}(f).start + 0.5\text{FrameInfo}(f).length \qquad (2)$$

Although any timing in [FrameInfo(f).$start$, FrameInfo(f).end] should produce the similar (not same because the user avatar's face or hands might move) comic frame, we empirically decided to use the above formula.

For a move-type frame, Frame(f), the snapshot timing is as follows:

$$t(f) = \text{FrameInfo}(f).start + \frac{\text{FrameInfo}(f).length}{\text{FrameInfo}(f).count + 1} \qquad (3)$$

3. At the frame layout module, each frame in the frame list is assigned a page and the position therein. This decision is based on the information on N, the page margins, and the number of frames per row and column.

4. At the renderer module, the system scans the visit log from the beginning and renders the snapshot image of the user experience with the snapshot timing $t(f)$ of each frame in the frame list. Each rendered image is then placed on the specific position in the predetermined comic page according to the information stored in the corresponding frame. Finally, all comic pages are outputted as an image file.

3   Evaluation

After implementing the proposed system, we conducted a user evaluation. The objective of this user evaluation is to examine if the proposed system can generate a comic that properly summarizes the user experience viewing exhibits at a metaverse museum.

Fig. 2. The metaverse museum used in this work

3.1   Implementation

We implemented the system by adding the above four modules to the open-source SL client program (SL viewer program). For this work, we adopted a typical comic layout where the order to read is in the raster order, from top-left to bottom-right, and all frames have the same size.

For the evaluation, we targeted user experiences at a SL museum[1] designed and operated by members of Global COE (Center of Exellence) Progam gDigital Humanities Center for Japanese Arts and Cultureh of Ritsumeikan Unviersity. This museum is aimed at a virtual exhibition of Kaga Okunizome Dyeing, kimono and bedding from the Ishikawa region in Japan during the latter part of the Edo period until the beginning of the Showa period. However, we would like to point out that our system is also applicable to other museums in SL as well as other metaverse.

Figure 2 shows the museum building in which 19 exhibits and two posters are located. Our reasons for extracting this museum are that (1) the exhibition therein has a high cultural value because Kaga Okunizome dyeing is famous in Japan and (2) there is no copyright problem because the authors belong also to the aforementioned Global COE. For other representative museums in SL, please refer to the work by Urban et al. (Urban et al. 2007).

---

[1] http://slurl.com/secondlife/rits%20gcoe%20jdh/167/189/22

3.2   Evaluation Outline

We compared comics whose frames were extracted by the proposed frame-extraction method and a baseline frame-extraction method. Each method was implemented in our comic-generation system. The latter method uses a fixed-and-equal interval for extraction of frames in a given comic. We first requested 26 participants, who are undergraduate or graduate students in our department, to watch each of the three video clips, showing typical visits of each of the three visitor types (Sparacino 2003): busy[2], selective[3], and greedy[4]. After watching a video clip, each participant was asked to compare two comics of the same number of frames that were extracted by the two methods from the corresponding visit log used for the video clip, without being told which method was used for a given comic. In particular, each participant was asked to answer the following four questions for each comic in the typical five-level Likert scale:

Q1 Does the comic have the content well representing the video-clip play?
Q2 Does the comic have the minimum amount of unnecessary frames?
Q3 Does the comic well display exhibits?
Q4 Does the comic have a dynamic story?

The participants were also asked to write a short reason behind their rating for each question.

3.3   Results and Discussions

Tables 1, 2, and 3 show the average score of each method for the busy-type, the selective-type, and the greedy-type, respectively. The proposed method outperforms the baseline in all questions, except Q3. Q3 is related to the quality of the camerawork and the relating parameters. Hence, the camerawork of the comic generation system should be improved.

Figure 3 shows the first page of the selective-type comics whose frames were extracted by the proposed method and the baseline method. It can be seen that the latter method extracted a number of similar frames while the former method did not. Similar comments were obtained from participants as their reasons for rating the former method higher than the latter one.

4   Related Work

A storytelling system was implemented that generates slides from images taken at multiple locations on a given map (Fujita & Arikawa 2008). For segmentation of videos, a method (Xu et al. 2009) exists that uses histograms of human motions, but this method aims at fast movements such as sports or dances, not slow movements typically seen in SL avatars while they are visiting museums.

---

[2] http://www.youtube.com/watch_popup?v=ijxwUrHejG8\&vq=medium
[3] http://www.youtube.com/watch_popup?v=H8eUJ6JZGXo\&vq=medium
[4] http://www.youtube.com/watch_popup?v=DtVFf5gAUZI\&vq=medium

Table 1. Average score of each method for the busy type

Question #	amp; Proposed Method	amp; Baseline Method
Q1	amp; 3.38	amp; 3.08
Q2	amp; 3.92	amp; 3.15
Q3	amp; 2.77	amp; 2.77
Q4	amp; 3.73	amp; 3.38

Table 2. Average score of each method for the selective type

Question #	amp; Proposed Method	amp; Baseline Method
Q1	amp; 4.19	amp; 3.35
Q2	amp; 3.73	amp; 2.19
Q3	amp; 3.92	amp; 3.62
Q4	amp; 3.81	amp; 2.38

Table 3. Average score of each method for the greedy type

Question #	amp; Proposed Method	amp; Baseline Method
Q1	amp; 4.23	amp; 3.85
Q2	amp; 3.69	amp; 2.19
Q3	amp; 3.5	amp; 4.04
Q4	amp; 3.58	amp; 2.31

A scripting system (Zhang et al. 2007.) was proposed for automating cinematics and cut-screens that facilitate video-game production processes such as the control of transitions between images, and the annotation of texts and sounds to image backgrounds.

There is a number of existing work in addition to the authors' previous work (Shuda & Thawonmas 2008; Thawonmas & Shuda 2008; Thawonmas and Oda 2010) on comic generation for online games. Such work includes comic generation for online games using a combination of screenshots (Chan et al. 2009), rather than using the targeted game's engine for re-rendering frames as done in our approach, and for first person shooters (Shamir et al. 2006). Comics were also used for summarization of experiences in a conference (Sumi et al. 2002), daily experiences (Cho et al. 2007), and videos or movies (Calic et al. 2007; Hwang et al. 2006; Tobita 2010).

5   Conclusions and Future Work

We proposed and implemented a system for generating comics from user experiences during their museum visits. The system enables extraction of frames that contain interesting exhibits, from the view point of a user of interest. This was

(a) Proposed Method
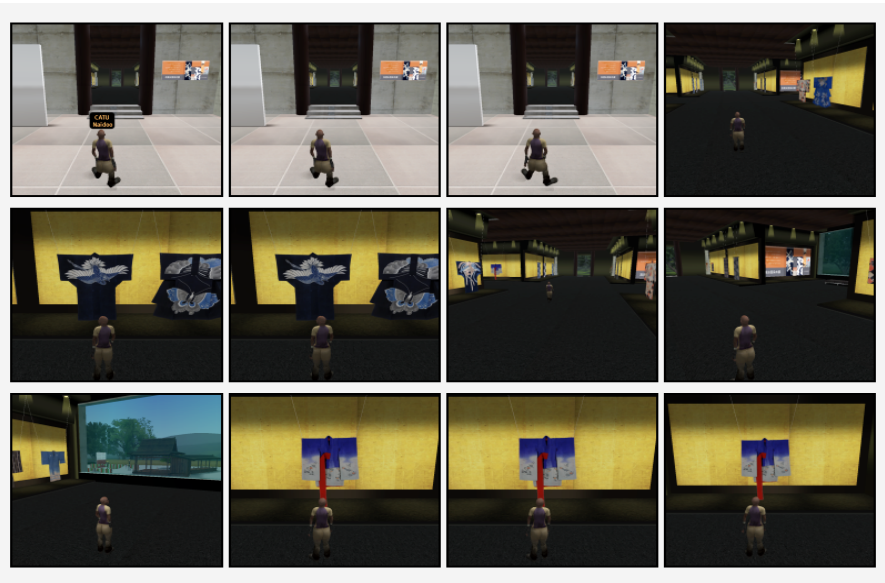
(b) Baseline Method

Fig. 3. Fist page of the generated comics for the selective type

achieved by taking into account the amount of displacement in the museum space and giving higher priorities to stop-type segments having long stop time. The user evaluation confirmed the effectiveness, in properly representing the afore-mentioned user experiences, of the proposed system for all visitor types. As our future work, we plan to increase the expressivity of comics. One possible research theme is to strengthen the comic layout mechanism. This would increase the va-riety of comic frames, such as large frames, small frames, slanted frames, etc. As already stated in 3.3, another theme is to improve the camerawork so that a more proper set of camera parameters, i.e., the camera angle, camera position, and zoom position, is applied to a given frame. The mechanisms proposed in our previous work for online games (Thawonmas and Shuda 2008; Thawonmas and Ko 2010) will be extended, respectively, for these two research themes.

Acknowledgment

References

1. Calic, J., et al.: Efficient Layout of Comic-like Video Summaries. IEEE Transac-tions on Circuits and Systems for Video Technology 17(7), 931–936 (2007)
2. Chan, C.-J., et al.: Automatic Storytelling in Comics: A Case Study on World of Warcraft. In: CHI Extended Abstracts 2009, pp. 3589–3594 (2009)
3. Cho, S.-B., Kim, K.-J., Hwang, K.-S.: Generating Cartoon-Style Summary of Daily Life with Multimedia Mobile Devices. In: Okuno, H.G., Ali, M. (eds.) IEA/AIE 2007. LNCS (LNAI), vol. 4570, pp. 135–144. Springer, Heidelberg (2007)
4. Fujita, H., Arikawa, M.: Animation of Mapped Photo Collections for Storytelling. IEICE Trans. Inf. &; Syst. E91-D(6), 1681–1692 (2008); Special Section/Issue on Human Communication III
5. Hwang, W.-I., et al.: Cinema comics: Cartoon generation from video stream. In: Proc. of GRAPP 2006, pp. 299–304 (2006)
6. Lowood, H.: High-performance play: The making of machinima. Journal of Media Practice 7(1), 25–42 (2006)
7. Prendinger, H.: The Global Lab: Towards a Virtual Mobility Platform for an Eco -Friendly Society. Transactions of the Virtual Reality Society of Japan 14(2), 163–170 (2009)
8. Shamir, A., et al.: Generating Comics from 3D Interactive Computer Graphics. EEE Computer Graphics and Applications 26(3), 53–61 (2006)
9. Shuda, T., Thawonmas, R.: Frame Selection for Automatic Comic Generation from Game Log. In: Stevens, S.M., Saldamarco, S.J. (eds.) ICEC 2008. LNCS, vol. 5309, pp. 179–184. Springer, Heidelberg (2008)
10. Sparacino, F.: Sto(ry)chastics: A Bayesian Network Architecture for User Modeling and Computational Storytelling for Interactive Spaces. In: Dey, A.K., Schmidt, A., McCarthy, J.F. (eds.) UbiComp 2003. LNCS, vol. 2864, pp. 54–72. Springer, Heidelberg (2003)

11. Sumi, Y., et al.: ComicDiary: Representing individual experiences in a comics style. In: Borriello, G., Holmquist, L.E. (eds.) UbiComp 2002. LNCS, vol. 2498, p. 16. Springer, Heidelberg (2002)
12. Thawonmas, R., Oda, K., Shuda, T.: Rule-Based Camerawork Controller for Automatic Comic Generation from Game Log. In: Yang, H.S., Malaka, R., Hoshino, J., Han, J.H. (eds.) ICEC 2010. LNCS, vol. 6243, pp. 326–333. Springer, Heidelberg (2010)
13. Thawonmas, R., Shuda, T.: Comic Layout for Automatic Comic Generation from Game Log. In: Proc. of 1st IFIP Entertainment Computing Symposium, vol. 279, pp. 105–115 (2008)
14. Tobita, H.: DigestManga: interactive movie summarizing through comic visualization. In: CHI 2010 Extended Abstracts, pp. 3751–3756 (2010)
15. Urban, R., et al.: A Second Life for Your Museum: 3D Multi-User Virtual Environments and Museums. In: Proc. of Museums and the Web 2007, San Francisco (April 2007)
16. Xu, J., et al.: Temporal segmentation of 3-D video by histogram-based feature vectors. IEEE Transactions on Circuits and Systems for Video Technology archive 19(6), 870–881 (2009)
17. Zhang, W., et al.: Story scripting for automating cinematics and cut-scenes in video games. In: Proc. of the 2007 Conference on Future Play, pp. 152–159 (2007)

Knowledge-Based Authoring Tool for Tutoring Multiple Languages

Maria Virvou and Christos Troussas

Department of Informatics, University of Piraeus
80 Karaoli & Dimitriou st., 18534 Piraeus, Greece
`{mvirvou,ctrouss}@unipi.gr`

Abstract. This paper describes the development of a knowledge-based authoring tool, which aims to be useful to teachers and students in multiple language learning. The tool provides assistance on the editing of exercises and also on the creating, updating or erasing a student's profile by the teacher. The system monitors the students while they are answering the exercises and provides appropriate feedback. One of the system's basic functions in to create a student model, which promotes the educational process. The tool incorporates an error diagnosis component, which handles the students' errors in the three languages. Multilingual support in Computer Assisted Language Learning (CALL) is a significant innovation in the related scientific literature.

Keywords: Knowledge-based authoring tool, Intelligent Tutoring Systems, Student Modeling, Error Diagnosis, Multiple Language Learning.

1   Introduction

In the European Union (EU), there are twenty-three official languages along with a wide range of others (www.europa.eu). The EU encourages all its citizens to become multilingual. More specifically, it encourages them to be able to speak two languages in addition to their mother tongue. Even though the EU has very limited influence in this area of educational systems, it offers a number of EU funding programs which actively promote language learning and linguistic diversity (www.europa.eu). Language learning has proved to be more effective with the use of Intelligent Tutoring Systems (ITSs), also called knowledge based tutors. ITSs, which have gained ground the last decades, are computer-based instructional systems that direct customized instruction or feedback to students. ITSs incorporate models of instructional content, which specify the teaching material, the teaching strategies and the way of teaching. ITS design is based on the individualized instruction, namely the instruction can be continuously adapted to best meet the needs of each student separately. Moreover, it offers students the possibility of learning in situations which more closely approximate the ones in which they will use their knowledge.

Although ITSs are becoming more common, building an ITS needs careful preparation in terms of describing the knowledge and possible behaviors of tutors. This procedure demands the technical knowledge of tutors and special skills.

However, it is quite difficult for tutors to be familiarized with the programming of such systems. Hence, these drawbacks can be resolved by the use of the authoring tools, which aim at providing environments for cost-effective development of tutoring systems that can be intelligent and adaptive to individual students (Virvou and Alepis, 2005). The design of an authoring tool addresses the following issues:

1. Using efficient methods, so that the build of an ITS is a cost-effective and a time-saving process.
2. Reducing the technical knowledge and decreasing the skill threshold (programming skills) for authoring an ITS.
3. Providing user interface friendliness.
4. Helping tutors adapt the teaching material to specific needs.
5. Designing methods of testing the tutors.

The resulting authoring tool incorporates a knowledge based approach, which includes the following design principles (Murray, 2000):

1. Represent instructional content and instructional strategies separately.
2. Modularize the instructional content for multiple use and re-use.
3. Creating generic teaching strategies that can be used with different instructional content.
4. Explicitly represent abstract pedagogical entities.
5. Design at the pedagogical level, when possible.

In view of the above, in this paper we present a knowledge based authoring tool for an ITS. This authoring tool gives a non-programmer teacher the possibility of configuring different aspects of the afore-mentioned components, namely the domain knowledge, the student modeling component, the tutoring component and the user interface. We have tried to reduce the time and cost required, along with the programming skills to author an ITS. Our authoring tool has a variety of purposes and intended users and its design promotes the learnability and the productivity.

This paper is organized as follows. First, we present the related work, concerning authoring tools in section 2. In section 3, we discuss our system's architecture. Following, in section 4, we present a description of our system, namely the way of authoring of crucial modules, a case study for the instructor and the student, an analysis concerning the student modeling and error diagnosis components and a modeling of the authoring process. Finally, in section 5, we come up with a discussion about the usability of the resulting authoring tool and we present our next plans.

2   Related Work

ITSs are computer-based systems that provide individualized tutoring to students. Once the ITS is created, authoring its learning material and exercises is an expensive job that requires the intervention of many people that are expert in different areas. Hence, the authoring tool allows the editing or updating of several parts of the ITS. For this reason, many researchers focus on the creation of effective authoring tools, some of which are presented below.

E-Adventure is an authoring tool, developed by Torrente et al (2008) which is not only focused on the abstraction of the programming tasks, leaving the technical features of the game to programmers and media designers, but also on supporting those aspects that are specific of the educational domain, usually not contemplated in toolkits for commercial videogames. Scenejo is an authoring tool, implemented by Spierling et al (2006). Its special emphasis is on the construction of conversational threads for virtual actors using pattern matching, employing transition graph representations as the main interface for authoring. U-CREATE is a graphical authoring tool, developed by Sauer et al (2006) which aims at promoting the creation of elaborated contents in a fast and easy way and of interactive contents for advanced experiencing systems that already exist. CTAT is the authoring tool, implemented by Aleven et al (2006) that addresses to easier and efficient creation of tutors for both real-world use and use in experimental scenarios. The web-based ITS authoring tool by Virvou and Moundridou (2000) is targeted to domains that make use of algebraic equations. In authoring systems such as in FlexiTrainer (Ramachandran et al, 2004) and VersaTutor (Kodaganallur et al, 2004), there are no features which would enable the reuse of learning materials. Furthermore, the authoring tool which was developed by Choksey (2004) is able to build cognitive tutors and focuses on the assistance of author in design, development, implementing, testing, verifying and maintaining cognitive models. The authoring tool, developed by Jokela (2003), is concentrated on mobile phone based multimedia content. Moreover, Mobile Author is an authoring tool, which was implemented by Virvou and Alepis (2005) and allows instructors to create and administer data-bases, concerning characteristics of students, of the domain to be taught and of tests and homework, through any computer or mobile phone. Finally, BioWorld is an authoring tool, developed by Lajoie et al (2001) which promotes scientific reasoning in the medical domain.

However, after a thorough investigation in the related scientific literature, we came up with the result that there was no implementation of authoring systems, which can be used for authoring multilingual systems that can create individual and personalized student models for each student and can perform error diagnosis to learners, through an intelligent way. The multilingual support enables students to learn three different languages in the same time and accept advice concerning the learning material and the exercises.

3   Architecture of Our System

The resulting tool incorporates knowledge about the construction of exercises and a mechanism for student error diagnosis, via a student model, that is applicable to many domains that make use of language learning. The development of our authoring tool addresses the design issues of an ITS as well as the issues of an authoring tool. The ITS that supports our system provides adaptive learning environment for the students by providing personalized instructions. It also offers possibilities of multilingual e-learning. It combines the learning of three languages, which are English, French and German. It is widely agreed that an ITS consists of four components, namely the domain knowledge, the student modeling component, the tutoring component and the user interface (Chou et al., 2003). The domain knowledge consists of a representation

of the domain to be taught (e.g. Languages, Mathematics, etc.). The student modeling component stores some information about the student's performance or behaviors, including his/her knowledge level, his/her preferences, his/her learning history and helps the ITS to personalize the teaching strategy. The tutoring component contains a representation of the teaching strategies of the system and provides adaptive instructions to the students. Finally, the user interface provides the means for the student to interact with the ITS, usually through a graphical environment and sometimes through a rich simulation of the task domain the student is learning.
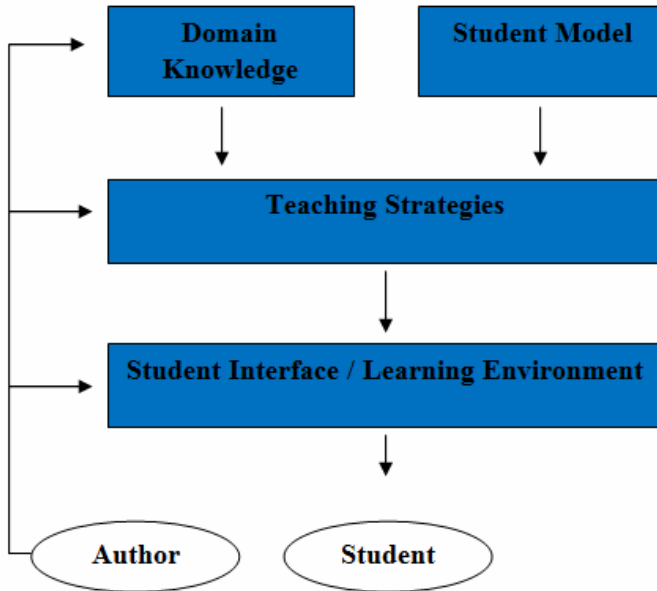
Fig. 1. ITS's components

The authoring system is a layer above the ITS, which helps the teachers configure the different authorable modules of the ITS with minimum effort. It provides user-friendly environments for human instructors to develop their own ITSs of author some special features in an easy way. Naturally, the reasoning abilities of the resulting ITSs have to be provided by the authoring tools. Therefore, the authoring tools incorporate generic and domain-independent methods that can be customized to the particular tutoring domain of each instructor author (Virvou and Alepis, 2005).

The resulting authoring tool is based on the following design principles (Murray, 2003):

1. It has a clear knowledge representation and uses appropriate representational formalisms.
2. It provides visual reification for the structure of the underlying representational framework, given that an authoring system is generally

      used by non-programmers with a low degree of knowledge engineering and programming expertise.
3. It provides tools which allow tutors to author using primitives that have instructional meaning.
4. It represents and authors instructional content modularity so that is can be used for multiple instructional purposes.
5. It provides features powerful enough to author generative content.
6. It provides a powerful set of interface components that is it allows for interface extensibility.
7. It facilitates an opportunistic design process.
8. It facilitates the evolution to more powerful paradigms that is it allows a smooth transition from traditional authoring paradigms to authoring more powerful intelligent tutors.
9. It includes customizable representational formalisms.
10. It promotes the usability and productivity.

Furthermore, the authoring process that is incorporated in our multilingual system, is promoted by the existence of a student model which keeps the students' profiles and provides individualized help and error proneness and diagnosis.

The system's underlying architecture is illustrated in Fig. 2. The teacher's input is the domain description in terms of all the information needed to construct an exercise. This information is stored and managed by the Exercise Answering component, which interacts with the student while s/he is answering a question of an exercise. The Error Diagnoser component is responsible for finding out the cause of a student's mistake. The system, via the student model, can hold information about the performance of the student in one language. Then it can provide the student with the appropriate feedback for the other two languages, via the Advice Generator component. The student model is updated at every interaction of the student with the system and keeps the latest information for him/her. This information can help the system perform adaptive navigation support, which is considered important especially in distance learning situations, where the learning system has to play the role of the teacher and be able to help the student navigate through the course and support him/her individually in an exercise answering process (Specht and Oppermann, 1998).
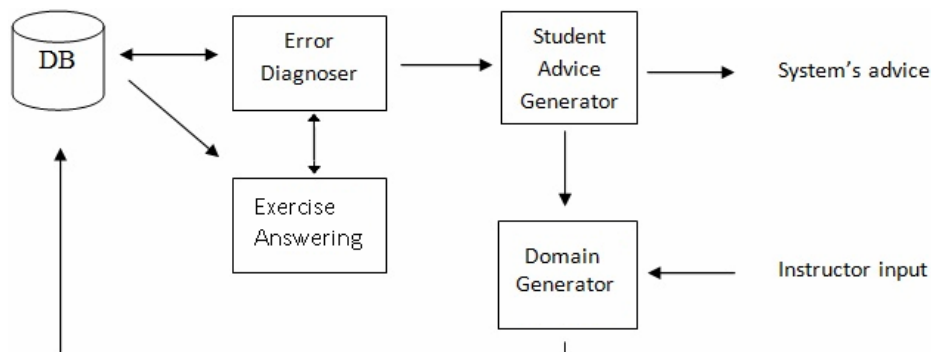

Fig. 2. The architecture of the system

4   Description of the System

The following subsections give in detail the way with which the crucial parts of an ITS are authored. Furthermore, we present a case study for the instructor and the student and we describe thoroughly the student modeling and error diagnosis components. Finally, we visualize our system's processes, by using the Unified Modeling Language (UML).

4.1   Authoring Domain Knowledge

Authoring domain knowledge comprises of creating new course structures as well as modifying the existing ones. The teachers have the permission to create exercises, define the right answer, erase an exercise or simply make several changes on them. Moreover, this module consists of setting up the repository (Chakraborty ey al., 2010). This process includes, inserting the available materials into the repository, as well as creating new learning materials. Given that our tool addresses to multiple language learning, we have implemented a platform where the teacher may edit the exercises of all the languages that are taught, in a simple way.

4.2   Authoring Student Model

Student Model is an abstract representation of the students' learning from the system, along with a student's profile, which includes personal information of him/her and other information concerning his/her performance and progress. It keeps track of student behaviors and makes inferences about what the student knows (Murray, 2003). A teacher can easily update a student's profile by interacting with the system, namely pressing buttons, choosing from a drop-down list and picking one from given multiple choices. Our system does not allow the teacher to make changes to the performance and the progress of the student. This prohibition ensures the integrity of the system. The student model can also be used as a "bug library", since misconceptions and buggy procedures can be defined to keep track of known classes of common deficiencies (Murray, 2003).

Furthermore, this module offers the possibility to the teachers to register a new student so that s/he is able to make use of the system and learn multiple languages. This registration can be conducted by filling in a simple form with the student's username, name and surname and gender.

4.3   Authoring of Teaching Model

This is an automated module, which works dynamically and makes pedagogic decisions, depending on the information of students' performance gathered at run time (Chakraborty et al., 2010). Hence, the teachers do not have any direct control over the decision making procedures of this module. They get informed about the performance of the teachers, so that they can adapt the difficulty of the exercises to the level of each one student. This module promotes the individual and personalized learning and renders the students capable of filling all requirements of the educational procedure.

4.4   Case Study for the Instructor

When the instructor is going to create multiple choice exercises, s/he can fill in the form with the sentence and then s/he can add four different choices and thereafter define the right choice. After the construction of a multiple choice exercise, the tool lets the instructor preview it and then s/he can add it into the system, by pressing a button of acceptance. While students are tackling the given exercises, the system collects evidence about the level of difficulty so that it can provide feedback to the instructor. In this way, the instructor may edit a multiple choice exercise in order to adapt it to the level of knowledge of students. Nevertheless, there is one final test in the system that cannot be edited by the teachers. The students are asked to fill in the gaps of twenty given sentences. These sentences are chosen at random each time. The system may choose these sentences from the database, so that they are different for each student. In this way, we promote the diaphaneity in the students' performance evaluation.

The authoring tool also incorporates a user modeling component that conducts diagnosis to the students' errors. This mechanism is used by the system when a student tackles the final exercise and is explained in detail in the next section.

4.5   Case Study for the Student

The system recognizes each student by his/her user name and his/her password, which were defined upon his/her registration. The domain knowledge of the resulting system consists of three languages, which are English, French and German. The system follows the same logical structure, for each one of these languages. This structure includes five novice level lessons for beginner students. The first lesson is the learning of the alphabet of the corresponding language. The alphabet is given both in capital and in minuscule letters. The second lesson encompasses the learning of months and days, along with their pronunciation. The third lesson encompasses the genders and the articles, so as to render the students capable of mastering these subjects. The fourth lesson describes in detail the personal and the possessive pronouns. The final lesson familiarizes students with the verbs "to be" and "have", as main verbs. An important issue considering these lessons is that the there is a multiple choice test for each one of the three last lessons, so that the students are evaluated and examined concerning their knowledge and comprehension of the previous lessons. If the students are not found to be adequately prepared to go on to the next lesson, they have to study again the relative theory. We used multiple choice exercises to evaluate their performance, as they are a mainstay of achieving testing and also provided us with the ability to measure the students' achievements (Rodriguez, 2005). The multiple choice questions can be authored by the teacher at any time, as it was described above.

The advice generator is activated not only when a student makes an error, but also in every lesson, when the student has to study the theory. Initially, when a student uses the system for the first time, there is only an advice in each lesson, concerning the theory. When the student answers the multiple-exercises, then he is evaluated and gets information from the system in an appropriate way about his/her grade and also is permitted to pass to the next lesson, if his performance is satisfactory.

Furthermore, when a student has completed both the theory and the exercises of a lesson in one particular language, s/he has the possibility to do the same lesson in another language. In this way, the system will provide the student with advice, concerning the new theory to be taught and also the differences with the other languages that s/he was taught.

A matter of great importance is the existence of the final test, which is an exercise where the student is asked to fill in the gaps of a specific sentence. After completing the test, the student can check his/her performance and take advice about the point where the error lies and the categories of error. The questions of this test are adapted to the student's performance so they are different each time. While the student is in the process of answering this exercise, the system monitors his/her actions. If the student makes a mistake, the diagnostic component of the system will attempt to diagnose the cause of it. The diagnosis of the underlying cause of a mistake is a difficult task for an ITS because the observable symptoms need to be analyzed further. As Hollnagel (1993) pointed out, there is an important distinction between the underlying cause or genotype of an error and the observable manifestation or phenotype of the error. In the final test, the system recognizes the errors and associates them with the lessons taught by the system. There are five categories of errors that can be recognized by the system:

1. Article and pronoun mistakes

   For example the user may have used "a" instead of "an" or "he" instead of "we".

2. Spelling mistakes

   A spelling mistake is a result of letter redundancy, letter missing or interchange of two neighboring letters.

3. Verb mistakes

   Verb mistakes occur when the user has typed another person than the correct one, for example s/he may have typed "I has" instead of "I have".

4. Unanswered questions

   The user may have no idea about what s/he should write and leave the question unanswered. That means that s/he has lack in theory.

5. Language Confusion

   Our system is a multilingual learning system, which means that a student may learn three languages at the same time. However, there is the possibility of student's getting confused, concerning the proper use of an article or verb. Namely, the student may have used "I am" instead of "Je suis", which is the French equivalent.

4.6 Student Modeling and Error Diagnosis

Student modeling represents the computer system's belief about the learner's knowledge, namely it is responsible for the creation of a student's profile with all the

crucial information about him/her. It is generally used in connection with applications computer-based instructional systems. Student modeling is crucial for an intelligent learning environment to be able to adapt to the needs and knowledge of individual students. Virvou et al. (2000) support that the student modeler is responsible for preserving the system's estimation of the learner's proficiency in the domain as well as his/her proneness to commit errors.

Our system constructs a student model, which gives assistance to the learner, providing feedback or interprets his/her behavior. One significant element is that before the student's starting a multiple-choice test in another language, the system informs him/her about his/her performance in the corresponding test of the lesson of the already taught language and gives him/her advice concerning the test s/he is about to do. Moreover, concerning the final test, the student modeler checks the student's answer and in case of an error and it performs error diagnosis. In this case, the system checks the complexion of the error and acts in a way that it will be described in the next section.

A matter of great importance is the existence of a long term user model for each student. The system includes also a form, which keeps information about the student's progress in the three languages, the total grade in each one of the three languages and all the results of the tests. Cook and Kay (1994) have shown that students benefit from viewing their own student models. For this reason, this form can be presented to students so that they stay aware of their advance of knowledge.

Chapelle (1998) points out that CALL programs must promote noticing (focus on form) that will result in the improvement of students' existing grammatical knowledge. This can be accomplished by evaluating the students' performance through several tests. In our system, there are two types of tests as mentioned above, which are the multiple choice exercises and the exercises where the user is asked to fill in the gap inside a sentence.

Multiple choice questions are the most widely used and highly regarded of the selection- type questions (Rodriguez, 2005). Multiple choice exercises are a form of assessment in which the user asked to select the best possible answer out of the choices from a list. One of them is the correct, while the others answers are erroneous. In our system, we have developed a library, which keeps all the erroneous answers and correlates them with a category of error, so that the student should have an integral idea of his/her knowledge and know exactly where s/he has weaknesses in.

In the final test, the student is asked to fill in the gap inside a sentence. The system examines the student's knowledge in all the lessons which have been taught in each language separately. There are twenty questions, which are different for each student and based on his/her user model. After completing the gaps, the system gives the results of the final examination. It shows the grade of the student and spots the erroneous answers. Furthermore, the system corresponds to the erroneous answers for each category of errors and gives a percentage that helps students understand in which lesson they have deficiencies.

Fig. 3 illustrates the relationship between the Student Modeling and Error Diagnosis components and gives an overview of the system.
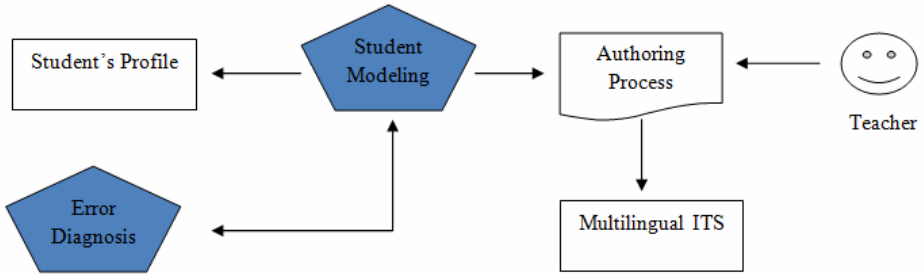
Fig. 3. Student Modeling and Error Diagnosis Components

4.7   Modeling the System's Authoring Process

The following sequence diagrams in UML describe the sequences of messages of our authoring system. Fig. 4 illustrates the updating of a student's profile by the teacher. Fig. 5 illustrates the editing of multiple choice exercises. The dotted lines extending downwards indicate the timeline. Time flows from top to bottom. The arrows represent the messages (stimuli) from an actor or object to other objects. The actor "Student" exists in all the diagrams, but s/he cannot send or receive messages, given that s/he does not participate in the authoring process.
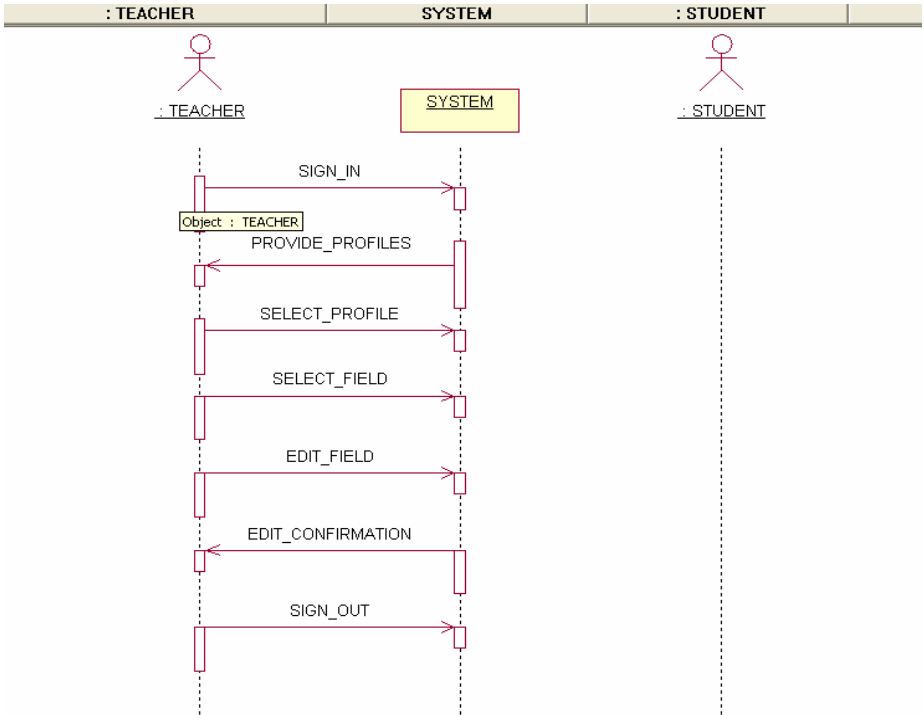
Fig. 4. Updating a student's profile
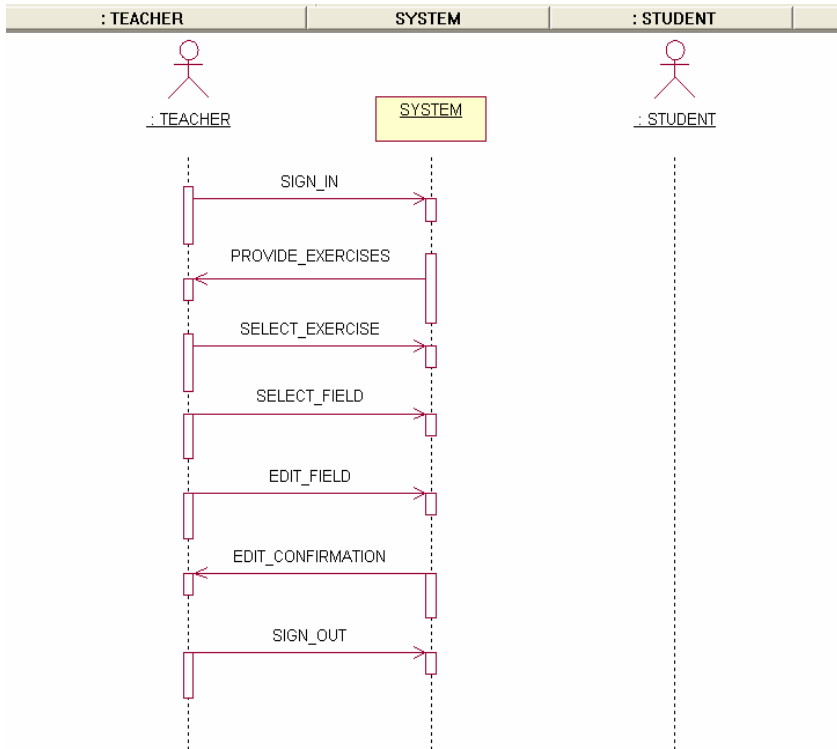
Fig. 5. Editing a multiple choice exercise

5   Conclusions

In this paper, we have described a knowledge-based authoring tool that models the domain of knowledge, the teaching strategies and the student states. The resulting knowledge-based tutor has the following features:

1. A representation of the knowledge to be taught.
2. Learning material is separate from learning strategies.
3. A student model exists, makes inferences about the student's state and is connected to an error diagnosis component which promotes the educational and authoring processes.
4. Instructional content is adapted to students' state.
5. Instructional decisions are predicated upon the inferred student state.

Our authoring system is a useful tool for teachers with no programming skills, who want to author the multilingual learning material. This authoring process can be conducted by using a sophisticated editor in a user friendly environment. In this way, we aim at decreasing the effort it takes to build instructional systems. With the same

effort, our system can help increasing the adaptivity, depth and effectiveness of resulting instructional systems.

It is in our future plans to evaluate our authoring tool in order to examine the degree of usefulness of the educational tool for the teachers who are going to author parts of the ITS. Furthermore, we are going to evaluate the usefulness of the student model and error diagnosis components for the learners who are going to use the multilingual educational system.

References

1. Murray, T.: An overview of intelligent tutoring system authoring tools: Updated analysis of the state of the art. In: Authoring Tools for Adv. Tech. Learning Env, Ainsworth and Blessing, pp. 493–546. Kluwer Academic Publishers, The Netherlands (2003)
2. Choksey, S.D.: Developing An Affordable Authoring Tool For Intelligent Tutoring Systems. MS Thesis, USA (2004)
3. Chakraborty, S., Roy, D., Bhowmick, P.K.: An Authoring System for Developing Intelligent Tutoring System. In: TechSym 2010 - Proceedings of the 2010 IEEE Students' Technology Symposium, pp. 196–205 (2010)
4. Virvou, M., Alepis, E.: Mobile educational features in authoring tools for personalised tutoring. Computers and Education, 53–68 (2005)
5. Ramachandran, S., Remolina, E., Fu, D.: FlexiTrainer: A visual authoring framework for case-Based intelligent tutoring systems. In: Intelligent Tutoring Systems, Maceio, Brazil, pp. 848–850 (2004)
6. Kodaganallur, V., Weitz, R., Rosenthal, D.: VersaTutor – An architecture for a constraint-based intelligent tutor generator. In: Proceedings of the Thirteenth Annual World Wide Web Conference, USA, pp. 474–475 (2004)
7. Virvou, M., Moundridou, M.: A Web-Based Authoring Tool for Algebra-Related Intelligent Tutoring Systems. In: EducationalTechnology & Society, pp. 61–70 (2000)
8. Murray, T.: Authoring Knowledge Based Tutors: Tools for Content, Instructional Strategy, Student Model, and Interface Design. International Journal of Artificial Intelligence in Education, 98–129 (2000)
9. Torrente, J., Moreno-Ger, P., Fernández-Manjón, B., Sierra, J.-L.: Instructor-oriented Authoring Tools for Educational Videogames. In: Proceedings of the 8th IEEE International Conference on Advanced Learning Technologies ICALT, Spain, pp. 516–518 (2008)
10. Aleven, V., Sewall, J., McLaren, B., Koedinger, K.: Rapid Authoring of Intelligent Tutors for Real-World and Experimental Use. In: Sixth IEEE International Conference on Advanced Learning Technologies (ICALT 2006), The Netherlands, pp. 847–851 (2006)
11. Sauer, S., Osswald, K., Wielemans, X., Stifter, M.: U-Create: Creative Authoring Tools for Edutainment Applications. In: Göbel, S., Malkewitz, R., Iurgel, I. (eds.) TIDSE 2006. LNCS, vol. 4326, pp. 163–168. Springer, Heidelberg (2006)
12. Jokela, T.: Authoring tools for mobile multimedia content. In: International Conference on Multimedia and Expo., vol. 2, pp. 637–640. IEEE, Los Alamitos (2003)
13. Spierling, U., Weiß, S.A., Müller, W.: Towards Accessible Authoring Tools for Interactive Storytelling. In: Göbel, S., Malkewitz, R., Iurgel, I. (eds.) TIDSE 2006. LNCS, vol. 4326, pp. 169–180. Springer, Heidelberg (2006)

14. Lajoie, S.P., Faremo, S., Wiseman, J.: A knowledge-based approach to designing authoring tools: From tutor to author. In: Moore, J.D., Redfield, C., Johnson, L.W. (eds.) Artificial intelligence in education: AI-ED in the wired and wireless future, pp. 77–86 (2001)
15. European Union, http://www.europa.eu
16. Chapelle, C.: Multimedia CALL: Lessons to be learned from research on instructed SLA. Language Learning and Technology 2, 21–39 (1998)
17. Cook, R., Kay, J.: The Justified User Model: A Viewable, Explained User Model. In: Proceedings of the Fourth International Conference on User Modelling, pp. 145–150 (1994)
18. Virvou, M., Maras, D., Tsiriga, V.: Student modeling in an intelligent tutoring system for the passive voice of English language. Educational Technology and Society 3(4), 139–150 (2000)
19. Rodriguez, M.C.: Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. In: Educational Measurement: Issues and Practice, pp. 3–13 (2005)
20. Hollnagel, E.: The Phenotype of Erroneous Actions. International Journal of Man-Machine Studies 39, 1–32 (1993)
21. Chou, C.–Y., Chan, T.–W., Lin, C.-J.: Redefining the learning companion: the past, present, and future of educational agents. Computers & Education 40, 255–269 (2003)
22. Specht, M., Oppermann, R.: ATS-Adaptive Teaching System a WWW-based ITS. In: Timm, U. (ed.) Proceedings of Workshop Adaptivität und Benutzermodellierung in Interaktiven Software System (1998)

Evaluating an Affective e-Learning System Using a Fuzzy Decision Making Method

Katerina Kabassi[2], Efthimios Alepis[1], and Maria Virvou[1]

[1] Department of Informatics, University of Piraeus,
80 Karaoli & Dimitriou St., 18534 Piraeus, Greece
`{kkabassi,mvirvou,geoatsi}@unipi.gr`
[2] Department of Ecology and the Environment,
Technological Educational Institute of the Ionian Islands,
2 Kalvou Sq., 29100 Zakynthos, Greece

Abstract. This paper presents how a fuzzy multi-criteria decision making theory has been used for evaluating an e-learning system with affective interaction capabilities. More specifically, as a test-bed for our evaluation we have used an e-learning system that teaches a programming language. This programming language is similar to the well known Logo programming language, while it incorporates a more user-friendly graphical user interface with the presence of animated tutoring characters. In our evaluation we have compared the affective system with a standard programming language learning environment and the results have revealed a general preference of the users in using the affective system than the standard system.

1 Introduction

In the last decade, the research field of affective computing has become a very important and popular because it focuses on recognizing and reproducing human feelings within human computer interaction. Human feelings are considered very important but only recently have started being taken into account in user interfaces. However, analyzing and eliciting human emotion is very complicated and as Picard points out (Picard, 2003) there is little hope that computers will accurately understand human feelings during the next decades. This derives from the fact that human emotions are very idiosyncratic and variable and emotion recognition is quite difficult even in human-human interaction. Thus, the area of affective computing is not yet well understood and needs a lot more research to reach maturity.

Cognitive science has opened new paths in exploring human behavior by technological means and software improvements, while in the past the development of affective computer interfaces was basically a technology-driven phenomenon (Picard, 2003). The specific research is quite promising considering that to date, most state-of-the art multimodal interfaces consist of more than one mode. Systems that combine speech and keyboard input and systems that combine speech and facial detection constitute two most extensively developed and supported areas within the fields of human-computer interaction (Pantic, M. & Rothkrantz, 2003).

In the past, the authors of this paper have developed systems that incorporate emotion recognition capabilities, based on several programming techniques and algorithmic approaches (Alepis et al. 2009, Alepis et al. 2010). These systems showed significant results in recognizing and reproducing emotions. However, our next goal was to evaluate such a system by exploring its benefits in terms of user interaction in comparison with a standard non affective interaction system.

In view of the above, we have used a programming language e-learning system that incorporated affective interaction mechanisms. The system was evaluated in comparison to a standard system that does not incorporate any intelligence using Fuzzy Simple Additive Weighting (FSAW) (Chou et al. 2008). Indeed, the evaluation of software involves several criteria and, therefore, can be considered as a multi-criteria problem. Especially, the FSAW theory was selected because it is rather simple and has the advantage of allowing multi-criteria problems to accommodate linguistic terms represented as fuzzy numbers. This facilitates the creation of a decision procedure that is more realistic than other existing theories (Chou et al. 2008).

2   Fuzzy Simple Additive Weighting

Zadeh (1965) pioneered the use of Fuzzy Set Theory (FST) to address problems involving fuzzy phenomena. In a universe of discourse X, a fuzzy subset $\tilde{A}$ of X is defined with a membership function $\mu_{\tilde{\alpha}}(x)$ that maps each element x in X to a real number in the interval [0, 1]. The function value of $\mu_{\tilde{\alpha}}(x)$ signifies the grade of membership of x in $\tilde{A}$. When $\mu_{\tilde{\alpha}}(x)$ is large, its grade of membership of x in $\tilde{A}$ is strong (Keufmann & Gupta, 1991).

A fuzzy set $\tilde{A} = (a, b, c, d)$ on R, a < b < c < d, is called a trapezoidal fuzzy number if its membership function is

$$\mu_{\tilde{\alpha}}(x) = \begin{cases} \dfrac{(x-a)}{(b-a)}, & a \le x \le b \\ 1, & b \le x \le c \\ \dfrac{(x-d)}{(c-d)}, & c \le x \le d \\ 0, & otherwise \end{cases}$$

where a, b, c, d are real numbers (Dubois & Prade 1978, Keufmann & Gupta 1991).

Trapezoidal fuzzy numbers are the most widely used forms of fuzzy numbers because they can be handled arithmetically and interpreted intuitively.

The FSAWS procedure based on above conceptual model is as follows:

Step 1: Form a committee of decision-makers. Choose the attributes and identify the prospective alternatives. A committee of decision-makers is formed to determine the most appropriate alternative.

Step 2: Determine the degree of importance (or reliability) of the decision-makers. If the degrees of importance (or reliability) of decision-makers are equal, then the group

of decision-makers is deemed a homogeneous group ($I_1 = I_2 = ... = I_k = \dfrac{1}{k}$); otherwise, the group of decision-makers is called a heterogeneous (non-homogeneous) group.

Step 3: Introduce linguistic weighting variables (Table 1) for decision-makers to assess attributes importance, and compute aggregated fuzzy weights of individual attributes. Let $\widetilde{W}_{jt} = (a_{jt}, b_{jt}, c_{jt}, d_{jt})$, $j = 1,2,...,n; t = 1,2,...,k$ be the linguistic weight given to subjective attributes $C_1, C_2,..., C_h$ and objective attributes $C_{h+1}, C_{h+2},..., C_n$ by decision-maker D_t. The aggregated fuzzy attribute weight, $\widetilde{W}_j = (a_j, b_j, c_j, d_j)$, $j = 1,2,...,n$ of attribute C_j assessed by the committee of k decision makers is defined as $\widetilde{W}_j = (I_1 \otimes \widetilde{W}_{j1}) \oplus (I_2 \otimes \widetilde{W}_{j2}) \oplus ... \oplus (I_{1k} \otimes \widetilde{W}_{jk})$, where $a_j = \sum_{t=1}^{k} I_t a_{jt}$, $b_j = \sum_{t=1}^{k} I_t b_{jt}$, $c_j = \sum_{t=1}^{k} I_t c_{jt}$, $d_j = \sum_{t=1}^{k} I_t d_{jt}$

Step 4: Defuzzify the fuzzy weights of individual attributes; compute the normalized weights and construct the weight vector.

To defuzzify the weights of the fuzzy attributes, the signed distance is adopted. The defuzzification of $\widetilde{W}_j$, denoted as $d(\widetilde{W}_j)$ is therefore given by

$$d(\widetilde{W}_j) = \frac{1}{4}(a_j + b_j + c_j + d_j), \quad j = 1,2,...,n$$

The crisp value of the normalized weight for attribute C_j denoted as W_j, is given by

$$W_j = \frac{d(\widetilde{W}_j)}{\Sigma_{j=1}^{n} d(\widetilde{W}_j)}, \quad j = 1,2,...,n,$$

where $\sum_{j=1}^{n} W_j = 1$. The weight vector $W = [W_1, W_2,..., W_n]$ is therefore formed.

Step 5: Use linguistic rating variables (Table 2) for decision-makers to assess fuzzy ratings of alternatives with respect to individual subjective attributes, and then pool them to obtain the aggregated fuzzy ratings. Let $\widetilde{x}_{ijt} = (o_{ijt}, p_{ijt}, q_{ijt}, s_{ijt})$, $i = 1,2,...,m$, $j = 1,2,...,h$, $t = 1,2,...,k$ be the linguistic suitability rating assigned to alternative location A_i for subjective attribute C_j by decision-maker D_t. Let us further define $\widetilde{x}_{ijt}$ as the aggregated fuzzy rating of alternative A_i for subjective attribute C_j, such that

$$\tilde{x}_{ij} = (I_1 \otimes \tilde{x}_{ij1}) \oplus (I_2 \otimes \tilde{x}_{ij2}) \oplus ... \oplus (I_{1k} \otimes \tilde{x}_{ijk})$$

which can subsequently be represented and computed as

$$\tilde{x}_{ij} = (o_{ij}, p_{ij}, q_{ij}, s_{ij}), \ i = 1,2,...,m, \ j = 1,2,...,h$$

where $o_{ij} = \sum_{t=1}^{k} I_t o_{ijt}$, $p_{ij} = \sum_{t=1}^{k} I_t p_{ijt}$, $q_{ij} = \sum_{t=1}^{k} I_t q_{ijt}$, $s_{ij} = \sum_{t=1}^{k} I_t s_{ijt}$.

Step 6: Construct a fuzzy rating matrix based on fuzzy ratings. The fuzzy rating matrix $\tilde{M}$ can be concisely expressed in matrix format

$$\tilde{M} = \begin{bmatrix} \tilde{x}_{11} & \tilde{x}_{12} & ... & \tilde{x}_{1n} \\ \tilde{x}_{21} & \tilde{x}_{22} & .. & \tilde{x}_{2n} \\ . & . & & . \\ . & . & ... & . \\ . & . & & . \\ \tilde{x}_{m1} & \tilde{x}_{m2} & ... & \tilde{x}_{mn} \end{bmatrix}$$

where $\tilde{x}_{ij}, \forall i, j$ is the aggregated fuzzy rating of alternative A_i, $i = 1,2,...,m$ with respect to attribute C_j.

Step 8: Derive total fuzzy scores for individual alternatives by multiplying the fuzzy rating matrix by their respective weight vectors.

Obtained total fuzzy score vector by multiplying the fuzzy rating matrix $\tilde{M}$ by the corresponding weight vector W, i.e.,

$$\tilde{F} = \tilde{M} \otimes W^T = \begin{bmatrix} \tilde{x}_{11} & \tilde{x}_{12} & ... & \tilde{x}_{1n} \\ \tilde{x}_{21} & \tilde{x}_{22} & .. & \tilde{x}_{2n} \\ . & . & & . \\ . & . & ... & . \\ . & . & & . \\ \tilde{x}_{m1} & \tilde{x}_{m2} & ... & \tilde{x}_{mn} \end{bmatrix} \otimes \begin{bmatrix} W_1 \\ W_2 \\ . \\ . \\ . \\ W_n \end{bmatrix} = \begin{bmatrix} \tilde{x}_{11} \otimes W_1 \oplus \tilde{x}_{12} \otimes W_2 \oplus ... \tilde{x}_{1n} \otimes W_n \\ \tilde{x}_{21} \otimes W_1 \oplus \tilde{x}_{22} \otimes W_2 \oplus ... \tilde{x}_{2n} \otimes W_n \\ . \\ . \\ . \\ \tilde{x}_{m1} \otimes W_1 \oplus \tilde{x}_{m2} \otimes W_2 \oplus ... \tilde{x}_{mn} \otimes W_n \end{bmatrix} = \begin{bmatrix} \tilde{f}_1 \\ \tilde{f}_2 \\ . \\ . \\ . \\ \tilde{f}_n \end{bmatrix} = [\tilde{f}_i]_{m*1}$$

where $\tilde{f}_i = (r_i, s_i, t_i, u_i)$, $i = 1,2,...,m$.

Step 9: Compute a crisp value for each total score using a defuzzification method and select the alternative(s) with the maximum total score.

Rank total fuzzy scores $\tilde{f}_1, \tilde{f}_2,..., \tilde{f}_m$ by the signed distance to determine the best location. Determine crisp total scores of individual locations by the following defuzzification equation:

$$d(\tilde{f}_i) = \frac{1}{4}(r_i + s_i + t_i + u_i), \ i = 1,2,...,m$$

where $d(\tilde{f}_i)$ gives the defuzzified value (crisp value) of the total fuzzy score of location A_i by using the signed distance. The ranking of the locations can then be preceded with the above crisp value of the total scores for individual alternatives.

3   Overall Description of the System

In this section we give a short overview of the affective e-learning system. As we have mentioned the system that incorporates the affective capabilities was a computer assisted programming language learning system. The affective system also incorporated an emotional animated tutoring character that was present during the e-learning process in order to make the interaction more human-like. The tutoring character would try to sympathise with the aims of each user as student and would try to elicit human emotions.
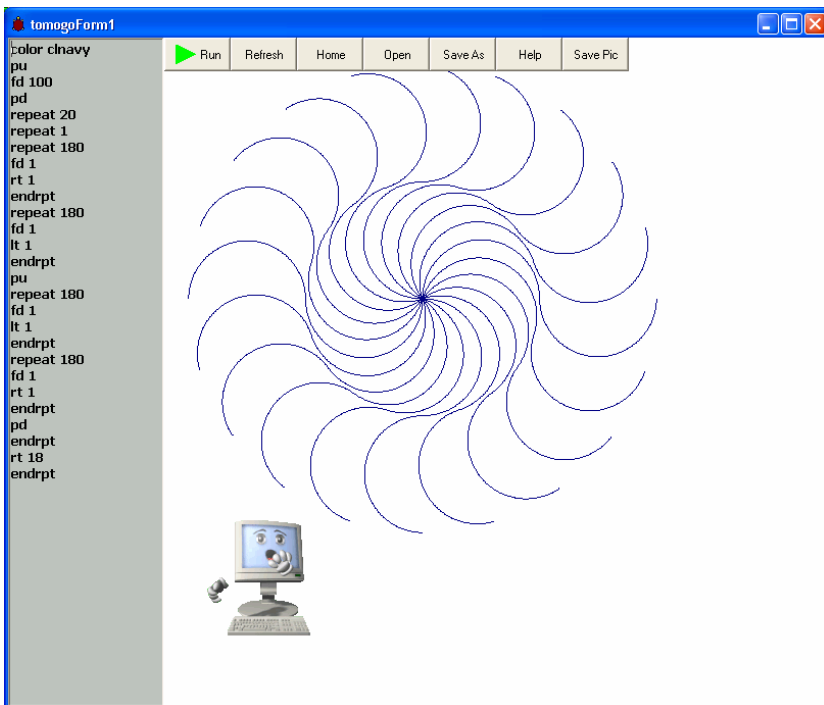
Fig. 1. A user is typing programming commands to create a specific drawing

Figure 1 illustrates a snapshot from user interaction where a user is typing programming commands in order to produce a complicated drawing. In figure 2 we may see the tutoring character that congratulates a user for producing a very complicated drawing by using the programming language. While the users interact with the
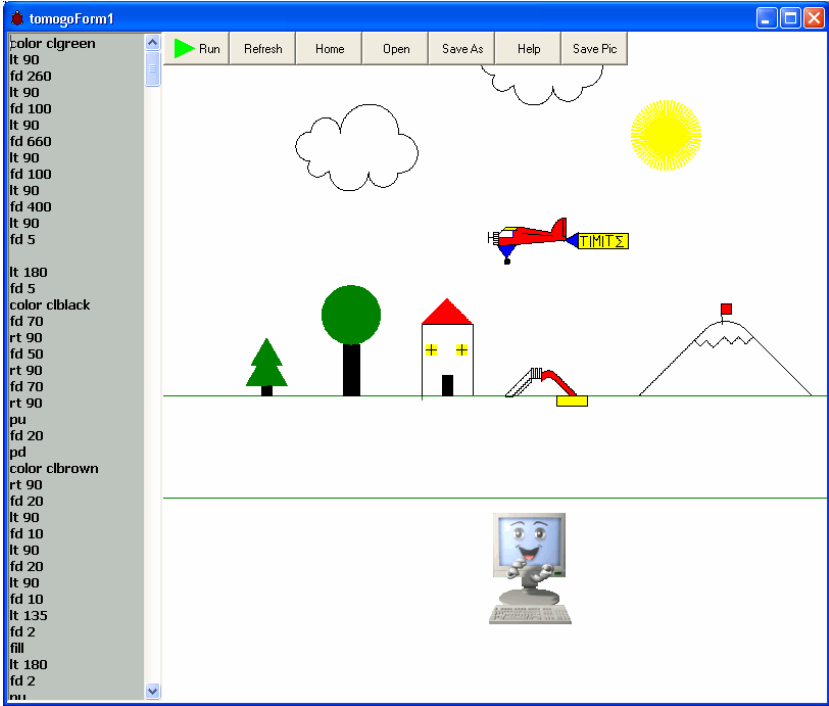
Fig. 2. The Tutoring character congratulates a user

affective system, both their oral and keyboard actions were recorded in order to be interpreted in emotional interaction terms. A more complicated description of the affective system is beyond the scopes of this paper that aims basically in the system's evaluation.

4   Evaluation Experiment

For the evaluation of the system, the evaluation experiment was designed taking into account the steps of the FSAW method. For this purpose a committee of the decision-makers was formed to evaluate the system. The committee of decision makers was consisted of 10 users that were randomly selected; 4 teachers, 4 students, and 2 software engineers. The group of decision makers was homogeneous as the reliability (importance) of the decision-makers was equal. The criteria that were taken into account for the evaluation of the system was:

- Affective Adaptation (Af): this criterion shows how successful the system was in adapting its interaction to the emotions to the user interacting to the system.

- User friendly (Uf): the specific criterion shows what the users think about the interaction with the adaptive system, whether it is intrusive or not, whether it provides natural interaction etc.
- Needs fulfillment (Nf): this criterion shows how successful the system was in addressing the users' needs and presenting the appropriate information to the user.
- Interest satisfaction (Is): this criterion reveals how successful the system was in addressing the users' interests.

The decision makers were first asked to evaluate the above mentioned attributes with respect to their importance in adaptive software evaluation. This procedure resulted in aggregated fuzzy weights of individual attributes. For the evaluation of the above mentioned criteria the decision makers used the linguistic variables presented in table 1.

Table 1. Linguistic variables and fuzzy numbers for the importance weight

Linguistic variables	Fuzzy numbers
Very low (VL)	(0, 0, 0, 3)
Low (L)	(0, 3, 3, 5)
Medium (M)	(2, 5, 5, 8)
High (H)	(5, 7, 7, 10)
Very high (VH)	(7, 10, 10, 10)

As soon as all the linguistic variables of the criteria weights were collected, the fuzzy weights of the criteria are calculated. More specifically, each linguistic variable is translated to a fuzzy number as this is presented in table 1. These values are used for calculating the aggregated fuzzy weight for each one of the four attributes:

$$\tilde{W}_{A_f} = (4.7, 7.3, 7.3, 9.4), \qquad \tilde{W}_{U_f} = (5.3, 7.7, 7.7, 9.8), \qquad \tilde{W}_{N_f} = (1.0, 3.4, 3.4, 6.1),$$

$\tilde{W}_{I_s} = (4.2, 6.8, 6.8, 9.2)$. These weights are defuzzified according to the step 4 of the theory: $d(\tilde{W}_{A_f}) = 7.2$, $d(\tilde{W}_{U_f}) = 7.6$, $d(\tilde{W}_{N_f}) = 3.5$, $d(\tilde{W}_{U_f}) = 6.8$. The crisp values of the normalized weights are calculated:

$$W_{A_f} = \frac{d(\tilde{W}_{I_s})}{\Sigma_{j=1}^{4} d(\tilde{W}_{I_s})} = \frac{7.2}{25.1} = 0.29 \quad , \quad W_{U_f} = \frac{d(\tilde{W}_{U_f})}{\Sigma_{j=1}^{4} d(\tilde{W}_{U_f})} = \frac{7.6}{25.1} = 0.30,$$

$$W_{N_f} = \frac{d(\tilde{W}_{N_f})}{\Sigma_{j=1}^{4} d(\tilde{W}_{N_f})} = \frac{3.5}{25.1} = 0.14, \quad W_{I_s} = \frac{d(\tilde{W}_{I_s})}{\Sigma_{j=1}^{4} d(\tilde{W}_{I_s})} = \frac{6.8}{25.1} = 0.27. \quad \text{The}$$

weight vector is, therefore, formed to $W = (0.29, 0.30, 0.14, 0.27)$.

5   Results

The 12 users were asked then to evaluate ADAPTIGIS using the linguistic rating variables presented in Table 2 with respect to the four attributes.

Table 2. Linguistic variables and fuzzy numbers for the ratings

Linguistic variables	Fuzzy numbers
Very poor (VP)	(0, 0, 0, 20)
Between very poor and poor (B. VP & P)	(0, 0, 20, 40)
Poor (P)	(0, 20, 20, 40)
Between poor and fair (B. P & F)	(0, 20, 50, 70)
Fair (F)	(30, 50, 50, 70)
Between fair and good (B. F & G)	(30, 50, 80, 100)
Good (G)	(60, 80, 80, 100)
Between good and very good (B. G & VG)	(60, 80, 100, 100)
Very good (VG)	(80, 100, 100, 100)

The linguistic rating variables were processed using the formulae of step 5 of the algorithm in order to obtain the aggregated fuzzy ratings. The detailed presentation of the calculations is beyond the scope of this paper and, therefore, only the results of this processing are presented in the fuzzy rating matrix (Table 3).

Table 3. Fuzzy rating matrix

	Af	Uf	Nf	Is
ADAPT IGIS	(61,81,85,97)	(61,81,85,97)	(51,71,77,91)	(68,88,96,100)
Standard GIS	(15,33,45,65)	(36,54,65,85)	(39,59,71,91)	(21,39,47,67)

What seems rather interesting is that the users not only value more the affective system in the attributes related to the affection but they are influenced when evaluating the adaptive system in other criteria that do not seem directly related. In order to derive the total fuzzy scores for individual alternative, the fuzzy rating matrix is multiplied by the corresponding weight vector W.

$$\tilde{F} = \begin{bmatrix} 61,81,85,97 & 61,81,85,97 & 51,71,77,91 & 68,88,96,100 \\ 15,33,45,65 & 36,54,65,85 & 39,59,71,91 & 21,39,47,67 \end{bmatrix} \otimes \begin{bmatrix} 0.29 \\ 0.30 \\ 0.14 \\ 0.27 \end{bmatrix} =$$

$$\tilde{F} = \begin{bmatrix} (61.5,81.5,86.9,97.0) \\ (26.3,44.6,55.2,75.2) \end{bmatrix}$$

The crisp value for each total score is computed using the defuzzification method.

$$d(\tilde{f}_1) = \frac{1}{4}(61.5 + 81.5 + 86.9 + 97.0) = 326.8$$

$$d(\tilde{f}_2) = \frac{1}{4}(26.3 + 44.6 + 55.2 + 75.2) = 201,2$$

The crisp value of the first alternative that corresponds to the affective system is much higher than the crisp value of the standard learning system. As a result, the users seem to prefer in general the affective system than a standard learning system.

6 Conclusions

In this paper, we described how a fuzzy multi-criteria decision making theory can be used for evaluating an affective learning system. The theory that was selected was the FSAW theory. The particular theory is very simple and uses linguistic terms. Indeed, previous studies (e.g. Virvou & Kabassi 2003) revealed that users had a difficulty in quantifying criteria while evaluating a software system. Therefore, in this case, FSAW was selected that uses linguistic terms that can be further translated into fuzzy numbers. This results in making the procedure more realistic, suitable, user friendly and effective.

The users as decision makers who participated in the evaluation experiment seem to prefer in general the affective system than a standard learning system. Moreover, the evaluation experiment revealed that the affective characteristic of the adaptive e-learning system influenced the users while evaluating other parameters of the system such as the satisfaction of the users' needs.

References

1. Malaka, R., Zipf, A.: Deep map-challenging IT research in the framework of a tourist information system. In: Proceedings of International Congress on Tourism and Communication Technologies in Tourism, pp. 15–27 (2000)
2. Laukkanen, M., Helin, H., Laamanen, H.: Tourists on the move. In: Klusch, M., Ossowski, S., Shehory, O. (eds.) CIA 2002. LNCS (LNAI), vol. 2446, p. 36. Springer, Heidelberg (2002)
3. Schmidt-Belz, B., Poslad, S., Nick, A., Zipf, A.: Personalized and location based mobile tourism services. In: Workshop on Mobile Tourism Support Systems, in conjunction with the Fourth International Symposium on Human Computer Interaction with Mobile Devices, pp. 18–20 (2002)
4. Gervais, E., Hongsheng, L., Nussbaum, D., Roh, Y.-S., Sack, J.-r., Yi, J.: Intelligent map agents – An ubiquitous personalized GIS. Phtogrammetry & Remote Sensing 62, 347–365 (2007)
5. Kabassi, K., Charou, E., Martinis, A.: Implementation Issues of a Knowledge-based Geographical Information System. Knowledge-Based Software Engineering. In: Frontiers in Artificial Intelligence and Applications (Proceedings of the 8th Joint Conference on Knowledge-Based Software Engineering JCKBSE 2008) (2008a)

6. Kabassi, K., Virvou, M., Charou, E., Martinis, A.: Software Life-cycle for an Adaptive Geographical Information System. In: International Conference on Signal Processing and Multimedia Applications (SIGMAP 2008), pp. 393–396 (2008b)

7. McTear, M.F.: Intelligent interface technology: from theory to reality? Interacting with computers 12, 323–336 (2000)

8. Chin, D.N.: Empirical Evaluation of User Models and User-Adapted Systems. User Modeling and User Adapted Interaction 11(1/2), 181–194 (2001)

9. Micarelli, A., Sciarrone, F.: Anatomy and Empirical Evaluation of an Adaptive Web-based Information Filtering System. User Modeling and User-Adapted Interaction 14, 159–200 (2004)

10. Cheng, R., Vassileva, J.: Design and evaluation of an adaptive incentive mechanism for sustained educational online communities. User Modeling and User Adapted Interaction 16, 321–348 (2006)

11. Chou, S.-Y., Chang, Y.-H., Shen, C.-Y.: A fuzzy simple addtive weighting system under group decision-making for facility location selection with objective/subjective attributes. European Journal of Operational Research 189, 132–145 (2008)

12. Zadeh, L.A.: Fuzzy sets. Information and Control 8, 338–353 (1965)

13. Keufmann, A., Gupta, M.M.: Introduction to Fuzzy Arithmetic: Theory and Application. Van Nostrand Reinhold, New York (1991)

14. Dubois, D., Prade, H.: Operations on fuzzy numbers. International Journal of Systems Science 9, 613–626 (1978)

15. Fishburn, P.C.: Additive Utilities with Incomplete Product Set: Applications to Priorities and Assignments. Operations Research (1967)

16. Hwang, C.L., Yoon, K.: Multiple Attribute Decision Making: Methods and Applications. Lecture Notes in Economics and Mathematical Systems, vol. 186. Springer, Heidelberg (1981)

17. Virvou, M., Alepis, E.: Mobile educational features in authoring tools for personalised tutoring. Computers and Education 44(1), 53–68 (2005)

18. Virvou, M., Katsionis, G., Manos, K.: Combining software games with education: Evaluation of its educational effectiveness, Educational Technology & Society. Journal of International Forum of Educational Technology & Society and IEEE Learning Technology Task Force (2005)

19. Virvou, M., Kabassi, K.: Experimental Studies within the Software Engineering Proc-ess for Intelligent Assistance in a GUI. Journal of Universal Computer Science 8(1), 51–85 (2003)

20. Picard, R.W.: Affective Computing: Challenges. Int. Journal of Human-Computer Studies 59(1-2), 55–64 (2003)

21. Pantic, M., Rothkrantz, L.J.M.: Toward an affect-sensitive multimodal human-computer interaction. Proceedings of the IEEE 91, 1370–1390 (2003)

22. Alepis, E., Virvou, M., Kabassi, K.: Recognition and generation of emotions in affective e-learning. In: ICSOFT 2009 Proceedings of 4th International Conference on Software and Data Technologies, vol. 2, pp. 273–280 (2009)

23. Alepis, E., Stathopoulou, I.-O., Virvou, M., Tsihrintzis, G.A., Kabassi, K.: Audio-lingual and visual-facial emotion recognition: Towards a bi-modal interaction system. In: Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI, vol. 2, Article number 5670096, pp. 274–281 (2010)

Performance Evaluation of Adaptive Content Selection in AEHS

Pythagoras Karampiperis[1,2] and Demetrios G. Sampson[1,3]

[1] Department of Digital Systems, University of Piraeus, Greece
[2] Institute of Informatics and Telecommunications,
National Center of Scientific Research "Demokritos", Athens, Greece
`pythk@ieee.org`
[3] Informatics and Telematics Institute, Centre for Research and Technology Hellas, Greece
`sampson@unipi.gr`

Abstract. Adaptive content selection is recognized as a challenging research issue in adaptive educational hypermedia systems (AEHS). Several efforts have been reported in literature aiming to support the Adaptation Model (AM) design by providing AEHS designers with either guidance for the direct definition of adaptation rules, or semi-automated mechanisms which generate the AM via the implicit definition of such rules. The goal of the semi-automated, decision-based approaches is to generate a continuous decision function that estimates the desired AEHS response, aiming to overcome the insufficiency and/or inconsistency problems of the defined adaptation rule sets. Although such approaches bare the potential to provide efficient AMs, they still miss a commonly accepted framework for evaluating their performance. In this paper, we propose an evaluation framework suitable for validating the performance decision-based approaches in adaptive learning objects selection in AEHS and demonstrate the use of this framework in the case of our proposed decision-based approach for estimating the desired AEHS response.

1 Introduction

Adaptive Educational Hypermedia Systems (AEHS) have been proposed as the underlying facilitator for personalized web-based learning with the general aim of personalizing learning experiences for a given learner [1]. In order to adaptively select and sequence learning objects in AEHS, the definition of adaptation behaviour, referred to as Adaptation Model, is required. The Adaptation Model (AM) contains the rules for describing the runtime behaviour of the AEHS. Typically, these rules include *Concept Selection Rules* which are used for selecting appropriate concepts from the Domain Model to be covered, *Content Selection Rules* which are used for selecting appropriate resources from the Media Space, as well as, *Sequencing Rules* which are used for generating appropriate learning paths (sequences of learning objects) for a given learner [2].

In the literature, there exist different approaches aiming to support the Adaptation Model design by providing AEHS designers with either guidance for the direct definition of adaptation rules, or semi-automated mechanisms which generate the AM via

the implicit definition of such rules [3], [4]. The main drawback of the direct definition of adaptation rules is that there can be cases during the run-time execution of AEHS where no adaptation decision can be made due to insufficiency and/or inconsistency of the defined adaptation rule sets [5]. This is due to the fact that, even if appropriate resources exist in the Media Space, the absence of a required rule (insufficiency problem) or the conflict between two or more rules (inconsistency problem), prevents the AEHS to select and use them in the generated learning resource sequence. The goal of the semi-automated approaches is to generate a continuous decision function that estimates the desired AEHS response, overcoming the above mentioned problems [6]. Although such approaches bare the potential to provide efficient Adaptation Models, they still miss a commonly accepted framework for evaluating their performance.

In this paper, we propose an evaluation framework suitable for validating the performance decision-based approaches in adaptive learning objects selection in AEHS and demonstrate the use of this framework in the case of our proposed decision-based approach for estimating the desired AEHS response. In [3], we have presented a semi-automated approach for extracting the content selection rules of the AM in AEHS. In this paper, we use this approach to demonstrate and verify the capacity of the proposed evaluation framework.

The paper is structured as follows: First, we discuss the performance evaluation metrics that have been proposed by the literature for validating the use of decision-based approaches. Then, we present the proposed performance evaluation methodology for decision-based content selection approaches in AEHS, and set up and report simulation-based experiments, following the above mentioned methodology, which aim to validate these evaluation metrics within the framework of our previously proposed method for estimating the desired AEHS response. Finally, we discuss our findings and the conclusions that can be offered.

2   Performance Evaluation Metrics for Decision-Based AEHS

In adaptive content selection several approaches have been proposed by the literature. The most commonly used are the following [7]:

− *Concept/Keyword-based Selection*. In these approaches, searching is performed based on a set of keywords, typically representing the desired concepts to be covered from the retrieved learning objects. In AEHS, these keywords are defined over the Domain Concept Ontology during the concept selection process. In this case, the ranking of learning objects is performed using a concept/keyword-based similarity formula [8].
− *Preference-based Selection*. In these approaches, selection is performed based on the comparison of the learner profile in hand with the metadata description of the learning objects. In this case, the ranking of learning objects is performed using a preference score [6], [9], which evaluates the utility/suitability of each learning object for the learner profile in hand.

In both techniques general purpose evaluation metrics are used from the field of information extraction [10]. More specifically, precision and recall measures are applied in order to evaluate the effectiveness of the learning objects selection technique, in terms of accuracy and completeness respectively. Precision is the ratio of correct responses to the sum of correct and incorrect responses. Recall is the number of correct system responses divided by the sum of correct, incorrect and missing system responses. In order to have a single evaluation metric, F-measure is used, which is a weighted combination of recall and precision.

However, AEHS implement a content selection strategy which limits the number of retrieved learning objects, aiming to restrict the amount of information provided to learners at a given time instance, due to the problem of learners' cognitive overload [11]. As a result, the precision should be measured not on the entire Media Space, but only on the desired sub-space which represent a set of the n most preferred learning objects, where n is the number of the desired learning objects. If not, the resulting precision would be higher or equal to the real one, since the number of retrieved learning objects is less or equal to the number of desired learning objects at a given time instance. Moreover, since the resulting LO space is restricted, the recall measure should also be measured over the space of the n most relevant learning objects, and not over the space of all relevant learning objects. This introduces the need for an alternative evaluation metric in adaptive content selection. In this paper we propose an evaluation framework, presented in next section, which uses a variation of the evaluation metric defined in [3]. The new metric is defined as follows:

$$\text{Selection Success} \ (\%) = 100 * \left(\frac{correct \ ranked \ Learning \ Objects \ selected}{requested \ Learning \ Objects} \right)$$

In this paper, we present simulation based experiments verifying this hypothesis, and demonstrate that the proposed metric is harder than the precision metric, since it measures the precision over a smaller value space.

3   Evaluation Methodology for Decision-Based AEHS

The underlying hypothesis of the design of a decision-based approach for content selection in AEHS is that it is feasible to construct a semi-automated algorithm, which generates a continuous decision function that estimates the desired AEHS response. Thus, the goal of evaluating such an approach is twofold: first, to examine whether a proposed semi-automated decision based approach is capable of extracting decision models which replicate the Adaptation Model (AM) of existing AEHS; and second, to verify via performance evaluation that this approach can be applied in cases where large-scale adaptation rule sets are needed to describe the desired AEHS response. To this end, the evaluation should be performed in two phases:

− *Phase A: Extracting the AM of existing AEHS.* In this evaluation phase, the Adaptation Model (AM) rules of existing AEHS are used for generating sample adaptation decisions. These decisions have the form of combinations of learning objects mapped to learner profiles, and are used to train the intelligent mechanism that fits the

response function on these data. The goal of this phase is to examine whether the proposed semi-automated decision based approach is capable of extracting the decision model of the AEHS in hand. In our experiments, we will try to extract the AM rules for content selection used in the INSPIRE [12] system.

−   *Phase B: Scaling up the experiments.* As already discussed, the problem of defining adaptation rules is a combinatorial problem, which means that in order to design sufficient and consistent adaptation rule sets, all the combinations of the adaptation decision variables should be covered. However, these combinations can be millions [3], leading to huge rule sets that is difficult to author, manage and verify their sufficiency and/or consistency. To this end, in order to keep the adaptation rule set human-maintainable, existing AEHS in the literature use few adaptation variables, typically 2-4 variables for describing learners' behaviour and 2-3 variables for describing educational content. The goal of this evaluation phase is to verify that the proposed approach can be applied in cases where large-scale adaptation rule sets are needed to describe the desired AEHS response. In order to do this, we simulate the existence of an AEHS that uses as many adaptation variables as possible. These variables (learner profile properties and educational description model properties) are selected from the elements of wide-spread Learning Technology standards. However, special attention should be given in generating learner profiles and educational content metadata records that simulate real-life conditions. Details on how such datasets are generated are given in next section.

4   Setting Up the Experiments

As already described, the adaptation model design is the process of defining (a) the concept selection rules which are used for selecting appropriate concepts from the *Domain Model* to be covered, (b) the content selection rules which are used for selecting appropriate resources from the *Media Space*, and (c) the sequencing rules which are used for generating appropriate learning paths (sequences of learning objects) for a given learner, based on learner's profile stored in the *Learner Model*. In this paper, we focus on content selection rules, thus, before executing our experiments for measuring the performance of adaptive selection of learning objects, we need to design (a) the Media Space, and (b) the Learner Model.

4.1   Designing the Media Space

In the first phase of the evaluation, we will extract the AM of the INSPIRE system [12]. The INSPIRE system uses two variables in the Educational Resource Description Model, namely, the Performance Level and the Learning Resource Type. In the second evaluation phase, we simulate the existence of an AEHS where large-scale adaptation rule sets are needed to describe the desired AEHS response. To do so, we have used as Educational Resource Description Model a subset of the IEEE Learning Object Metadata standard elements [13].

   In both evaluation phases, we need to simulate real-life conditions. This means that the simulated learning object metadata records should have a distribution over their value spaces similar to the metadata value distribution found in real-life learning

object repositories. Najjar and Duval [14], presented a statistical analysis of the actual use of IEEE LOM metadata elements in the ARIADNE learning object repository. The results were derived from analyzing the empirical data (usage logs) of 3,700 ARIADNE metadata instances.

Table 1. Usage Percentage of Data Elements in ARIADNE Repository

IEEE LOM Element	Value Provided (%)	Most used Vocabulary value (M)	% of M (filled-in)	%M among all cases
Aggregation Level	91.9	Lesson	92.7	85.2
Context	53.5	University Degree	69.7	37.2
Interactivity Level	53.2	Medium	67.7	36.1
Semantic Density	52.4	Medium	76.4	40.0
Difficulty Level	52.2	Medium	72.8	38.0

Table 1 shows the percentage of times each ARIADNE data element was filled in by indexers during the indexing process. From the data shown in Table 1, we notice that only one data element is almost always used: the Aggregation Level element. Other elements are used in about 50 % of the descriptions. For the values of data elements, we can see that indexers often use just one value. As a result, in order to simulate in our experiments the metadata of a real-world repository, we will generate metadata records with normal distribution over the metadata elements value space, simulating that not all metadata elements and their corresponding vocabulary terms are used equally. Normal distribution is a continuous probability distribution that is often used to describe random variables that tend to cluster around a single mean value.

4.2 Designing the Learner Model

In the first phase of the evaluation, we will extract the AM of the INSPIRE system [12]. The INSPIRE system uses two variables in the Learner Model, namely, the Learner's Knowledge Level and the Learner's Learning Style). In the second evaluation phase, we simulate the existence of an AEHS where large-scale adaptation rule sets are needed to describe the desired AEHS response. To do so, for the design of the Learner Model in our simulations we have used an overlay model for representing the Learners Knowledge Space and a stereotype model for representing learners' preferences. More precisely, for the learners' knowledge level we assume the existence of a related certification for each node of the Learners Knowledge Space, the evaluation score in testing records and the number of attempts made on the evaluation. For modelling of learners' preferences we use learning styles according to Honey and Mumford model [15], as well as modality preference information consisting of four modality types, namely, the visual modality, the textual modality, the auditory modality and any combination of the three modality preferences [16]. Each element of the Learner Model was mapped to the IMS Learner Information Package specification [17].

In order to simulate in our experiments the profiles of real learners we generated profile records using truncated standard lognormal distribution with [sigma]=1 and reduced by factor 1/5. This distribution is often used in the literature for simulating learner behaviour [18].

4.3 Simulating the AM of an AEHS

The goal of our experiments is to evaluate the suitability of the set of performance evaluation metrics, presented in the previous section, for validating the use of decision-based approaches for adaptive learning objects selection in AEHS, and assess the use of these metrics in the case of our previously proposed statistical method for estimating the desired AEHS response.

Performance evaluation in this context means measuring (a) how well our semi-automated approach fits the decision function to the provided model adaptation decisions (training data), and (b) how well this decision function responds to decision cases not known during the training process (generalisation capacity).

As a result, we need to produce model adaptation decisions and compare them with the corresponding response of our decision-based approach. Some of these model adaptation decisions will be used for training our method, and some will be used for measuring its' generalisation capacity. In the first evaluation phase, the Adaptation Model (AM) rules of an existing AEHS are used for generating sample adaptation decisions. In the second evaluation phase, we need to simulate the existence of an AEHS that uses as many adaptation variables as possible. Since such an AEHS does not exist, we will simulate model adaptation decisions via the use of simulated instructional designers' preference models. These models have been selected in such a way that the preference surface is complex, thus, it would be a difficult task for the decision based algorithm to fit the training data.

To achieve this, we use as an instructional designers' preference model a multi-variable function, with 18 variables (k). These variables model the eleven (11) elements of the Educational Resource Description Model in use (that is, the elements used from the "General" and the "Educational" IEEE LOM categories) and the seven (7) elements of the Learner Model in use [10]. We assume that the response of this function expresses the utility of a given learning object for a given learner profile (preference-based selection problem). In our experiments, we simulate the preference models of five (5) instructional designers, using the functions presented in Table 2.

Table 2. Multi-variable functions used as simulated instructional designers' preference models

Rosenbrock function	$f(x) = \sum_{i=1}^{k-1} \left[100 \cdot \left(x_{i+1} - x_i^2 \right)^2 + \left(1 - x_i \right)^2 \right]$
Schwefel function	$f(x) = \sum_{i=1}^{k} \left[-x_i \cdot \sin\left(\sqrt{|x_i|} \right) \right]$
Rastrigin function	$f(x) = 10 \cdot k + \sum_{i=1}^{k} \left[x_i^2 - 10 \cdot \cos\left(2 \cdot \pi \cdot x_i \right) \right]$
Griewangk function	$f(x) = \sum_{i=1}^{k} \frac{x_i^2}{4000} - \prod_{i=1}^{k} \cos\left(\frac{x_i}{\sqrt{i}} \right) + 1$
Sum of different powers function	$f(x) = \sum_{i=1}^{k} |x_i|^{(i+1)}$

For evaluating the performance, we have generated a set of 1.000 learning object metadata records and a set of 100 learner profiles. In each experiment, 50% of the available learning objects metadata records, randomly selected, were used for algorithmic training and the rest 50% for measuring the generalisation, that is, the estimation capacity, of the algorithm. Similarly, in each experiment 50% randomly selected of the available learner profiles were used for algorithmic training and the rest 50% for measuring the generalisation of the algorithm.

5   Experimental Results and Discussion

In this section, we present experimental results from the execution of the above mentioned evaluation methodology for the case of our previously proposed statistical method for estimating the desired AEHS response [3]. The results are presented per evaluation phase.

5.1   Extracting the AM of existing AEHS

Our first experiment was the application of our decision-based approach for replicating the Adaptation Model (AM) of an existing AEHS. To this end, we simulated the AM of the INSPIRE [[HYPERLINK \l "Pap03" 12], produced sample adaptation rules in the form of combinations of learning objects mapped to learner profiles, and applied our methodology to extract the AM. The INSPIRE system uses two variables from the Learner Model (namely, the Learner's Knowledge Level and the Learner's Learning Style) and two variables from the Educational Resource Description Model (namely, the Performance Level and the Learning Resource Type), for performing adaptation decisions according to Table 3.

Table 3. INSPIRE Adaptation Model Rules

Learner Attributes		Proposed Learning Objects	
Knowledge Level	Inadequate (1)	Performance Level	Remember (1)
	Mediocre (2)	Performance Level	Use (2)
	Advanced (3)	Performance Level	Find (3)
	Proficient (4)	Performance Level	-
Learning Style	Activist (1)	Learning Resource Type	Activity-oriented (1)
	Reflector (2)	Learning Resource Type	Example-oriented (2)
	Theorist (3)	Learning Resource Type	Theory-oriented (3)
	Pragmatist (4)	Learning Resource Type	Exercise-oriented (4)

Figure 1, presents the INSPIRE's AM dependencies of the Learning Style and Learning Resource Type in the LO utility space, whereas Figure 2 presents the same dependencies of the produced AM when our decision based approach is applied. From these figures we can observe that the produced adaptation model is a super class of the INSPIRE's AM, since it contains more adaptation rules (dependencies between learning object and learner characteristics). Moreover, we can observe that the produced AM has a continuous contour in the Utility Space, which means that this AM has the ability to always propose learning objects.
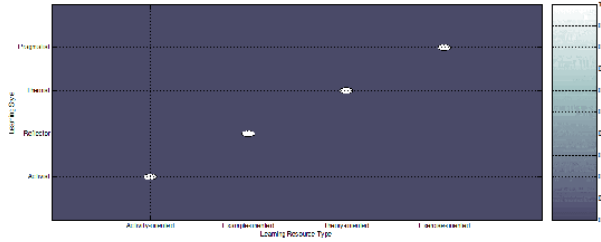
Fig. 1. INSPIRE: Learning Style and Learning Resource Type Utility Space
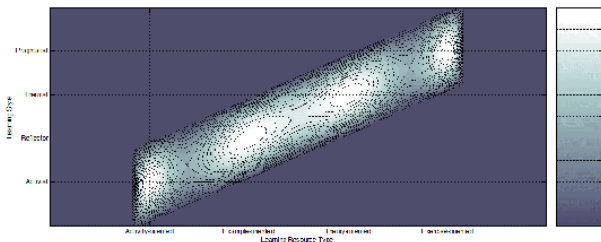
Fig. 2. Generated Learning Style and Learning Resource Type Utility Space from INSPIRE

After the above experiment, the research question was to investigate if the proposed decision based approach has the capacity of learning more complex Adaptation Models, consisting of many free variables with complex preference surfaces, thus, it would be a difficult task for the decision based algorithm to fit the training data. This is the goal of the second evaluation phase, which is presented in next section.

5.2 Scaling Up the Experiments

Before proceeding with the performance evaluation of our decision-based approach, we have conducted an additional experiment, aiming to assess the use of the performance evaluation metrics proposed by the literature.

Our semi-automated approach for adaptive content selection uses a preference-based learning objects selection mechanism based on the use of a suitability function that estimates the utility of a given learning object for a given learner [3].

In order to compare the performance evaluation metrics discussed in previous section, we evaluate the performance using randomly generated datasets which serve as model adaptation decisions and vary in size. The size of these datasets depends on the number of ranked learning objects for a given number of learner profiles. In real conditions, these rankings would be requested from an instructional designer. In our experiments, these rankings are the result of the application of the simulated instructional designers' preference models presented in Table 4. As already described, the datasets were divided into two subsets: the training dataset, which was used for algorithmic training and for evaluating the performance during the training process, and
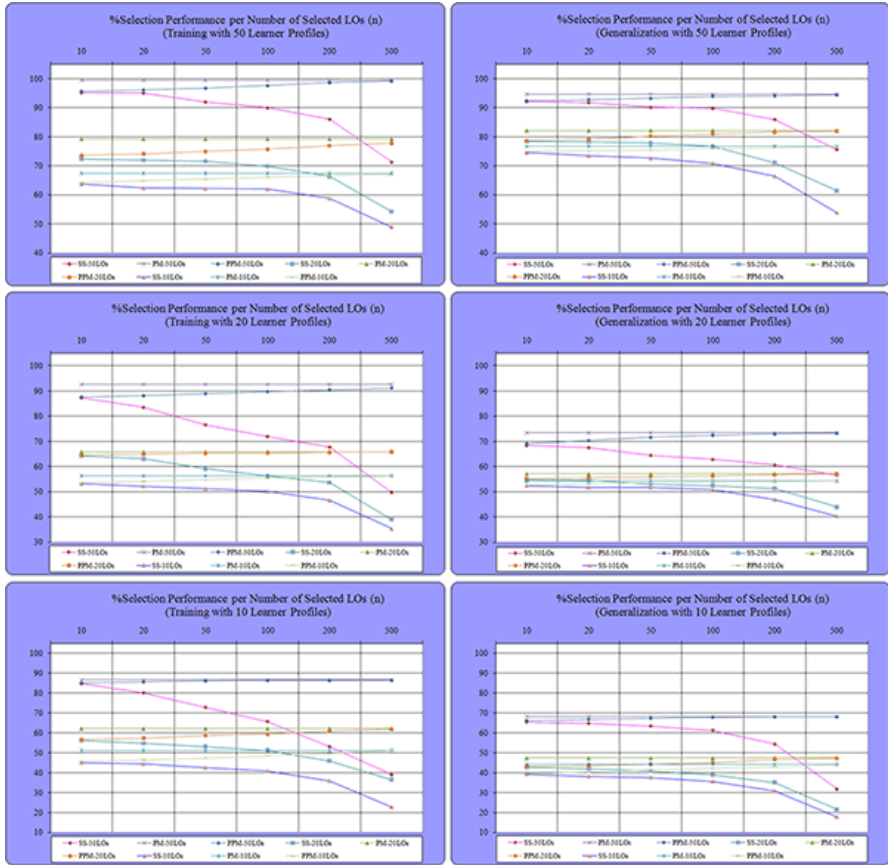
Fig. 3. Adaptive Selection of Learning Objects. Left column presents training results, whereas, right column presents generalization results

the generalization dataset, which was used for measuring the generalization capacity of the algorithm. Each experiment was executed 100 times using a randomly selected instructional designers' preference model.

Figure 3 (left column) presents average selection performance results during algorithmic training, when using different simulation parameters regarding the number of learner profiles and the number of learning object metadata records used. In each experiment, the selection performance was measured when using different values of the parameter n (varying from 10 to 500), which expresses the maximum number of requested learning objects from the Media Space. In this figure the performance evaluation was measured using the typical Precision Metric (PM), the proposed alternative metric for Selection Success (SS), as well as, by applying the PM metric only on the desired sub-space of the Media Space (Partial Precision Metric, PPM). From these results we observe the following:

a) Precision when measured with PM metric is independent from the maximum number of requested learning objects from the Media Space (selection space), as well as, from the ranking of the selected learning objects.
b) Precision when measured with PPM metric is independent from the ranking of the selected learning objects, but depends on the volume of the selection space.
c) The PPM metric tends to be equal to the PM metric when the selection space becomes bigger (n increases).
d) Performance evaluation using the PM metric is higher or equal to the performance when using the PPM metric. Also performance evaluation using the PM metric is higher or equal to the performance when using the SS metric.
e) The SS metric tends to be lower as the searching space increases, whereas PPM metric becomes higher as the searching space increases. This is due to the fact that, when the searching space increases the probability of introducing ranking errors also increases. Since the PPM metric is not dependent by the ranking of the selected learning objects, the PPM metric behaves differently from the SS metric.

The same observations apply also when measuring the generalization capacity, as depicted in Figure 3 (right column). These observations verify the hypothesis that by definition the SS metric is harder than the PM or the PPM metric, which means that in the case of AEHS, where the ranking of the selected learning objects is critical, the SS metric should be used.

6   Conclusions

In this paper, we propose an evaluation framework suitable for validating the performance decision-based approaches in adaptive learning objects selection in AEHS and demonstrate the use of this framework in the case of our proposed decision-based approach for estimating the desired AEHS response. More precisely, we discussed the limitations of the performance metrics used by the literature for the problem of adaptive content selection, introduced the need for an alternative evaluation metric and presented a metric, which although seems similar to the precision metric in information retrieval systems, its difference is critical. This metric evaluates the precision of selecting learning objects not on the entire space of the Media Space, but only on the desired sub-space, and also it takes into consideration the ranking of the selection process.

References

[1] Knutov, E., De Bra, P., Pechenizkiy, M.: AH 12 years later: a comprehensive survey of adaptive hypermedia methods and techniques. New Review of Hypermedia and Multimedia 15(1), 5–38 (2009)
[2] Henze, N., Nejdl, W.: A Logical Characterization of Adaptive Educational Hypermedia. New Review of Hypermedia and Multimedia (NRHM) 10(1), 77–113 (2004)
[3] Karampiperis, P., Sampson, D.G.: Adaptive Learning Resources Sequencing in Educational Hypermedia Systems. Educational Technology & Society 8(4), 128–147 (2005)

[4] Ras, E., Ilin, D.: Using decision models for the adaptive generation of learning spaces. In: Nejdl, W., Kay, J., Pu, P., Herder, E. (eds.) AH 2008. LNCS, vol. 5149, pp. 153–162. Springer, Heidelberg (2008)

[5] Brusilovsky, P., Wade, V., Conlan, O.: From Learning Objects to Adaptive Content Services for E-Learning. In: Architecture Solutions for E-Learning Systems, pp. 243–261. Idea Group Inc, USA (2007)

[6] Karampiperis, P., Sampson, D.G.: Adaptive Learning Object Selection in Intelligent Learning Systems. Journal of Interactive Learning Research, Special Issue on Computational Intelligence 15(4), 389–409 (2004)

[7] Sampson, D., Karampiperis, P.: Decision Models in the Design of Adaptive Educational Hypermedia Systems. In: Graf, S., et al. (eds.) Intelligent and Adaptive Learning Systems: Technology Enhanced Support for Learners and Teachers. IGI Global (2011)

[8] Biletskiy, Y., Baghi, H., Keleberda, I., Fleming, M.: An adjustable personalization of search and delivery of learning objects to learners. Expert Systems with Applications 36(5), 9113–9120 (2009)

[9] Dolog, P., Simon, B., Nejdl, W., Klobucar, T.: Personalizing Access to Learning Networks. ACM Transactions on Internet Technology 8(2), p. Article 8 (2008)

[10] Ochoa, X., Duval, E.: Relevance Ranking Metrics for Learning Objects. IEEE Transactions on Learning Technologies 1(1), 34–48 (2008)

[11] Brusilovsky, P.: Adaptive navigation support. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) Adaptive Web 2007. LNCS, vol. 4321, pp. 263–290. Springer, Heidelberg (2007)

[12] Papanikolaou, K., Grigoriadou, M., Kornilakis, H., Magoulas, G.: Personalising the Interaction in a Web-based Educational Hypermedia System: the case of INSPIRE. International Journal of User Modeling and User-Adapted Interaction 13(3), 213–267 (2003)

[13] IEEE. IEEE Learning Object Metadata Standard (IEEE LOM), IEEE P1484.12.1. (2002), `http://ltsc.ieee.org/wg12/`

[14] Najjar, J., Duval, E.: Actual Use of Learning Objects and Metadata: An Empirical Analysis. IEEE Technical Committee on Digital Libraries Bulletin (TCDL) 2(2) (2006)

[15] Honey, P., Mumford, A.: The manual of Learning Styles.: Peter Honey, Maidenhead (1992)

[16] Razmerita, L.: User modeling and personalization. In: Adaptable and Adaptive Hypermedia Systems. Idea Group Inc., USA (2005)

[17] IMS, IMS Global Learning Consortium Inc., Learner Information Package (LIP) Final Specification v1.0 (2001)

[18] McCalla, G.: Smart recommendation for an evolving e-learning system: architecture and experiment. International Journal on E-Learning 4(1), 105–129 (2005)

AFOL: Towards a New Intelligent Interactive Programming Language for Children

Efthimios Alepis

Department of Informatics, University of Piraeus,
80 Karaoli & Dimitriou St., 18534 Piraeus, Greece
`talepis@unipi.gr`

Abstract. This paper introduces a new programming language for children named AFOL. The creation of the AFOL programming language has been based on the idea of the well-known Logo programming language. However, AFOL extends Logo's basic programming concepts such as sequential and functional programming by introducing the more modern concepts of Object Oriented programming. Furthermore, AFOL incorporates highly sophisticated user interaction mechanisms, namely affective interaction through emotion recognition and through the use of animated tutoring agents. In this way, the interaction of children is more user-friendly and constructive since children learn programming and have fun at the same time.

1   Introduction

The well-known "Logo" programming language was introduced in 1967 (Frazier, 1967). The Logo developers' main objective was to take the best practices and ideas from computer science and computer programming and produce an interface that was good and suitable for the education of young children. Hence, the authors of Logo aimed to create a friendly programming language for the education of children where they could learn programming by playing with words and sentences. The first implementation of Logo was written in LISP for the purposes of creating a programming language as a math land where kids could play by giving commands that produced nice and colorful drawings. Logo programming language may be seen as a compromise between a sequential programming language with block structures, and a functional programming language. Logo has been used mainly in the past as a teaching language for children but its list handling facilities made it remarkably useful for producing useful scripts. A detailed study on the "Logo" programming language from its early stages and also recent work on Logo-derived languages and learning applications can be found in (Feurzeig, 2010).

Modern programming languages try to provide as much user-friendliness as possible while retaining their full programming functionality. Hudlicka (Hudlicka, 2003) points out that an unprecedented growth in human-computer interaction has led to a redefinition of requirements for effective user interfaces and that a key component of these requirements is the ability of systems to address affect. Learning a programming language is a complex cognitive process and it is argued that how people feel may play an important role on their cognitive processes as well (Goleman, 1981). At the

same time, many researchers acknowledge that affect has been overlooked by the computer community in general (Picard and Klein, 2002). A remedy in the problem of effectively teaching children through educational applications may lie in rendering computer assisted e-learning systems more human-like and thus more affective. To this end, the incorporation of emotion recognition components as well as the incorporation of animated tutoring agents in the user interface of the educational application can be quite useful and profitable (Elliott et al., 1999). Indeed, the presence of animated, speaking agents has been considered beneficial for educational software (Johnson et. al., 2000, Lester et. al., 1997).

In the past, the author of this paper has participated in the development of prototype systems that incorporate emotion recognition modules, based on artificial intelligence techniques and multi-attribute decision making approaches (Alepis et al. 2009, Alepis et al. 2010). The resulting systems showed significant efficiency in recognizing and reproducing emotions.

After a thorough investigation in the related scientific literature we found that there is a shortage of educational systems that incorporate multi-modal emotion recognition, while we did not find any existing programming languages that incorporate emotion recognition and/or emotion generation modules. Perhaps the most relevant work is that of Kahn (Kahn, 1996), where an animated programming environment for children is described. The author of this paper has developed a programming interface called ToonTalk in which the source code is animated and the programming environment is a video game. The aim of this project was to give children the opportunity to build real programs in a manner that was easy to learn and fun to do. However, this approach did not incorporate any affective interaction capabilities.

In view of the above, this paper presents a new programming language for children, which is highly interactive and intelligent since it provides affective interaction during programming. The programming language is called AFOL which is the acronym for "Affective Object Oriented Logo Language". In the implementation of the AFOL language there is an added programming dimension that of object oriented programming (Pastor et al., 2001, Alepis & Virvou, 2011). Through the language's object-oriented architecture, an AFOL program may thus be viewed as a collection of interacting objects, as opposed to the conventional Logo model, in which a program is seen as a list of tasks. Furthermore, the initial goal to create a programming language suitable for the needs and the limitations of children is further improved through the AFOL language, by a highly user-friendly user interface, designed for affective interaction between computers and children. The incorporation of emotion recognition capabilities, as well as the presence of speaking animated emotional tutoring agents both introduce a novelty in the area of computer assisted programming language learning.

2   General Architecture of the AFOL Programming Environment

In this section, we describe the overall functionality and emotion recognition features of AFOL. The architecture of AFOL consists of the main educational programming environment, a user monitoring component, emotion recognition inference mechanisms and a database. Part of the database is used to store data related to the

programming language. Another part of the database is used to store and handle affective interaction related data.. The programming environment's architecture is illustrated in figure 1.
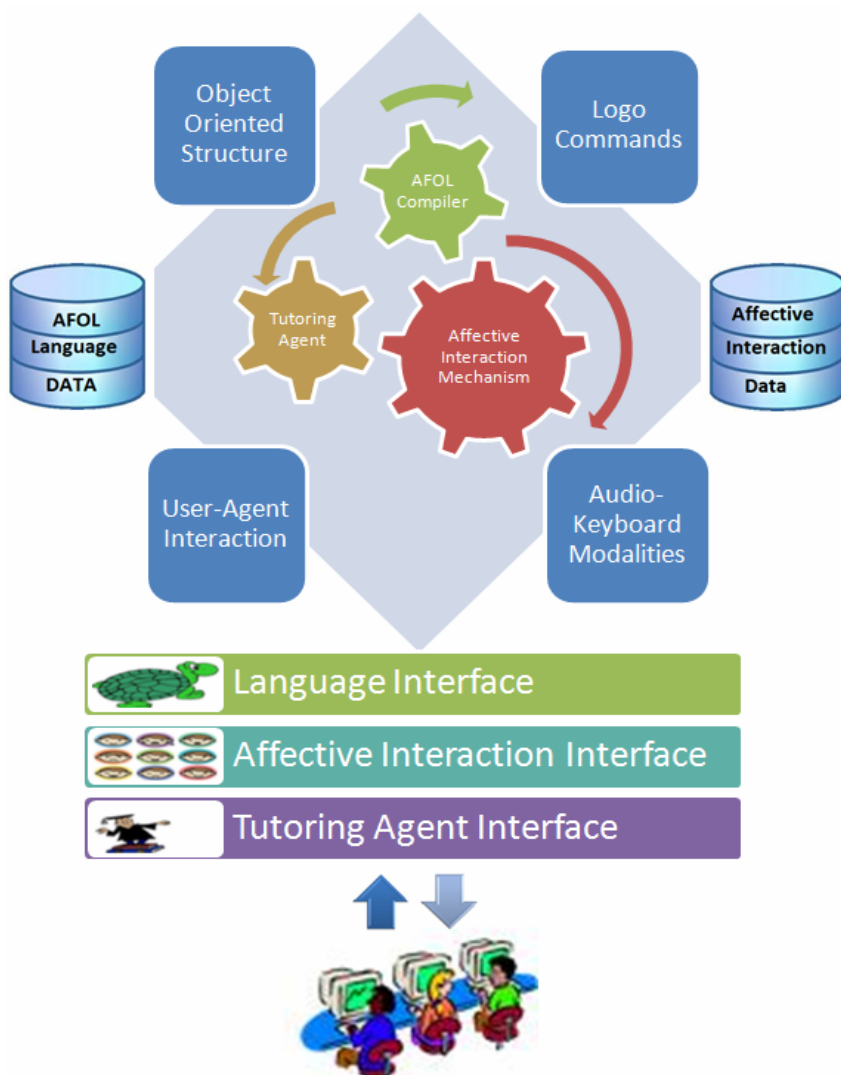
Fig. 1. General architecture of the AFOL programming language

As we can see in figure 1, the students' interaction can be accomplished either orally through the microphone, or through the keyboard/mouse modality. The educational systems consists of three subsystems, namely the affective interaction subsystem, the programming language's compiler subsystem and the subsystem that reasons for and handles the animated agent's behaviour.

3   Overview of the AFOL Programming Learning System

While using the educational application from a desktop computer, children as students have the oportunity of being taught particular programming courses. The information is given in text form while at the same time an animated agent reads it out loud using a speech engine. Students are prompted to write programming commands and also programs in the AFOL language in order to produce drawings, shapes and particular objects. The main application is installed either on a public computer where all students have access, or alternatively each student may have a copy on his/her own personal computer. Two examples of using the AFOL's programming interface are illustrated in figures 2 and 3. The animated agent is present in these modes to make the interaction more human-like. The tutoring agent would also try to sympathise with the aims of each user as student and would try to elicit human emotions.

Figure 1 illustrates a snapshot from user interaction where a user is typing programming commands that contain pre-stored objects in order to produce a complicated drawing. In figure 2 we can see a tutoring character that congratulates a user for producing a quite complicated drawing by using the programming languages commands and object oriented features. While the users interact with the affective system, both their oral and keyboard actions are recorded in order to be interpreted in
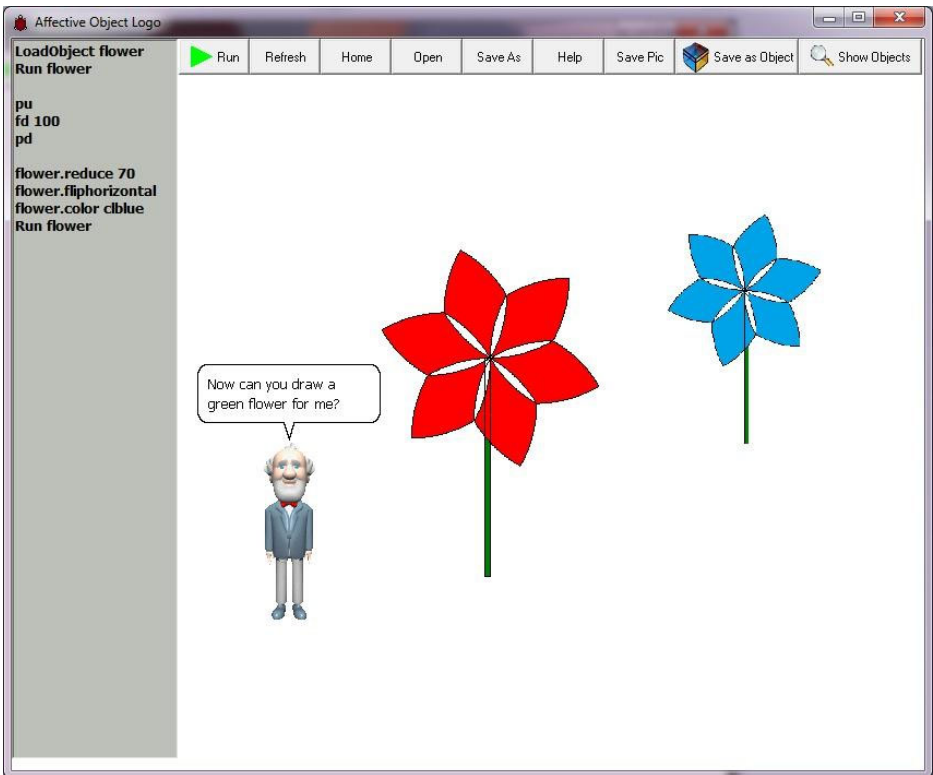
Fig. 2. A user is typing programming commands to create a specific drawing
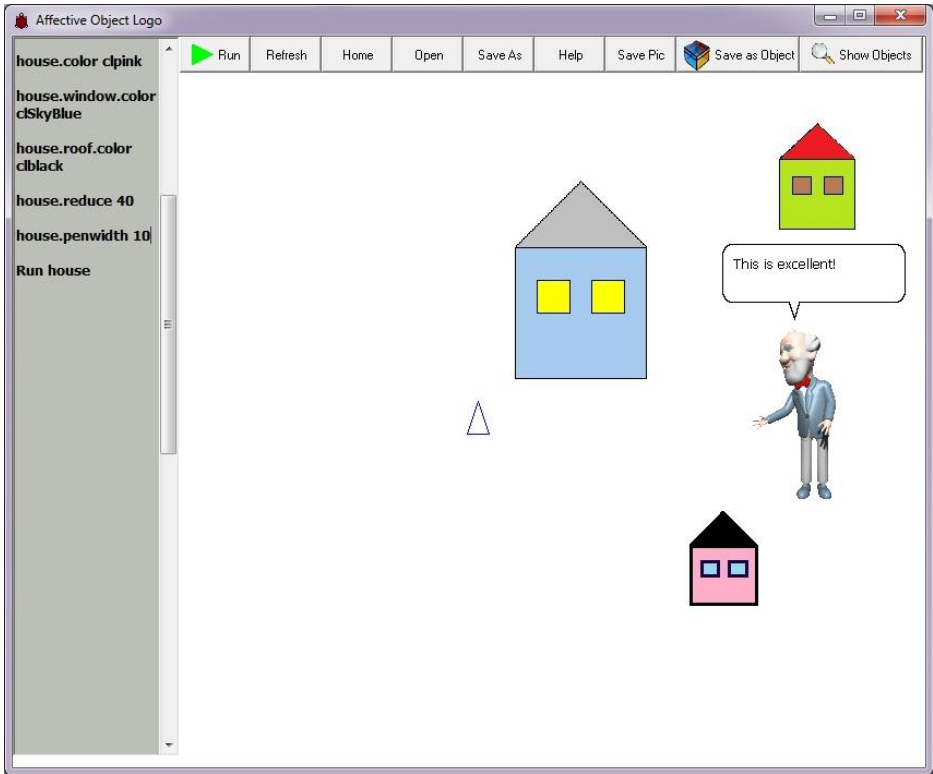
Fig. 3. The Tutoring character congratulates a user

emotional interaction terms. A more complicated description of the affective module is beyond the scopes of this paper that aims basically in the presentation of the new programming environment. However, the overall functionality of these modules which lead to multi-modal recognition of emotions through multimodal data, can be found in (Tsihrintzis et al. 2008) and in (Alepis & Virvou, 2006). The results from these studies were quite promising and encouraged as to test this approach to a more demanding domain that belongs to the area of teaching programming to young children.

4   AFOL Language Commands and Object Oriented Structure

In this section we give an overview of the AFOL system's supported programming commands, as well as the functionality that derives from the language's Object Oriented perspective. In the implementation of the AFOL system we have used the Logo's well-known feature which is the turtle. The turtle is an on-screen cursor, which can be given movement and drawing instructions, and is used to programmatically produce line graphics and colored shapes. Programming code snippets that produce drawing objects can be saved and stored as Objects of the AFOL system. Stored

objects can be reloaded and used within the programming language as existing pieces of programming code. Tables 1 and 2 show the AFOL's class structure and the programming language's commands respectively.

Table 1. AFOL language class attributes and operations

Class Attributes	
Attribute	**Description**
Color	This attribute assigns a color value to the specified Object. The color value can be used by the "drawing" and the "filling" commands.
Penwidth	This attribute assigns a penwidth (width of pen) value to the specified Object. The width of the pen value can be used by the "drawing" commands.
Class Operations	
Operation	**Description**
Reduce	This operation is used to reduce the size of a specified Object by a user specified percentage. Affects all forward and backward turtle drawing movements.
Expand	This operation is used to expand the size of a specified Object by a user specified percentage. Affects all forward and backward turtle drawing movements.
Fliphorizontal	This operation is used to flip horizontally a specified Object. Affects all right and left turning commands.
Flipvertical	This operation is used to flip vertically a specified Object. Affects all right and left turning commands.
Rotate	This operation is used to draw a specified Object rotated clockwise by a user specified angle. Affects the turtle's initial angle.
Save	This operation is used to save a specified Object in the system's database.

Table 1 illustrates a list of attributes and operations that can be used within the Object Oriented structure of the AFOL language. Each existing object can be called by its name in the system's command-line interface. Correspondingly, each specified object has a number of attributes and operations that can take values or can be called on run time by the users. Both the attributes and the operations constitute each objects characteristics and functionalities and are members of a general template class. In accordance with the implementation of the AFOL language, if an object consists of

other objects (for example a drawn house may contain drawn windows and doors), then these sub-objects also inherit the super-object's characteristics. However, all inherited characteristics are the default values for each sub-object and these values can be changed in each sub-object's own implementation.

At this point we may remark that each object's attributes provide an alternative way of assigning values while drawing in the AFOL's interface. For example by writing "TomHouse.color clred" we command the AFOL's turtle to use the red color when drawing the "TomHouse" object. This can also be achieved by assigning the red color as the value of the turtle's current drawing pen before drawing the specified object. However, each object's operations have a more complicated structure and it would be quite difficult to provide an alternative way for their implementation rather than their calls through the existing objects. As an example, if we wanted to reduce the size of a specified object, the reduce call would easily do that ("newobject.reduce 50", which means that the new object will have 50% of the size of the initial object), while a user would have to change a big number of commands in the object's implementation in order to have the same effect. As a result, the OO structure within the AFOL's commands not only provides a better, logical structure for the existing programming language, but also offers more functionality to the users.

Table 2. AFOL language commands

Affective Object Logo Commands	Description of command	Example
Fd *value*	This command is used to move the turtle forward by a specified by the "value" variable range (move forward).	Fd 100
Bd *value*	This command is used to move the turtle backward by a specified by the "value" variable range (move backward).	Bd 80
Rt *value*	This command is used to rotate the turtle clockwise by a specified by the "value" variable range (turn right).	Rt 45
Lt *value*	This command is used to rotate the turtle counterclockwise by a specified by the "value" variable range (turn left).	Lt 180
Fill	This procedure is used to fill a shape with the turtle's current color.	Fill
Color *color-variable*	This command is used to change the turtle's current drawing color to the specified by the "colorvariable" variable color.	Color clBlue
PenWidth *value*	This command is used to change the turtle's current drawing pen width to the specified by the "value" variable value.	PenWidth 5
Pu	This command is used to hold the turtle's pen up, so that the turtle can move without leaving any trace (pen up).	Pu

Table 2. *(continued)*

Pd	This command is used to restore the turtle's pen down (after a pu command), so that the turtle can move leaving its trace (pen down).	Pd
Refresh	This command is used to refresh/clear the drawing area.	Refresh
Home	This command is used to move the turtle to a default position in the center of the drawing area.	Home
New *Object*	This command is used when creating a new Object. The new Object inherits its functionality by a pre-stored to the system's database Object.	New　　　　House TomHouse
Run *Object*	This command is used to run the programming code of a specified pre-loaded Object.	Run TomHouse
Repeat *value*	This command is used to specify the beginning of a loop. Each loop contains commands and is repeated by a number of repetitions specified by the "value" variable.	Repeat 4 Fd 50 Rt 90 Endrpt
Endrpt	This command is used to indicate the end of a "repeat" loop.	

Table 2 shows a listing of the AFOL programming language interface commands, as well as an example of programming code for each command. In the AFOL's GUI the turtle moves with commands that are relative to its own position, for example Rt 90 means rotate right by 90 degrees. Students who use the AFOL system would easily understand (and predict and reason about) the turtle's motion by imagining what they would do if they were the turtle. Some commands produce a drawing effect while the turtle moves on the system's screen, while other commands are used to handle objects and to refresh the drawing area. Furthermore, some commands affect the turtle's available pen that is used for drawing a trace when the turtle moves. Finally, the "repeat" command is used to include a sequence of commands that is executed repeatedly for a specified number of times.

5　Conclusions

In this paper, we presented a new programming language that has been implemented for the needs of teaching programming skills to young children. The resulting sophisticated programming environment is called AFOL. AFOL reflects its author's attempt to preserve well-known past approaches towards the education of children in the area of programming, while at the same time enrich the language's architecture with modern architectures and approaches and also provide a highly attractive interface for young children. The "older" Logo programming language has been used as a foretype for the resulting system which has been extended by the incorporation of the wide spread

model of object oriented programming. The programming language's environment is also very user friendly since it includes affective interaction components, namely bi-modal emotion recognition and emotion elicitation through interactive tutoring agents that participate in the whole educational process.

It is in our future plans to evaluate the AFOL system as an educational tool and as programming environment in order to examine the degree of its usefulness as an educational tool for the teachers, as well as the degree of usefulness and user-friendliness for the young children as students who are going to use the system. Moreover, a future evaluation study is expected to reveal specific affective characteristics of the e-learning environment that may influence the children in becoming more effective students and more satisfied users.

References

1. Picard, R.W.: Affective Computing: Challenges. Int. Journal of Human-Computer Studies 59(1-2), 55–64 (2003)
2. Pantic, M., Rothkrantz, L.J.M.: Toward an affect-sensitive multimodal human-computer interaction. Proceedings of the IEEE 91, 1370–1390 (2003)
3. Alepis, E., Virvou, M., Kabassi, K.: Recognition and generation of emotions in affective e-learning. In: ICSOFT 2009 Proceedings of 4th International Conference on Software and Data Technologies, vol. 2, pp. 273–280 (2009)
4. Alepis, E., Stathopoulou, I.-O., Virvou, M., Tsihrintzis, G.A., Kabassi, K.: Audio-lingual and visual-facial emotion recognition: Towards a bi-modal interaction system. In: Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI, vol. 2, Article number 5670096, pp. 274–281 (2010)
5. Tsihrintzis, G., Virvou, M., Stathopoulou, I.O., Alepis, E.: On improving visual-facial emotion recognition with audio-lingual and keyboard stroke pattern information. In: Web Intelligence and Intelligent Agent Technology, WI-IAT 2008, vol. 1, pp. 810–816 (2008)
6. Alepis, E., Virvou, M.: Emotional intelligence: Constructing user stereotypes for affective bi-modal interaction. In: Gabrys, B., Howlett, R.J., Jain, L.C. (eds.) KES 2006. LNCS (LNAI), vol. 4251, pp. 435–442. Springer, Heidelberg (2006)
7. Feurzeig, W.: Toward a Culture of Creativity: A Personal Perspective on Logo's Early Years and Ongoing Potential. International Journal of Computers for Mathematical Learning, 1–9 (2010) Article in press
8. Frazier, F.: The logo system: Preliminary manual. BBN Technical Report. Cambridge, MA BBN Technologies (1967)
9. Goleman, D.: Emotional Intelligence. Bantam Books, New York (1995)
10. Hudlicka, E.: To feel or not to feel: The role of affect in human-computer interaction. International Journal of Human-Computer Studies, 1–32 (July 2003)
11. Elliott, C., Rickel, J., Lester, J.C.: Lifelike Pedagogical Agents and Affective Computing: An Exploratory Synthesis. In: Veloso, M.M., Wooldridge, M.J. (eds.) Artificial Intelligence Today. LNCS (LNAI), vol. 1600, pp. 195–211. Springer, Heidelberg (1999)
12. Kahn, K.: ToonTalk - An Animated Programming Environment for Children. Journal of Visual Languages and Computing 7(2), 197–217 (1996)
13. Johnson, W.L., Rickel, J.: Lester, J, Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environments. International (2000)

14. Lester, J., Converse, S., Kahler, S., Barlow, S., Stone, B., Bhogal, R.: The Persona Effect: affective impact of animated pedagogical agents. In: Pemberton, S. (ed.) Proceedings of Human Factors in Computing Systems Conference, CHI 1997, pp. 359–366. ACM Press, New York (1997)
15. Pastor, O., Gómez, J., Insfrán, E., Pelechano, V.: The OO-Method approach for information systems modeling: From object-oriented conceptual modeling to automated programming. Information Systems 26(7), 507–534 (2001)
16. Alepis, E., Virvou, M.: Multimodal Object Oriented user interfaces in mobile affective interaction. Multimedia Tools and Applications, 1–23 Article in Press

Multimedia Session Reconfiguration for Mobility-Aware QoS Management: Use Cases and the Functional Model

Ognjen Dobrijevic and Maja Matijasevic

University of Zagreb, Faculty of Electrical Engineering and Computing,
Unska 3, HR-10000 Zagreb, Croatia
ognjen.dobrijevic@fer.hr, maja.matijasevic@fer.hr

Abstract. Evolution of communication networks envisages providing mobile users with seamless Quality of Service (QoS) support. This paper presents a session reconfiguration approach that dynamically manages QoS for multimedia services in response to session mobility and terminal mobility. It considers changes (referred to as *events*) that are induced by mobility and utilizes those affecting QoS to steer the session reconfiguration. As the result, basic operations, called *reconfiguration primitives*, are applied to maintain or adapt QoS. Use cases that illustrate motivation for this work are described, along with the proposed functional model. An initial performance evaluation of the model is also outlined.

1 Introduction

The need to deliver multimedia services "anywhere-anytime" has largely increased over the last ten years [3]. This highlights one of the key requirements envisaged by the evolution of communication networks – providing mobile users with seamless Quality of Service (QoS) support for multimedia services. Such a support must consider various mobility aspects. In this work, we focus on two aspects. First aspect, referred to as *session mobility*, allows a user to continue communication when replacing her terminals, while the second aspect, i.e. *terminal mobility*, enables a user terminal to maintain communication while changing its network attachment point. In order to achieve an adaptive QoS management, the support must regard variable mobility-induced constraints [1,2,5], thus achieving a form of *mobility-awareness*.

Mobility-aware QoS support described in literature mainly focuses on applying transport/network-level operations that consider individual mobility aspects and "conventional" mobility-induced changes. The *Trigger Management Framework* [4] handles notifications caused by conventional changes, e.g. based on received signal strength indications and network load, but also by "higher-level" changes, such as end-to-end QoS degradation. The *Multi-User Session Control* solution [2] provides mobile users with access to sessions, while supporting QoS adaptation against terminal mobility. This approach is centered around a transport/network-level signaling between the specified entities and the adaptation in terms of assigning different QoS classes to session flows or adding/dropping flows from a session.

A service delivery framework for multimedia applications described in [1] is adaptable to terminal and session mobility, which includes customizing session parameters

to bandwidth of access networks and encodings/formats supported by user terminals, but does not involve QoS negotiation and network resource reservation. A solution that provides session continuity between different access networks is presented in [5]. It enables transfer of agreed QoS parameters across the access networks, but assumes that the QoS settings stay the same after the change.

The work presented in this paper complements the current research efforts and presents a session reconfiguration approach that dynamically manages QoS for multimedia services in response to mobility. It employs the application level, which offers independence of an access technology or a service scenario, and ability to make application-dependent decisions. The approach considers changes (referred to as *events*) that stem from session and terminal mobility, and utilizes those affecting QoS to steer the reconfiguration. As the result, basic operations, called *reconfiguration primitives*, are applied to maintain or adapt QoS. The latter involves interacting with control entities that reserve network resources to assign the required allocation.

The remainder of the paper is organized as follows. In Section 2, we define the mobility events and reconfiguration primitives, and describe use cases that illustrate application of the approach. The proposed functional model is presented in Section 3, with the emphasis on necessary functional entities, and on the signaling that facilitates event notification and session reconfiguration. Section 4 depicts an initial performance evaluation of the model. We then conclude the paper.

2 Session Reconfiguration and Use Cases

For the purposes of this work, we define a *multimedia session* as a collection of *media flows* within a group of *participants*. *Media flow* is a stream of media data (e.g., audio or video), to which these parameters are assigned: identifiers of *participants* between which the flow is established, chosen media format and encoding, and a QoS specification defining required network bandwidth, delay, delay variation, and packet loss ratio. *Session participant* relates either to a user terminal, which enables a user to engage in the session, or a server, which acts as a media source.

Mobility Events and Reconfiguration Primitives

Session mobility may lead to heterogeneous terminals being used for accessing the services, while terminal mobility may result in changes to network connectivity and access technology. Using a concept of representing changes with *events*, similar to that proposed in [4], we identify changes induced by the mobility and model them as *mobility events* (illustrated in Figure 1) to be considered by the reconfiguration:

1. *Change of terminal* – represents a change of the user terminal due to *session mobility*;
2. *Change of location* – represents a change in user terminal's location due to *terminal mobility*; and
3. *Change of access network* – represents a change of access network (i.e., underlying access technology) for the user terminal due to *terminal mobility*.

The goal of the session reconfiguration is to modify the elements of a session in order to maintain or adapt QoS. It introduces three *reconfiguration primitives*: (a) *start*
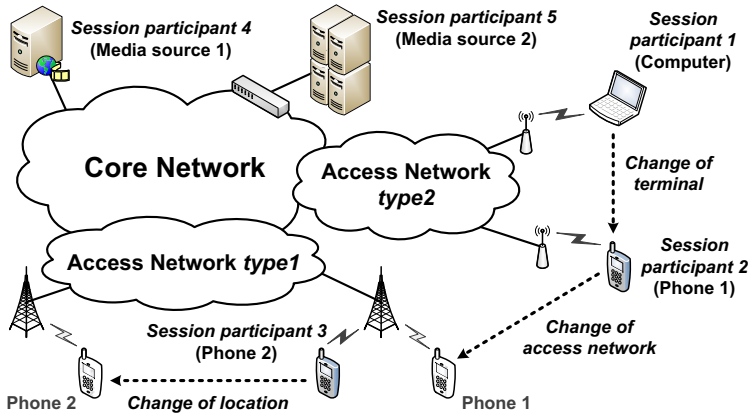
Fig. 1. Illustration of the mobility events

media flow, (b) *stop media flow*, and (c) *modify media flow*. Starting a media flow includes several steps. First, the associated session participants need to agree upon a QoS specification for the considered flow, after which the allocation of required network resources is invoked. Finally, the participants set up media transceivers with the agreed format/encoding and transport parameters, and start the flow. On the other hand, when participants reach a decision to stop a flow, the reserved resources are released and the corresponding media transceivers suspended. Modifying a media flow refers to changing its format/encoding and QoS specification, which leads to adjusting the allocation of the resources for the flow.

Use Cases

Each of the use cases, which illustrate motivation for and application of the approach, is associated to an introduced mobility event.

1) Use case for *Change of terminal*

Maria establishes a video conference with an associate while traveling to work. She uses her smartphone to go over the materials for a meeting. Once at work, she goes to a room with a high-definition plasma screen. Maria decides to transfer the video flow of the conference to the plasma screen and to leave the audio flow on the phone. Prior to the flow transfer, new flow parameters (with a QoS specification) are produced regarding the hardware and software features of the targeted terminal. The reconfiguration applies *start media flow* to initiate video flow to the screen and *stop media flow* for the same flow on the phone.

2) Use case for *Change of location*

After boarding a train, Sofia establishes a QoS-supported session with a "local" game server and starts playing. The session involves several media and data flows. As the train travels to the destination, Sofia's laptop computer changes location and network attachment point, which leads to seamless reconnection to another game server instance, to maintain QoS. The session flows are transferred by invoking *start media flow* and

stop media flow, but the applied QoS specifications are kept. Prior to the transfer, flow parameters are updated to match the new server instance.

3) Use case for *Change of access network*

While going home by a subway, Zoe is playing a multi-player game on her smartphone, which supports several access technologies. The only available access at the time is UMTS (Universal Mobile Telecommunications System). After a while, a HSPA (High Speed Packet Access) network comes to reach. Since Zoe's preferences state that HSPA is preferred, due to a higher bandwidth and improved QoS, her phone changes the access. Then, *modify media flow* is applied to produce new flow parameters (with QoS specifications) and reserve resources in the new network.

3   Functional Model

In order to achieve a mobility-aware QoS management at the application level, the presented approach encompasses several functional aspects:

- session-level signaling for negotiation and dynamic renegotiation of the flow parameters (notably, QoS specifications), in particular due to the mobility events;
- producing QoS specifications for media flows within a multimedia session;
- interacting with control entities that reserve access network resources; and
- generating and conveying the mobility event notifications, as well as deciding upon the reconfiguration primitives to apply in response.

Identified functional model (Figure 2) comprises generic control entities that provide the given aspects. *User Terminal Entity* (UTE) represents a *session participant* used for accessing multimedia services. Each UTE is described with a hardware and software configuration, including supported access technologies, which is stored in a *user profile* at *User Profile Repository Entity*. The *user profile* holds a "knowledge" about the user in terms of her preferences and the available terminals. UTE contains these functions: *Session control function* (SCF), *Event analysis function* (EAF), *QoS monitoring function* (QSMF), and *Media transmission function* (MTF).

EAF processes *User inputs* and produces the *Change of terminal* notifications. SCF is then invoked to signal the notifications to *Session Configuration Management-Support Entity* (SCM-SE), which decides about the reconfiguration. SCF also performs signaling for session establishment and QoS (re)negotiation. When the media flow parameters are agreed, SCF invokes MTF to set up and start, modify, or stop media delivery. QSMF measures actual QoS performance for a media flow.

Mediating Session Control Entity (MSCE) is a signaling proxy for each UTE that wants to establish the sessions. It forwards session control messages between UTEs and the chosen *Serving Session Control Entity* (SSCE), and invokes the allocation of access network resources by interacting with *Resource and Admission Control Entities* (RACEs). MSCE extracts required flow parameters from the control messages and conveys them to RACEs. SSCE is the central signaling proxy, which forwards the control messages between *session participants*. SSCE includes, e.g., SCM-SE in the signaling path, thus invoking functions that SCM-SE provides.
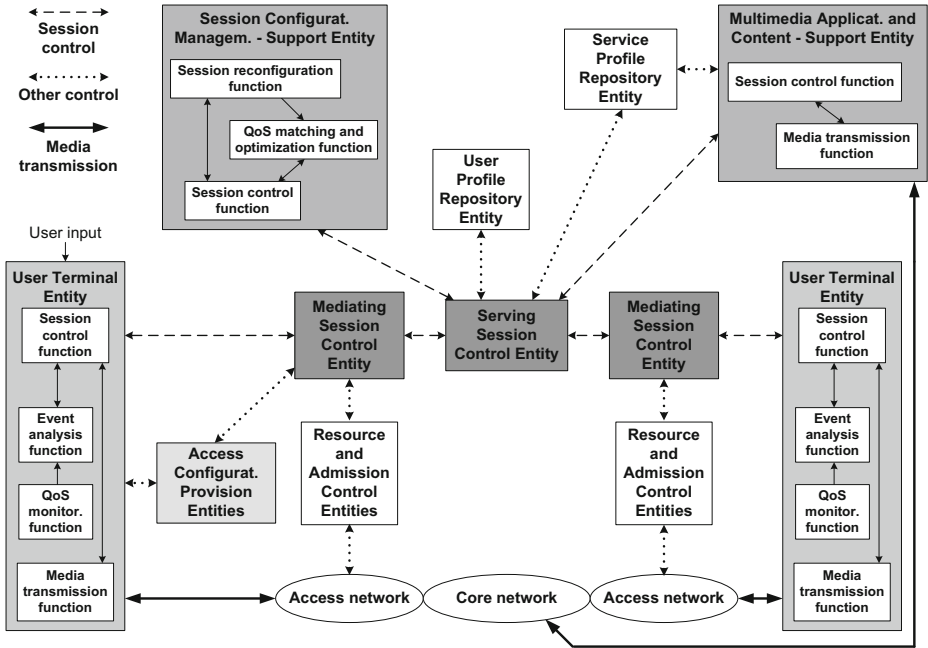
Fig. 2. Functional model for the mobility-aware QoS management

Multimedia Application and Content-Support Entity (MAC-SE) is a *session partic-ipant*, which refers to a server that executes multimedia applications and hosts media content. Each application is described with its requirements and constraints, which are stored in a *service profile* at *Service Profile Repository Entity*. The *service profile* rep-resents a "knowledge" about a multimedia service that the application offers. MAC-SE also incorporates SCF and MTF. Similarly to UTE, SCF handles session control mes-sages, but MAC-SE exchanges them directly with the chosen SSCE. SCF invokes MTF to control media delivery (e.g., streaming).

SCM-SE is the key QoS management component, which includes the *QoS match-ing and optimization function* (QMOF) [6] and *Session reconfiguration function* (SRF). QMOF produces QoS specifications based on information obtained from considered *user* and *service profiles*. An event notification signaled to SCM-SE is delivered to SRF, which analyzes the change and decides about the primitive(s) to be applied. SCM-SE includes SCF to exchange session control messages with the given SSCE, allowing it to engage in QoS negotiation and receive the notifications.

RACEs authorize and allocate resources in access networks, and include *Policy De-cision Entity* (PDE) and *Policy Enforcement Entity* (PEE). While PDE makes decisions regarding authorization, reservation and release of the resources, PEE imposes the deci-sions. *Access Configuration Provision Entities* provide information about UTE's access network and location, including a unique identifier of the network, thus notifying for *Change of location* and *Change of access network*.

Signaling Procedures

This QoS management approach defines five signaling procedures. *Media flow establishment* specifies signaling that establishes one or more media flows, while *Media flow termination* relates to stopping the flows and releasing the allocated resources. The three remaining procedures define signaling in response to the corresponding events, and the *Change of user terminal* procedure is described in details.

1) The *Media flow establishment* signaling

This procedure creates an association of two participants for exchanging media flows, during which capabilities and demands of the participants are matched, and the flow parameters are agreed upon. Figure 3 depicts the Unified Modeling Language (UML) sequence diagram that specifies the procedure. It illustrates the flow establishment between a UTE and a MAC-SE. For a clearer presentation, the *User* and *Service Profile Repository Entities* are merged into the *Profile Repository Entity*.

Signaling is invoked when a user decides to access a multimedia service, which leads to establishing a session between user's UTE and the hosting MAC-SE. After the UTE sends a *session request* message to the MAC-SE (Figure 3, step 1), it traverses the UTE's MSCE and the given SSCE, which acquires the needed *user profile*, authorizes the user, and retrieves *service profile* for the requested service. Then, the MAC-SE accepts the request (*session response* is sent to the UTE), which triggers the SSCE to invoke the SCM-SE (step 10). The QMOF produces a *feasible service profile* by matching parameters from the given profiles. The feasible profile contains flow parameters that both the UTE and the MAC-SE support. It offers a set of media formats/encodings for each flow, with the associated QoS specifications.

The *feasible service profile*, which is referred to as an offer, is then sent to the MSCE to invoke authorization of the resources. The procedure continues at the UTE (step 17), where the user can accept, modify or deny (a subset of) parameters from the feasible profile (e.g., she can choose which media flows are to be established). The chosen parameters form a *feasible service profile answer* (step 18), which is conveyed to the MAC-SE and, then, to the SCM-SE. The QMOF performs the *optimization* (step 23), to determine an optimal usage of the resources that are needed for providing the chosen QoS specifications. The resulting *optimized service profile* is used by the PDE/PEE to allocate the resources (steps 26 and 27). As there are multiple combinations of the formats/encodings and QoS specifications which can be applied, the PDE decides of the reservation considering current resource availability. The optimized profile is then sent to the UTE and the MAC-SE, to set up their media transceivers with the selected parameters and to initiate media flow(s).

2) The *Change of user terminal* signaling

Change of user terminal enables transfer of one or more media flows between session participants, without breaking the established communication. The procedure is depicted by the UML sequence diagram shown in Figure 4.

It illustrates signaling between three UTEs, assuming that media flows are established between UTE1 and UTE2, and that a user wants to transfer the flows from UTE1 to UTE3, which are both assisted by the same MSCE. The reconfiguration first establishes the flows between UTE3 and UTE2, and then terminates the corresponding flows
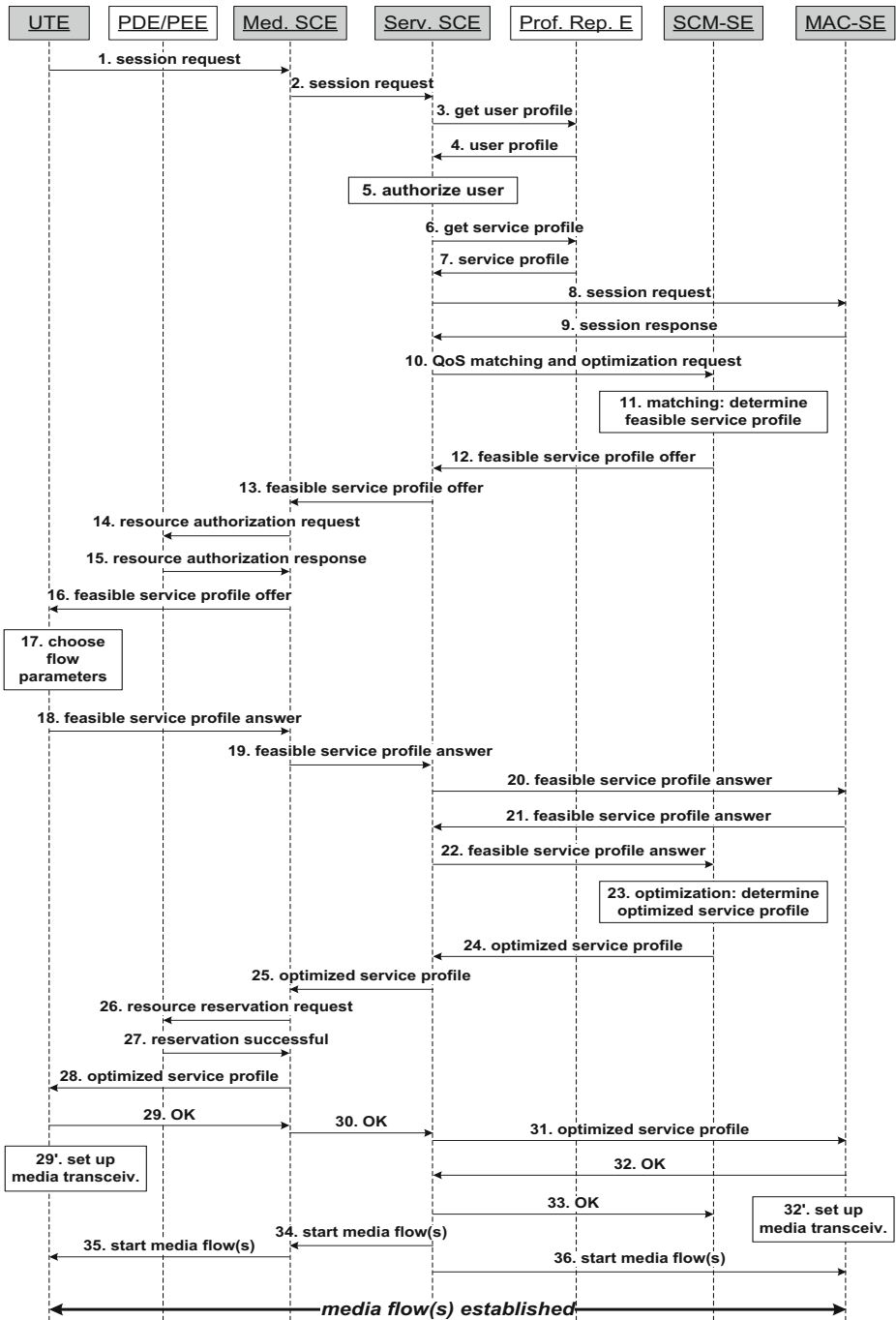
Fig. 3. The *Media flow establishment* signaling procedure
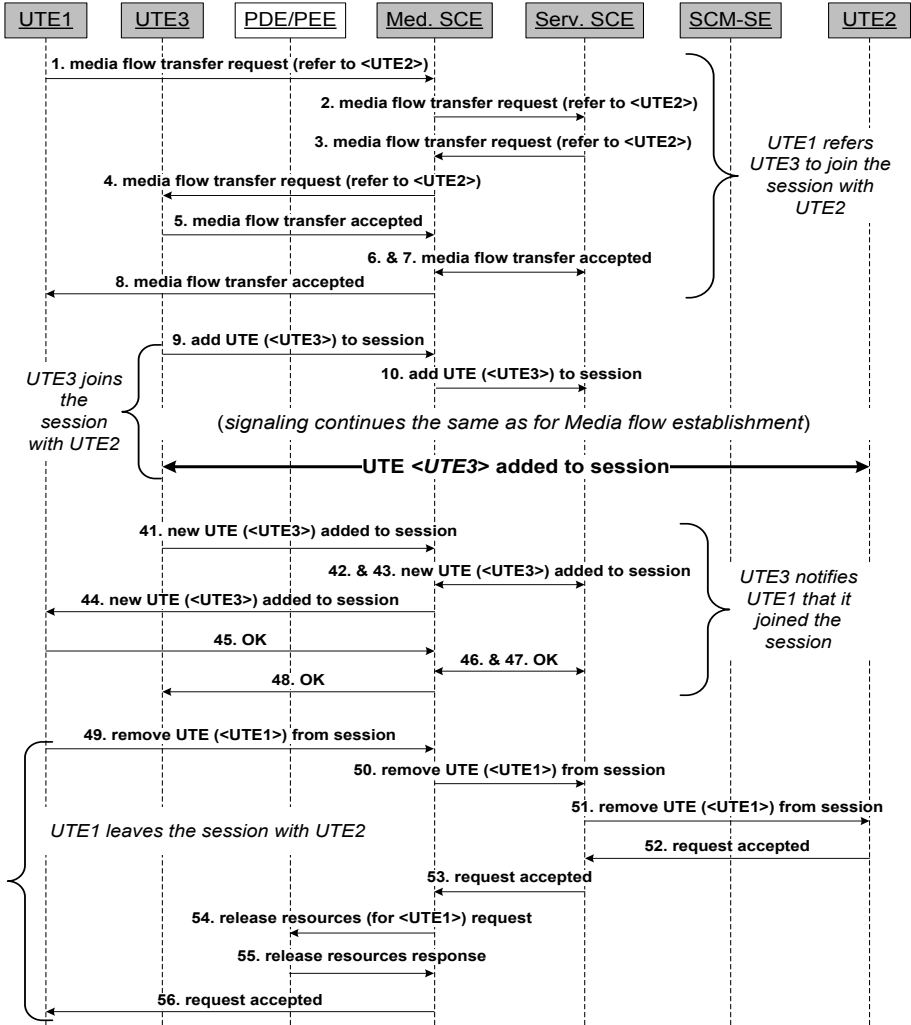
Fig. 4. The *Change of user terminal* signaling procedure

between UTE1 and UTE2. For a clearer presentation, a MSCE and a PDE/PEE support-
ing UTE2 are omitted from the sequence diagram, as is the *Profile Repository Entity*.

Signaling is invoked when a user decides to replace her terminal (referred to as
UTE1) in an ongoing session and identifies another terminal (UTE3) as the transfer
target. User's request results in the associated *Change of terminal* notification being
produced, which UTE1 sends as a *media flow transfer request* message to UTE3 (Fig-
ure 4, steps 1-4). This message contains identifier of the correspondent UTE (referred
to as UTE2). After the request is accepted by UTE3 (steps 5-8), it sends an *add UTE to
session* message to UTE2.

When UTE2 accepts this "join request" from UTE3, the procedure continues similarly as for *Media flow establishment*: the SRF and the QMOF produce the *feasible* and *optimized service profiles* for UTE3, which are used for reserving the network resources, and the media transmission is initiated towards UTE3 (thus adding the latter to the session). Afterwards, UTE3 initiates removal of UTE1 from the session (steps 41-44), which prompts UTE1 to end its participation (steps 49-56). Before the removal of UTE1, the resources allocated to it are released (steps 54-55).

4   Performance Evaluation

The proposed functional model is formally defined by using *Discrete Event System Specification* (DEVS) [7]. DEVS is a formal framework for simulation of general discrete event systems, which offers *atomic DEVS models* to capture systems' behavior and *coupled DEVS models* to build their structure. Each proposed model entity is realized as an atomic DEVS (aDEVS), while the associated signaling messages are "mapped" to input and output events of the aDEVS. Defined aDEVS models are integrated by using coupled DEVS. The definition is the basis for an implementation and simulation of the proposed model in the DEVS-Suite Simulator [8].

An initial performance evaluation of the model is conducted to assess its scalability. It introduces the *duration* metric, which is measured in relation to the number of UTEs that simultaneously execute a particular signaling procedure. Each procedure defines *duration* as the interval to exchange all the messages (e.g., for *Media flow establishment*, this is the interval between *1. session request* and *36. start media flows*), with its *reference value* implying single procedure execution. As the proposed model is generic, these values may be set as seen fit. For this evaluation, the values for *Media flow establishment* and *Change of user terminal* are set to 14.0 and 25.5 "time units" (with regards to message number ratio between these procedures). The results (Table 1) indicate that average *duration* for the analyzed procedures increases almost linearly with the number of UTEs, thus offering a good scalability.

Table 1. Average *duration* for the analyzed signaling procedures

Number of UTEs	1100	1400	1700	2000
Media flow establishment [time units]	17.5	19.4	22.3	25.0
Number of UTEs	50	100	150	200
Change of user terminal [time units]	30.0	30.5	31	31.7

5   Conclusions and Future Work

This paper presents use cases and a proposed functional model of mobility-aware QoS management for multimedia services. The approach employs session reconfiguration at

the application level to maintain or adapt QoS dynamically, and independently of an access type or a service scenario. Generic mobility events and the applicable reconfiguration operations are defined. We illustrate model application in the signaling procedure for managing QoS against session mobility, while initial evaluation indicates a good scalability to number of the participants. Future work includes an extensive model evaluation and application in a real-network prototype.

References

1. Balakrishna, C., Al-Begain, K.: Towards a user-centric and quality-aware multimedia service delivery implementation on IP multimedia subsystem. In: Al-Begain, K. (ed.) Proc. of the 2007 Int. Conference on Next Generation Mobile Applications, Services and Technologies (NGMAST 2007), pp. 36–42. IEEE Computer Society, Los Alamitos (2007)
2. Cerqueira, E., Veloso, L., Neto, A., Curado, M., Monteiro, E., Mendes, P.: Mobility managementfor multi-user sessions in next generation wireless systems. Comput. Commun. 31, 915–934 (2008)
3. Gomes, D., Aguair, R.L., Sargento, S.: A cross-system approach for multimedia services with IP multicast in 4G networks. Wireless Pers. Commun. 52, 651–668 (2010)
4. Mäkelä, J.: Towards seamless mobility support with cross-layer triggering. In: Papaoulakis, N., Tragos, E. (eds.) Proc. of the 18th IEEE Int. Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2007), pp. 1–5. IEEE, Los Alamitos (2007)
5. Renier, T., Larsen, K.L., Castro, G., Schwefel, H.-P.: Mid-session macro-mobility in IMS-based networks. IEEE Veh. Technol. Mag. 2, 20–27 (2007)
6. Skorin-Kapov, L., Matijasevic, M.: Modeling of a qoS matching and optimization function for multimedia services in the NGN. In: Pfeifer, T., Bellavista, P. (eds.) MMNS 2009. LNCS, vol. 5842, pp. 55–68. Springer, Heidelberg (2009)
7. Zeigler, B., Prähofer, H., Kim, T.G.: Theory of Modeling and Simulation. Academic Press, New York (2000)
8. DEVS-Suite Simulator, http://sourceforge.net/projects/devs-suitesim/ (cited April 15, 2011)

LSB Steganographic Detection Using Compressive Sensing

Constantinos Patsakis[1], Nikolaos Aroukatos[1], and Stelios Zimeras[2]

[1] Department of Informatics, University of Piraeus
[2] Department of Statistics and Actuarial - Financial Mathematics,
University of Aegean

Abstract. People have always been using several techniques in order to protect their privacy. For centuries, steganography has been used, but only in the last decades the proper mathematical background has been developed. Due technological advances, steganography has found many applications, with most important the protection of digital assets through DRM.

This work proposes a new detection method of steganographic content through compressive sensing. The proposed method is a probabilistic filter that can detect steganographic content in images with increased probability, if this is made with the LSB method, after applying a filter with compressive sensing technique.

Keywords: Steganography, compressive sensing, steganalysis.

1 Introduction

Our constantly increasing needs for communication through Internet and mobile telecommunications have also increased our need for privacy. The two main methods for protecting our data is cryptography and steganography. The former tries to hide the contents of our data, while the latter tries to hide their very own existence. In practice, we never use steganography on its own, we use cryptography as well as another layer of security, in order to protect our data in case the existence of the content becomes known.

The two most widely used techniques in images are LSB and frequency domain. The first method is based on embedding the data in the least significant bits of the medium, something that introduces some distortions in the picture. These however which may be untraceable, if the pixels used are properly selected. The other methods take the waveform of the image and embed the data in its DCT or DWT form, by quantizing the coefficients.

In our work we make use of compressive sensing algorithms, specially BM3D algorithm. The core idea of compressive sensing is that if a signal is sparse then it can be described by a smaller basis, while containing most of the useful information. The idea was introduced in 2004 by Donoho [5]and was extended later by Candès, Romberg and Tao[14].

The current work presents a new algorithm that tries to detect LSB steganographic embedding of data in images by the use of compressive sensing

algorithms. The algorithm that has been used is bm3d, but as it will become apparent, others can be used, possibly with better results.

2   Steganalysis

Steganalysis is the opposite of steganography, it is the methods that are used in order to detect embedded content in a medium. Depending on the knowledge that we have about the steganographic system and regardless of the medium that is used as a stego-object, the steganalysis techniques can be categorized to the following main three groups:

Blind identification. In this group, we do not have any knowledge of the steganographic system and we try to detect steganographic data based solely on statistical properties of the medium.

Parametric statistical steganalysis. In some cases we may know that the used steganographic system changes some properties of the medium with specific patterns on some of its properties. The techniques of this group try to detect the existence of steganographic data by detecting these patterns.

Supervised learning based steganalysis. These techniques use statistical classifiers to see whether the tested image is a stego image. In this group we have a set of known clean set of object and one of stego-object. These sets are used for training our algorithm to detect whether a medium is stego-object or not.

Since the focus of this work is on image steganalysis and specially on LSB hiding, we will point out some already used methods. One of the first methods used was χ^2 method [6], where we examine whether the tested image's histogram is close to the histogram of the same image with embedded data.

Fridrich, Goljan and Du [7] proposed in 2001 the RS (Regular/Singular) scheme. Their approach counts the number of occurrences of pairs in sets. The idea is that spatially adjacent image pixels, which are highly correlated, can give us a lot of information whether LSB has been applied in the examined image, meaning that if LSB has been applied then areas where embedding has been made then adjacent image pixels would appear to have many different properties compared to where no tampering has been made.

Related sample pair analysis [8] is another steganalysis method for detecting LSB data hiding in images. In this approach we take sample pairs of values and we divide them into subsets depending on the relation of the two values to one another. This technique makes the assumption that if it is not a stego image, then the number of pairs in each subset are almost equal and that LSB embedding is quantizing the subsets by changing their numbers, something that can be easily detected.

For more an introduction to general steganalysis techniques one may refer to [1, 2] and for more advanced techniques and extensions of the above to [9, 10, 11, 12].

3   Compressive Sensing and BM3D

Compressive sensing, also known as compressed sensing and compressive sampling is a method to reconstruct a signal. This method is used by several scientific fields in Information technology. Donoho first and then Candès, Romberg and Tao devised a method to reconstruct an image from an amount of data much less from the number of data that would be deemed sufficient by the Nyquist–Shannon sampling theorem [13]. Their approach was that in many cases the signals are sparse, hence using another smaller basis we might be able to recover the most usefull part of the signal by using a smaller sample. Several applications of these algoriths have so far been made, raging from MRI reconstruction to coding theory.
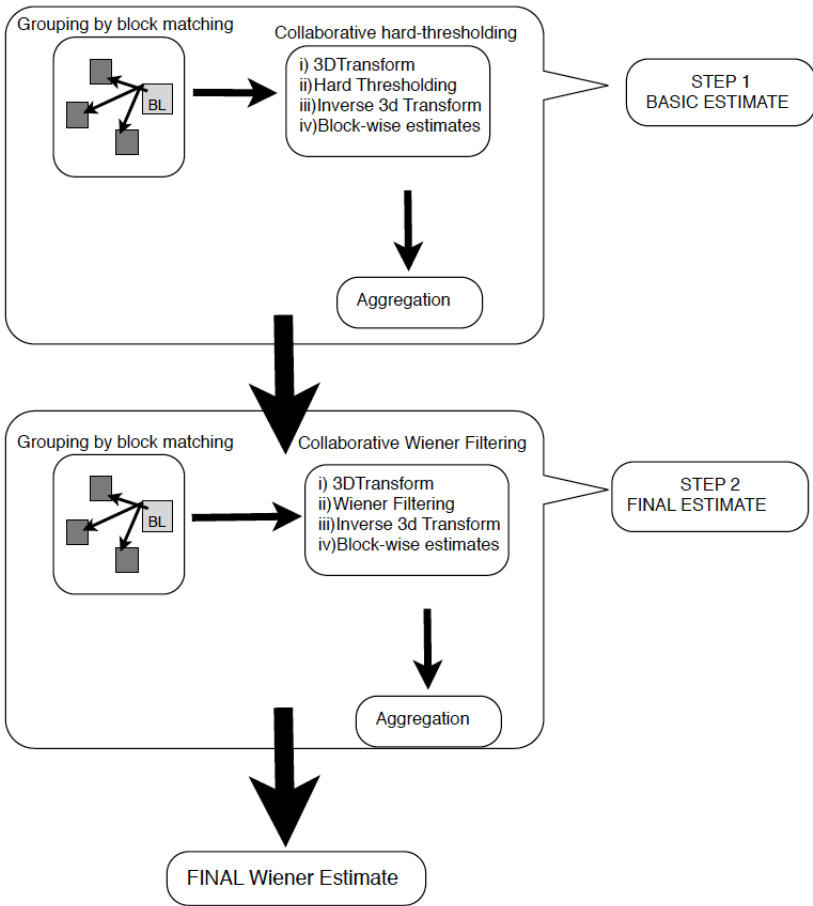
Fig. 1. The BM3D method

An algorithm that belongs to this family of compressive sampling is the BM3D technique, that we use in our work. In this method, a group of image blocks is created around to a selected block. Filtering takes place in blocks grouped in 3d arrays. This is considered to give spatial adaptively to the algorithm. Filtered blocks are then ungrouped and returned to their original coordinates. This method of collaborative filtering preserves local details in the image and achieves high quality denoising. BM3D method's effectiveness does not depend on image weights calculation, which is computationally expensive, rather it is based on the similarities between groups of image parts, which is consider to be more efficient. The BM3D algorithm by Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian is implemented in two steps, the basic estimate and the final estimate. A draft outline of the algorithm is presented in Figure 1.

4   The Proposed Method

The main idea of our proposal is to regard each image to be examined as an image with embedded noise. The noise our image has, the more the image is probable to have LSB embedded information. The question that is raised now is what we mean with a lot of noise, and how do we measure it?

As discussed above, compressive sampling is a method for reconstructing our signals, taking a small sample of it. The assumption that we make is that if the image has no LSB embedding, then the reconstruction with a compressive sampling will be more close to the original, than to one stego-image. The reason for making this assumption is that when parsing an image from a compressive sampling algorithm, if it has LSB data embedded, then the algorithm will not perform as well, because it has extra noise to remove.

The method now is very straight forward, firstly we select a compressive sampling algorithm, in our case BM3D. Afterwards we take the image I to be

Fig. 2. The original Lena Picture, embedded with data and BM3D processed

examined and parse it to BM3D to recover image I'. We then compare the differences of images I and I' in pixel colors.

5   Results

In order to test our method, we made several tests using Matlab and a set of classic test images for image processing. In our tests we took each picture I, created a stego-image out of it, with LSB and we processed both images using BM3D. We then took each processed picture and compared it to the original one, keeping their percent color difference. The results can be seen in Tables 1 and 2.

For parameter sigma equal to 1 it becomes apparent that we have a big difference between "clean" images and stego images. The results are more clear for sigma equal to 2, where we can have an obvious criterion for LSB stego images, which is 40%.
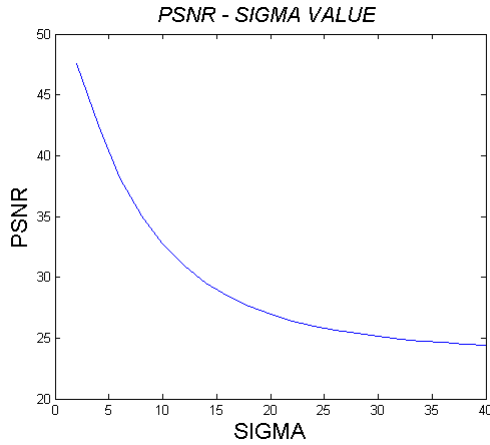
Fig. 3. PSNR-Sigma Value Diagram

6   Conclusions

The proposed method performs with extremely good percentages in finding stego-images, with minimal false-positives and true-negatives, setting specific bounds on when an image should be regarded as a stego-image. The applied filter is very quick and does not need any special hardware installed. One drawback of the method as presented here is that BM3D is an algorithm for grayscale images, yet this can be easily circumvented by parsing each color layer on its own and detecting LSB existence on each one.

Table 1. Results table with parameter sigma equal to 1

camera256		lena512		barbara512		boat512	
Stego-image	self	Stego-image	self	Stego-image	self	Stego-image	self
12.6785	3.5461	7.2735	2.1938	6.3694	1.9016	2.1450	0.6226
12.6724	3.6194	7.3082	2.1770	6.3522	1.8887	2.1580	0.6153
12.6663	3.5675	7.3376	2.1698	6.3393	1.8967	2.1885	0.6203
12.7441	3.5767	7.3650	2.2011	6.3660	1.9070	2.1935	0.6157
12.6389	3.6911	7.3418	2.1770	6.4304	1.8661	2.1790	0.6134
12.7151	3.5629	7.3009	2.2072	6.3995	1.9062	2.1687	0.6275
12.6831	3.6301	7.3254	2.1992	6.3465	1.9138	2.1610	0.6001
12.7579	3.6026	7.3372	2.1996	6.3892	1.8829	2.1824	0.6363
12.6343	3.5172	7.3704	2.1847	6.3488	1.8887	2.1919	0.6310
12.7792	3.5782	7.3563	2.1645	6.3854	1.8501	2.1652	0.6275
house512		hill512		house256		f16 512	
Stego-image	self	Stego-image	self	Stego-image	self	Stego-image	self
13.7711	3.9795	2.7973	0.9209	13.7177	4.1214	13.6410	3.7518
13.7192	3.9795	2.8141	0.9212	13.6993	4.1489	13.6063	3.8128
13.6841	4.0085	2.8297	0.9312	13.7329	4.0009	13.5990	3.7708
13.6734	3.9047	2.8099	0.9480	13.6810	4.2007	13.5761	3.8303
13.7909	3.9185	2.8172	0.9281	13.8107	4.0344	13.6692	3.7991
13.6856	3.8986	2.7718	0.9422	13.7009	4.1748	13.6467	3.7914
13.7939	3.9200	2.8049	0.9308	13.7604	4.1351	13.5857	3.7815
13.7375	3.9841	2.7851	0.9430	13.7482	4.1275	13.6761	3.7796
13.6780	3.9124	2.7779	0.9415	13.7207	4.1214	13.6066	3.7937
13.7070	3.9154	2.8038	0.9377	13.6658	4.0985	13.6459	3.8197

Table 2. Results table with parameter sigma equal to 2

camera256		lena512		barbara512		boat512	
Stego-image	self	Stego-image	self	Stego-image	self	Stego-image	self
40.3076	15.7578	51.6136	22.9492	40.4022	16.3910	33.0948	11.2705
40.4037	15.8493	51.7689	22.8558	40.4358	16.4154	33.0269	11.3544
40.5121	15.8386	51.6273	22.9527	40.4747	16.3254	33.0128	11.3163
40.3961	15.7898	51.7071	22.9961	40.4057	16.4505	32.9964	11.3148
40.3244	15.8722	51.6781	23.0061	40.4091	16.4417	33.0132	11.3434
40.4922	15.9988	51.6968	22.9591	40.4026	16.4261	32.9857	11.3171
40.5380	16.0004	51.6750	22.9942	40.4678	16.4295	32.9704	11.3155
40.2603	15.8127	51.6808	22.9862	40.4591	16.3738	32.9750	11.2873
40.3946	15.7013	51.6468	22.9397	40.4305	16.4524	33.0158	11.2595
40.3076	15.7196	51.6842	22.9893	40.4610	16.3269	33.0318	11.2869
house512		hill512		house256		f16 512	
Stego-image	self	Stego-image	self	Stego-image	self	Stego-image	self
44.7510	15.5579	34.7626	13.2710	44.2902	15.4190	43.7489	16.9132
44.8563	15.6723	34.7836	13.2141	44.6243	15.5716	43.7405	16.9407
44.7525	15.6052	34.8049	13.1409	44.8288	15.5914	43.8583	17.0349
44.9631	15.5136	34.7996	13.1508	44.1925	15.5396	43.9098	16.9735
44.8730	15.7425	34.8301	13.1710	44.4580	15.5807	43.8236	16.9411
44.7845	15.4694	34.7305	13.1321	44.5419	15.5853	43.7756	16.9014
44.9738	15.6906	34.7878	13.1653	44.5465	15.6082	43.7210	16.9521
44.6198	15.8340	34.8412	13.0959	44.7235	15.6204	43.7336	17.0258
45.0974	15.6906	34.7809	13.1603	44.5862	15.3336	43.7870	16.8785
44.9158	15.7990	34.7622	13.1874	44.4595	15.6586	43.7050	16.9323

It is obvious that the method can be applied with other compressive sensing algorithms, besides BM3D, possibly with better results. As an extension to the method, we will try to test it against DCT and DWT steganographic schemes and on other mediums, like audio.

References

[1] Wayner, P.: Disappearing Cryptography: Information Hiding: Steganography & Watermarking, 3rd edn. Morgan Kaufmann, San Francisco (2008)

[2] Cox, I., Miller, M., Bloom, J., Fridrich, J., Kalker, T.: Digital Watermarking and Steganography, 2nd edn. Morgan Kaufmann, San Francisco (2007)

[3] Donoho, D.L.: Compressed Sensing. IEEE Transactions on Information Theory 52(4), 1289–1306 (2006)

[4] Candés, E.J., Wakin, M.B.: An Introduction To Compressive Sampling. IEEE Signal Processing Magazine 21 (March 2008)

[5] Donoho, D.L.: Compressed Sensing. IEEE Transactions on Information Theory 52(4), 1289–1306 (2006)

[6] Westfeld, A., Pfitzmann, A.: Attacks on steganographic systems. In: 3rd International Workshop on Information Hiding (1999)

[7] Fridrich, J., Goljan, M., Du, R.: Reliable detection of LSB steganography in color and grayscale images. In: Proc. ACM Workshop on Multimedia Security, Ottawa, Canada (2001)

[8] Dumitrescu, S., Wu, X., Wang, Z.: "Detection of LSB steganography via sample pair analysis. IEEE Trans. on Signal Processing 51(7), 1995–2007 (2003)

[9] Dumitrescu, S., Wu, X.: A new framework of LSB steganalysis of digital media. IEEE Trans. on Signal Processing, Supplement on Secure Media 53(10), 3936–3947 (2005)

[10] Dumitrescu, X.: LSB Steganalysis Based on High-order Statistics. In: Proc. of ACM Multimedia Security Workshop 2005, pp. 25–32 (August 2005)

[11] Fridrich, J., Pevny, T.: Novelty Detection in Blind Steganalysis. In: Proc. ACM Multimedia and Security Workshop, Oxford, UK, September 22-23, pp. 167–176 (2008)

[12] Fridrich, J., Pevny, T., Ker, A.D.: From Blind to Quantitative Steganalysis. In: Proc. SPIE, Electronic Imaging, Media Forensics and Security XI, San Jose, CA, January 18-22, pp. 0C 1-0C 14 (2009)

[13] Shannon, C.E.: Communication in the presence of noise. Proc. Institute of Radio Engineers 37(1), 10–21 (1949)

[14] Candés, E.J., Romberg, J., Tao, T.: "Stable signal recovery from incomplete and inaccurate measurements". Comm. Pure Appl. Math. 59, 1207–1223 (2006)

Analysis of Histogram Descriptor for Image Retrieval in DCT Domain

Cong Bai, Kidiyo Kpalma, and Joseph Ronsin

Université Européenne de Bretagne, France
INSA de Rennes, IETR, UMR 6164, F-35708, RENNES
{cong.bai,kidiyo.kpalma,joseph.ronsin}@insa-rennes.fr

Abstract. Many researches of content-based image retrieval appear in transform domain. We analyze and enhance a histogram method for image retrieval in DCT domain. This approach is based on 4×4 block DCT. After preprocessing, AC and DC Patterns are extracted from DCT coefficients. After various experiments, we propose to use zig-zag scan with fewer DCT coefficients to construct the AC-Pattern. Moreover adjacent patterns are defined by observing distances between them and merged in AC-Pattern histogram. Then the descriptors are constructed from AC-Pattern and DC-Pattern histograms and the combination of these descriptors is used to do image retrieval. Performance analysis is done on two common face image databases. Experiments show that we can get better performance by using our proposals.

Keywords: Content-based image retrieval, DCT, zig-zag scan, adjacent patterns, face recognition.

1 Introduction

As the majority of the images are stored in compressed format and most of compression technologies adopt transforms to achieve a large amount of compression, image retrieval in transform domain has been widely studied in many researches. Comparing to traditional approaches of indexing compressed images in which they need to decode the images to the pixel domain first; working directly in transform domain has the advantages in time consuming and computational load. Discrete Cosine Transform (DCT) is used in JPEG compression standard. In this aspect, DCT is also used as a tool to extract features in image retrieval. In last few years, many researches appeared in this field.

The DCT coefficients represent some directional information at different resolutions, so they can be used directly as texture features [1]. In a block the upper left DCT coefficients are categorized into 4 groups: one is DC coefficient and three other ones include the coefficients which have vertical, horizontal and diagonal information. In [2], author groups coefficients into several vectors to represent colour and texture features. In [3], the statistical information of DCT coefficients is used as texture features. In [4], the histogram of DCT patterns is constructed and it is used to do retrieval.

Histogram is a typical statistical tool that extracts global information of the whole image. In this paper, based on the histogram method presented in [4], we propose to use a zig-zag scan with fewer coefficients to construct AC-Pattern. Furthermore, we define adjacent patterns by observing distances and merge them in AC-Pattern histogram. Experiments in GTF [10] and ORL [11] database demonstrate the outperformance of our proposal.

The paper is organized as follows. Descriptions of constructing descriptor and parameter analysis are given in section 2 and section 3. Section 4 presents the results of experiments and a conclusion is given in section 5.

2   Description of the Method

The flowchart of the generation of the descriptor is as follows:

Image

DCT Block transforms

Preprocessing

AC-Pattern and DC-Pattern
Histogram generation
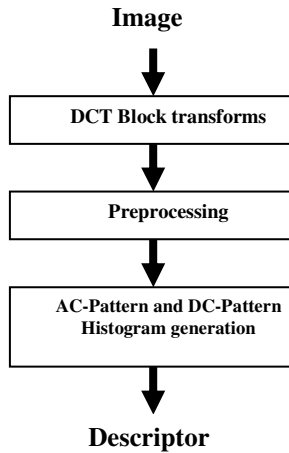
Descriptor

Fig. 1. Descriptor generation

Images are divided into $P \times Q$ blocks and each block is transformed independently by 2-D DCT. In this study, we use $P = Q = 4$, therefore, 4×4 block DCT transform is obtained.

AC-Pattern is referred as the AC coefficients in one DCT block. The total number of AC coefficients in a DCT block is 15, but the number of coefficients that are used to construct the AC-Pattern can be adjusted. Time and performance advantages can be got by this adjustment. There are two methods of scanning to range the coefficients in AC-Pattern. The first way is a row-by-row manner, as used in [4]. We call it linear scan. The second way is a zig-zag scan.

The histogram of AC-Pattern (H_{AC}) is defined as numbers of appearance of AC-Patterns in the image and use DC-DirecVec histogram [4] as DC-Pattern histogram (H_{DC}). The descriptors are constructed from these two histograms.

2.1 Pre-processing

To reduce the impacts of luminance variations, pre-processing steps need to be done before constructing the histograms [4].

Assume there are N DCT blocks in an image i, and the DC value for each block is denoted by $DC_j(i), 1 \le j \le N$. From these DC values, we can calculate the mean DC values for this image:

$$DC_{mean}(i) = \frac{1}{N} \sum_{j=1}^{N} DC_j(i) \tag{1}$$

Then the average luminance $DC_{allmean}$ of all images in a database is calculated:

$$DC_{allmean} = \frac{1}{M} \sum_{i=1}^{M} DC_{mean}(i) \tag{2}$$

where M denotes the total number of images in the database. Then, the ratio of luminance rescaling for image i is calculated:

$$R_i = \frac{DC_{allmean}}{DC_{mean}(i)} \tag{3}$$

Then all the DCT coefficients are normalized with respect to R_i by rescaling them:

$$\overline{F_i(u,v)} = F_i(u,v) \times R_i \tag{4}$$

After normalization, the DCT coefficients are quantized by using a quantization parameter QP:

$$\overline{\overline{F_i(u,v)}} = \frac{\overline{F_i(u,v)}}{QP} \tag{5}$$

2.2 Construction of the AC-Pattern Histogram

Observed AC-patterns from blocks are numerous. So an objective is the reduction of their number. To do this, adjacent patterns will be defined and merged by observing distances between coefficients in AC-Patterns i and j. The concept of adjacency is defined as follows:

$$\text{If } \left|C_i(1) - C_j(1)\right| \le Th \text{ or} \left|C_i(2) - C_j(2)\right| \le Th \text{ or} \cdots \text{or} \left|C_i(m) - C_j(m)\right| \le Th \tag{6}$$

then these two patterns will be classified as adjacent patterns. Where $C_i(k)$ and $C_j(k)$ ($1 \le k \le m$, m indicates the number of coefficients in AC-Pattern) represent AC coefficients. Th is the threshold. In our proposal, $Th = 1$.

To merge the adjacent patterns, we arrange the bins in adjacent order firstly and then merge the adjacent patterns by histogram equalization. Furthermore, there is one special AC-Pattern in which all the AC-coefficients are zero. We exclude this pattern from the AC-Pattern histogram.

2.3   Construction of DC-Pattern Histogram

We use DC-DirecVec histogram [4] as DC-Pattern histogram. As shown in Fig.2, nine differences between DC coefficients of the current block and its 8 neighbours are calculated. The difference of direction 9 is the difference between current DC value and the mean of all the nine neighbouring DC values. The absolute value of the differences are ordered in descend order and the first γ direction-values with largest differences are taken to form direction vectors. γ is a parameter which can be adjusted to get the best retrieval result. These direction vectors are used as DC-Patterns and histograms H_{DC} are constructed based on these patterns.
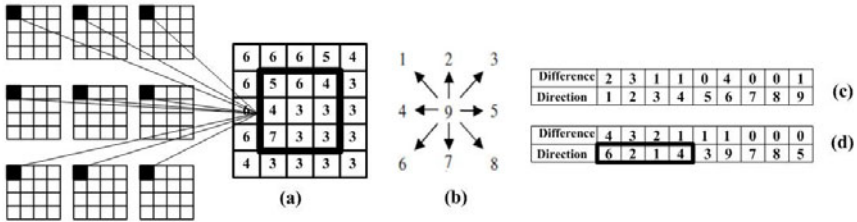
Fig. 2. Forming Direction vector (a) DC coefficients are extracted from neighbouring blocks (b) Directions (c) Differences for each direction (d) Directions with largest differences from direction vector

As there is only part of patterns which appear in large quantities and a large number of patterns which appear rarely, so in consideration of time-consuming and efficiency, we just select those DC-Patterns which have higher frequencies as the bins of histogram.

2.4   Application to Image Retrieval

To evaluate this proposal in image retrieval, we do three classes of experiments by using AC-Pattern histogram, DC-Pattern histogram and their combination. For AC-Pattern histogram alone or DC-Pattern histogram alone, the descriptor is defined as follows:

$$D = H_{AC} \qquad \text{or} \qquad D = H_{DC} \qquad (7)$$

For the combination,

$$D = [(1-\alpha) \times H_{AC}, \alpha \times H_{DC}] \qquad (8)$$

where α is a weight parameter that controls the impact of AC-Patterns histogram and that of DC-Patterns histogram.

To measure the similarity between two descriptors we use the Manhattan distance:

$$Dis_{i,j} = \sum_{k=1}^{m} \left| D_i(k) - D_j(k) \right| \tag{9}$$

where $D(k)$ demonstrates the vectors of the descriptor and i, j demonstrate the images to compare, m indicates the total number of bins in the descriptor.

3   Paramaters of Descriptor

This paragraph is dedicated to the parameters that can influence the performance of retrieval: the scanning method, the number of coefficients used in AC-Pattern, the quantization parameter, the number of bins used in the descriptor, the parameter of DC-Pattern γ and the weight parameter α.

The scanning methods for arranging AC coefficients in the AC-Pattern: linear scan and zig-zag scan. In the linear scan, used in [4], AC coefficients are ordered from left-to-right and top-to-bottom. For most images, much of the signal energy lies at low frequencies coefficients that appear in the upper left corner of the DCT. So we propose to use zig-zag scan to order the AC coefficients in order to take into account the highest coefficients first. By doing this, the coefficients are in the order of increasing frequency. So comparing with linear scan, zig-zag scan gains more advantages in using coefficients that have higher energies. So we propose to use zig-zag scan instead of linear scan for constructing AC-Patterns. These two methods are shown in Fig.3.
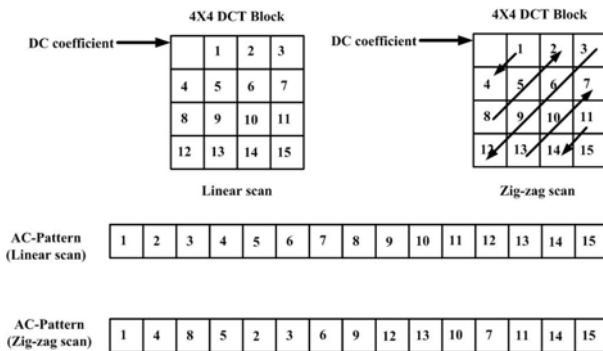
Fig. 3. Linear scan and zig-zag scan

The one is the number of coefficients (nc) used in the AC-Pattern. When this value is big, more high-frequency coefficients are included in the AC-Patterns. As high frequency coefficients have a weak content, it means that the retrieval performance will be more sensitive to the noise. Furthermore, the total number of different AC-Patterns is also bigger and that leads to more time-consuming in image retrieval. When this value is small, the

total number of different AC-Patterns will be small too. Although it leads to less time-consuming, it will decrease the performance of the retrieval too. Unlike skipping the coefficients at the end that have zero values to reduce the size of AC-Pattern [6], we try to find a best value of nc for best performance of retrieval and less time-consuming.

The one is the quantization parameter QP. High QP truncates coefficients leading to zero values. If this value is low, the total number of different AC-Patterns will be very high, that will make the processing of generating histogram more time-consuming and complicated. In contrast, if this value is rather high, there will only be a small number of different AC-Patterns, even all the coefficients could be zero. This will decrease the performance of image retrieval obviously. So there will be a trade-off in quantization parameter between performance and time efficiency.

The one is the number of bins of the histogram that are used to do retrieval (AC-bins for AC-Patterns and DCbins for DC-Patterns). For AC-Pattern histogram, it indicates the total number of bins after merging adjacent patterns. For DC-Patterns histogram, it represents the number of different DC-Patterns chosen to construct DC-Pattern histogram.

Parameter γ and α can also be adjusted. These two parameters have the effect on the performance too.

4   Performance Analysis

GTF (Georgia Tech Face) and ORL (AT&T Laboratories Cambridge) databases are two commonly used databases for face recognition. To evaluate the contribution of our proposal, we implement the described method in [4] and our proposal on GTF database and also compare the performance of our proposal with other statistics-based approaches applied to ORL database.

For performance evaluation we use EER (Equal Error Rate) method [4] [5]. If a value is used to express the similarity between query images and images in the database, so given a certain threshold, an input image of certain class A, may be recognized falsely as class B. Then the ratio of how many images of class A have been recognized as other class is called FRR (False Rejected Rate), while the ratio of how many images of other classes have been recognized into class A is call FAR (False Accept Rate). When both rates take equal values, an equal error rate (EER) is got. The lower the EER is, the better is the system's performance, as the total error rate is the sum of FAR and FRR. One example of EER is showed in Figure 4.
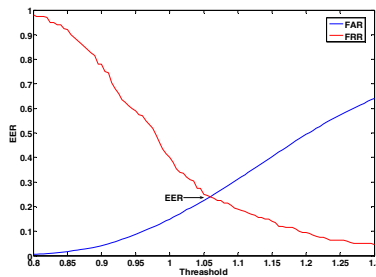

Fig. 4. Example of an Equal Error Rate (EER)

4.1 Application to GTF Database

The GTF database includes 15 different face images of 50 peoples, from both men and women. Most of the images were taken in various illumination conditions, different facial expressions, different scales and orientations, as shown in Fig. 5 and the max degree of the face rotation is about 30°. For experiments, we select the first 11 images of each person as training database and remain 4 images as test images for retrieval. Therefore, the total number of images in the training database is 550 (11×50) and that of test images is 200 (4×50).

Fig. 5. 15 different faces of one person

Firstly, we use AC-Pattern histogram alone for retrieval. We use the method presented in [4] to construct AC-Pattern histograms but using different scanning method and different number of coefficients. Figure 6 shows the curve of comparison. As we can see, zig-zag scan outperforms the linear scan and best result is EER=0.141 when $nc = 4$ .Furthermore, if we use our proposal to construct AC-Pattern histogram, that means not only use zig-zag scan to arrange AC-coefficients in AC-Patterns but also merge adjacent patterns, a lower EER=0.111 is got when $nc = 7$.

Secondly, we use DC-Pattern histogram alone for retrieval. As our proposal focus on AC-Pattern histogram, the purpose of these experiments is to find the best parameters of the histogram for retrieval. We change the values of γ , $QPDC$ and $DCbins$ to see the influence on the performance. We can observe that when $\gamma = 4$, $QPDC = 26$ and $DCbins = 400$, the best performance can be got, the lowest EER is 0.152.

Finally, we use the combination of the AC-Pattern and DC-Pattern histogram to do image retrieval. We give the name to the method presented in [4] as 'linear scan' and the name to the method in which AC-Patterns histogram is constructed by our proposal as 'adjacent zig-zag'. For both methods, we tested different sets of parameters to find the one that can assure the best performance. Finally, for 'linear scan', the best result is obtained when $nc = 4, QPAC = 10, ACbins = 550$. And for 'adjacent zig-zag'

method, the best result is observed when $nc = 7, QPAC = 10$, $ACbins = 35$. And we adjust the weight parameter α to see the global comparison of the performance (Fig. 7 is the curve). We can conclude, from this figure, that the proposal can improve the performance obviously in GTF database. And the best EER of our proposal is 0.087 when $\alpha = 0.25$.
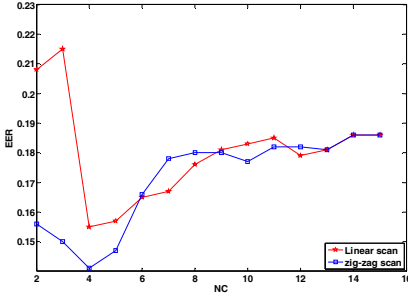
Fig. 6. Performance of AC-Pattern histogram   **Fig. 7.** Performance of different methods
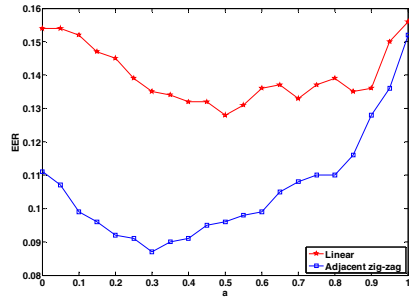
4.2   Application to ORL Database

The ORL database includes 10 different images of 40 peoples. For tests, we use first 6 images as training database and remain 4 images as test images for retrieval. Therefore, the total number of images in the training database is 240 and that of test images is 160. We do similar experiments as we did for GTF database. Again, the adjacent zig-zag scan method outperforms that of linear scan also and the best EER is 0.05.

To illustrate the contribution of our proposal, we compare the performance of the proposed method with those of other methods. Principal component analysis (PCA) is a general method used for identification. It provides an optimal transformation from the original image space to an orthogonal eigen space with reduced dimensionality. However, based on 2D image matrix, two dimension principal component analysis (2D PCA) is proposed for face recognition in recent years. Linear discriminate analysis (LDA) is another typical method that uses the global information of whole image to do face recognition. Table 1 shows the comparison of lowest EER of PCA [7], 2D PCA [8] and LDA [7] [9] reported in recent years and the one we got. The best result of ORL is obtained when: $nc = 4$, $QPAC = 30$, $ACbins = 27$, $\gamma = 3$, $QPDC = 70$, $DCbins = 250$ and $\alpha = 0.5$.

From this comparison, we can conclude that adjacent zig-zag method outperforms all the other methods.

Table 1. Comparison of EER in ORL database with other methods

Method	PCA[7]	2D PCA[8]	LDA[7]	LDA[9]	Our proposal
EER	0.095	0.165	0.096	0.11	0.05

5   Conclusions

In this paper we have presented a new way to construct the histogram of DCT patterns in the context of face recognition. Zig-zag scanning and adjacent patterns merging are proposed for AC-Pattern histogram constructing. When applied for image retrieval on two face databases, widely used for evaluating face recognition algorithms, and by evaluating comparatively their performance, we can conclude that this proposal for AC-Patterns histogram constructing improves retrieval performance. Moreover this approach outperforms other statistics-based approaches.

References

[1] Tsai, T., Huang, Y.-P., Chiang, T.-W.: Image Retrieval Based on Dominant Texture Features. In: 2006 IEEE International Symposium on Indus-trial Electronics, vol. 1, pp. 441–446 (July 2006)

[2] Theoharatos, C., Pothos, V.K., Laskaris, N.A., Economou, G.: Multivariate image similarity in the compressed domain using statistical graph matching. Pattern Recognition 29, 1892–1904 (2006)

[3] Feng, G., Jiang, J.: JPEG compressed image retrieval via statis-tical features. Pattern Recognition 36, 977–985 (2003)

[4] Zhong, D., Defée, I.: DCT histogram optimization for image database retrieval. Pattern Recognition Letters 26, 2272–2281 (2005)

[5] Bolle, R.M., Pankanti, S., Ratha, N.K.: Evaluation techniques for biomet-rics-based authentication systems (FRR). In: Proc. International Conf. on Pattern Recognition, vol. 2, pp. 831–837 (2000)

[6] Daidi, Z.: Image database retrieval methods based on feature histograms. PhD thesis. Tampere University of Technology (May 2008)

[7] Naz, E., Farooq, U., Naz, T.: Analysis of Principal Component Analysis-Based and Fisher Discriminant Analysis-Based Face Recognition Algorithms. In: 2006 International Conference on Emerging Technologies, pp. 121–127 (November 2006)

[8] Xu, Z., Zhang, J., Dai, X.: Boosting for Learning a Similarity Measure in 2DPCA Based Face Recognition. In: 2009 World Congress on Com-puter Science and Information Engineering, vol. 7, pp. 130–134 (2009)

[9] Goudelis, G., Zafeiriou, S., Tefas, A., Pitas, I.: Class-Specific Kernel-Discriminant Analysis for Face Verification. IEEE Transactions on Informa-tion Forensics and Security 2, 570–587 (2007)

[10] Georgia Tech Face Database,
http://www.anefian.com/research/face_reco.htm
(accessed March 2010)

[11] ORL database,
http://www.cl.cam.ac.uk/research/dtg/attarchive/
facedatabase.html (accessed March 2010)

A Representation Model of Images Based on Graphs and Automatic Instantiation of Its Skeletal Configuration

Ali Benafia[1,3], Ramdane Maamri[2,3], Zaidi Sahnoun[3,4],
Sami Saadaoui[1], and Yasmina Saadna[1]

[1] University of El-Hadj Lakhdar Batna - 05000 Batna – Algeria
[2] Computer science department | Campus N.V Ali Mendjli 25000 Constantine – Algeria
[3] LIRE laboratory, C.E.Rsas, University of Mentouri 25000 - Constantine – Algeria
[4] King Faisal University in Al Ahsa - Kingdom of Saudi Arabia
{ali_bnfa,rmaamri,zsahnoun,saadaoui_sam,
saadna_yasmina}@yahoo.fr

Abstract. In this article, we propose an original model of images representation. This work contributes to the subject of the general framework for image indexing and retrieval on the WEB. We describe a method to construct and quantify images from a semantic model, based on graphs, using a set of descriptors. This quantification of the skeleton graph is used for indexing images. In the case of images, the retrieval is mostly done by using text information (image's name, surrounding text, annotations,...) but the information related to content images (color, shape, texture, structure,...) are not considered. It is within this context that the proposed approach allows automatically to calculate for each object of the image (represented by a region) a local descriptor that allows to define its visual and semantic characteristics. This approach therefore presents an extension of the classical model of visual graph used in information retrieval.

Keywords: Representation of graphs, Semantic graph, Image attributes, Low level Descriptors, Semantic concepts, Image segmentation.

1 Introduction

A good choice for image descriptors is critical for the system to be effective and efficient since the indexing techniques become obsolete if the dimension of descriptors increases greatly [6]. To overcome this problem we must: either reduce the number of dimensions of attributes or eliminate descriptors that are less important. In this phase of determination of image descriptors, there must be a compromise between expressiveness, which is provided by many characteristics and performance of the system, which is reduced by the large number of descriptors dimensions. It should also be noted here that the correct choice of features affects not only the indexing phase, but also the procedure to establish the degree of correspondence or similarity between the query and images. Once the descriptors are selected, we associate to the images an index structure which contains all the features of these images. It is within this context that we present here a definition of the image content

with a simple model using a semantic graph used as a support for indexing. For this, we propose to rely on a set of existing work in the state of art. The remainder of this paper is organized as follow; in Section 2 we present the related work in the field of image representation by graphs. We propose our model in Section 3. Section 4 defines the approach to instantiate the skeleton model configuration according to images that will be represented. Section 5 introduces the results of this quantification of the model while comparing the automatic quantification with expert opinion on the basis of empirical practice and finally we end with a conclusion.

2 Related Works

Models based on graphs for the description of the images have widely been studied by different authors, we present in the following section a brief state of the art on the issues discussed.

Early works focused on the use of spatial relationships between image regions for their indexing and their retrieval, descriptions by 2D strings [10] capture the sequences of appearance of objects along one or more reading directions. Thereafter, these works were quickly criticized.

Ounis and al, in [7], describes the original idea of moving a portion of data processing at the time of indexing, so as to speed the matching process. The implementation of this idea is to generate to the indexing a data structure containing all the specializations of sub-graphs of possible queries. The projection problem is thus divided into two sub-problems, one being treated during the indexing and the other at the time of retrieval, it remains to identify sub-graphs composing the query to retrieve all indexes specializing the query. It follows that the processing time of the request becomes polynomial.

In [5], the authors uses a star graph for define the specific types of concepts. Star graphs are simpler than the conceptual graphs and their representations preserves the semantics of conceptual graph formalism .However , their modeling by means of a vector space and expansion of data is likely to generate many irrelevant dimensions, and produce side-effects due to the high dependency between the dimensions of space, in addition, there is an aspect of weighting image objects not yet defined, we can ask the question of completeness of the chosen criteria, and dependencies between these criteria (e.g. a big object has necessarily parts close to the center of the image) from which a lack of an evaluation of the weighting, so as to have starting points to govern the key points of the weighting.

Finally, recent works like in [13] which consider the use of parts of images based on 2D extended HMM (hidden Markov models). Performances of these approaches depend on the network structure and the a priori probabilities in the sense that a network with a good graph structure and a good definition of probabilities gives reliable results.

Other works conducted by Pham and al [8], present an approach that integrates elements of different types in a representation based on graphs, and uses a language model on these graphs for image retrieval. The graph models the image as a set of regions (regions associated with visual concepts), and spatial relationships between

these regions. This representation handles different scenarios, as isolated or grouped images are used as bases for learning or testing.

The results obtained on a problem of categorization of images show first that the automatic procedure that combines the concepts to an image is effective, and secondly that the use of spatial relations, in addition to concepts, improves the quality of classification. This approach presents therefore an extension of the classical model of language in information retrieval to address the problem of unannotated image retrieval and categorization, represented by graphs.

All these models have confirmed not only the stability of visual graphs but also the benefits of integration of the relationship between different types of concepts. Unfortunately they suffer from this lack of difference between the semantic level and signal level of the image content as well as spatial relationships between different types of concepts and they especially suffer from the lack of consideration of the relationships within the facets and between facets of graphs. To address this problematic, we propose a new approach of images representation based on a semantic graph structure.

3   A Model for Images

The model that we present in this section is based on the linguistic content of operating in the ground truth and on visual content extracted from the images.

A process of indexing must use a set of heterogeneous descriptors to answer to a panel of request; this is why we are going to define a model of images description used as a platform for indexing.

Definitions: Formally, S is a set of labels of vertices s and A is a set of labels of arcs a, a labeled graph for a given image is defined by a quadruplet = (S, A, Nodes, Arcs) as:

- **S**: set of vertices denoting all regions of the image by concepts.
- **A**: set of arcs annotated from the vocabulary associated with the image.
- **Nodes** = $\bigcup \text{Desc}(S_i) \subseteq I \times S \times N \times N \times D$: set of quintuples associated with each region **i**, with :

$\text{Desc}(S_i) = ((x_1)_i, (x_2)_i, (x_3)_i, (x_4)_i, (x_5)_i))$ with :

S_i: vertex of the graph that models a region of the image using a concept. We will see its representation later.

$(x_1)_i$: instance of the concept associated with the object that represents the region **i** or NULL in case when the concept is abstract.

$(x_2)_i$: concept associated with region **i**.

$(x_3)_i$: rank of region **i** which corresponds to the address of the first pixel located on the extreme left in the matrix of image pixels.

$(x_4)_i$: plan's level corresponding to the topology of the region i in the image.

$(x_5)_i$: normalized vector (low-level descriptors) of the region **i**.

I: set of instances of concepts, each empty instance is denoted by the mark 'NUL'.

N: set of integer**.**

D: set of descriptors low-level.

   - Arcs $= \bigcup \text{Rel}(S_i) \subseteq A \times S$: set of doublets with $\text{Rel}(S_i)_k = \{((y_1)_i, (y_2)_i)_k\}$ as

k: number of links of vertex S_i with the other vertex.

$(y_1)_i$: label associated with the arc that connects the vertex **$(y_2)_i$**.

$(y_2)_i$: vertex in liaison with vertex S_i via the label **$(y_1)_i$.**

Support model: To describe the skeleton of the graph used for indexing images, it is necessary to associate a support to the description model.

The defined support here is a pair (C, A) with:

C: Set of concepts partially ordered by the subsumption relationship.

A: Set of associations partitioned into subsets of association of the same arity.

These two sets form the conceptual vocabulary for our model.

   ► Concepts: a concept is modeled by the <u>quadruplet</u> (ident, synony, descend, ancest) with:

ident: name given to the concept.

synony: different synonymous terms associated with the concept.

descend: terms of inheriting concept.

ancest: general terms encompassing the concept.

   ► Associations: an association is expressed either by a verb or a nominal entity. It is modeled by the <u>triplet</u> (ident , typ_ass, semant) with:

ident: name given to the association.

typ_ass: verb or nominal entity.

semant: semantic relation associated with the association (subject, location, time, etc....).

4   Instantiating the Model

This process permit to calculate automatically the different characteristics associated with the image and the corresponding regions. We consider three main steps:

Firstly: segmenting the image into regions.

Secondly: extracting low-level visual indices.

Thirdly: interpretation of these indices and their relationship with high-level concepts.

This automatic retrieval process of features is made more workable given the learning method used, starting from a subset of annotated images and to learn correspondence between visual indices and concepts (see details in the following paragraphs). For this, we begin by dividing the image into regions using image segmentation algorithm. For each region we calculate the low-level descriptors and then move to the calculations of semantic descriptors and finally construct a standard vector descriptor for our standard model. All this work leads to the instantiation of our model for the representation of images. This model considered as a solid support for image indexing and retrieval.

1- Segmentation: It must allow us to obtain, from an original source image, a cut-out into regions serving as base for all future processing (indexing, classification ...). It is therefore not to find the best possible segmentation method, but to implement a solution that allows to partition an image to regions so that they roughly correspond at least to certain objects in the image or obvious parts of an object. There are several approaches of segmentation in the literature and the approach retained in our study is the homogeneous region segmentation called CSC [9], this choice is justified by the fact that this method provides better perceptual vision and ensures textual and color homogeneity of image regions (consistency of pixels corresponding to regions on the visual and perceptual framework).

2- Low-level visual descriptors: An image is considered in our model as a whole composite region. Each region generated by segmentation is associated to a descriptor of low level compound of characteristics of color, texture and shape.

Color: We chose the sampled space RGB and additive cumulating of the histograms weighted by Laplacien. This choice is justified by the fact that the method ensures a spatial distribution of colors and offers good performance. The regions issued from the segmentation are heterogeneous and to standardize these regions, we applied the algorithm of "minimum bounding box" for each image region [14]. We associate for each image the corresponding histogram and we cumulate, for the set of regions, the associated histograms.

Texture: There is no relevant definition of texture having regard to irregularity distribution of the basic elements of textures (grid, wall, tissue ...) and random textures (sand, cloud, grass ...).We opt for the choice of the method "local contour patterns" providing histograms with several components (512 in our case). The direction of edges goes through the calculation of the patterns of local contours, then by using a representation of a histogram with 512 possible values. This is justified by the simplicity of the method and especially its proven effectiveness in many applications.

Shape: The shape considered here, use the Freeman coding of translating the course of the contours of the regions following directions (**8-connectedness in our case**). This type of descriptor is used to establish a correlation between regions in the process of image matching. The correlation of two regions r_1 and r_2 represented by the series a_1 , a_2 , ..., a_n and b_1 , b_2 , ..., b_m (n<=m) is defined by:

$$C_{r1r2}(j) = \frac{1}{n}\sum_{i=1}^{n}\cos(a_i - b_{i+j})\times\frac{\pi}{4} \quad (1) \qquad \text{with } j=1 \dots m$$

Where a_i, b_j , represent the various codes associated to directions **8- connectedness**

We add this feature to the region the perimeter and the regions coordinates which are often necessary to establish a connection and a more effective comparison between regions.

3- Calculation of rank region: We call the rank of a region **i** the number of regions above the region **i** from left to right.

The algorithm considered for this computation case helps to determine the status of a region compared to other regions of the image according to a benchmark that is chosen horizontally from left to right. The calculation of rank proceeds through the determination of the first pixel on the left-most among all the pixels constituting the examined region. Two regions that have the same index column will have the same rank. To clarify this method of calculation we consider the following example:

Fig. 1. Matrix of pixels representing regions of a segmented image

In this matrix of pixels there are three regions represented by the respective values 10, 2 and 3.

The rank corresponding to the region 10 is 1.

The rank corresponding to the region 2 is 1.

The rank corresponding to the region 3 is 3.

This means that the region 3 is located after the regions 10 and 2 by going horizontally from left to right.

4- Calculation of plan levels: The approach of separating regions into treated plans in our case is based primarily on the theory of binocular [11] and takes into account the arrangement and dispersion of regions across the surface and the framing of the image. This last , containing all the regions is defined by four borders (top, bottom, left and right) .The definition of the various plans corresponding to different regions is from top to bottom of the frame. At perceptual level, we tend to see an overlay of different backgrounds as a successive cover of widths or lengths of the image and make contact with the boundaries of the image framing. The region that covers the full extent of the border of framing or that covers the greater part of that boundary is able to be a part of the background.

Propriety 1: We say that two regions r_1 and r_2 are located in the same plan if and only if $D_h(r_i) = D_h(r_j)$ with $D_h(r_i)$, $D_h(r_j)$ distances of the peak (highest) regions r_1 and r_2 over the border 'top' the frame.

Algorithm description: The general algorithm of our approach is summarized as follows:

- For each region r_1:

Determine the position of the peak pixel compared to the border of the framing 'top', $D_h(r_j)$ is this position.

Determine the position of the pixel having the minimum distance compared to the border of the framing 'left', $D_g(r_j)$ is this position.

Determine the position of the pixel having the minimum distance compared
to the border of the framing 'right' $D_d(r_j)$ is this position.
- Sort the triplets $(D_d(r_i), D_h(r_i), D_g(r_i))$ according to the associated values to $D_h(r_j)$
- If there are identical triples $(D_d(r_k),D_h(r_k),D_g(r_k))$ (with $k \in [1,n]$ such as $i = k$
and n is the number of segmented regions) then Apply the coverage
procedure two to two between these regions.
- Establish the sequence of triplets $D_d(r_i)$, $D_h(r_i)$, $D_g(r_i)$ already sorted.

Heuristic coverage regions: The originality of our heuristic is based on the theory
of the concept of visual field and binocular vision of each eye that is limited to 62°
which gives an intersection of two fields in the form of a W. This heuristic uses the
calculation of surfaces of regions and their locations compared to the framings 'top'
and 'bottom'.
The principle of the visual field that we baptize **W_space** to decompose the image,
allows to put two peaks on each framing border 'top' and 'bottom' as :

$$peak1 = L/4 \tag{2}$$

$$peak2 = 3*L/4 \tag{3}$$

We note that **L** is the length of the frame 'top' and 'bottom'.
By aligning these four peaks with the two ends of the framing 'top' and 'bottom' of
the image to reach a **W** form called **W_space** .
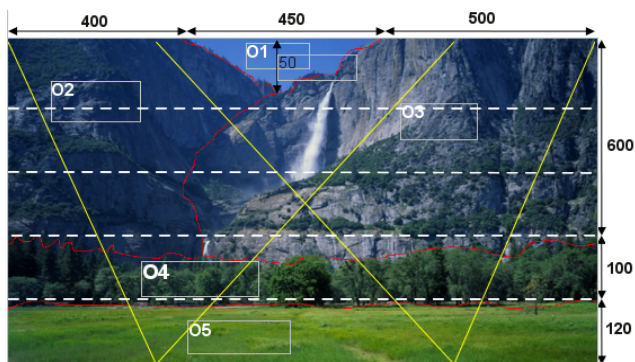
Fig. 2. Landscape drawing with a structure **W_space** (shape W)

Propriety 2: r_i and r_j are the two regions to cover and N(r) is the number of pixels
located on the set of borders 'top', 'left' and 'right' of the region r.

Description of the heuristic:

   - Decompose the image into homogenous horizontal sections depending on the
      number of regions r_i.
   - Calculate the surface Sj representing the number of pixels from the

intersection of different sections with w_space.
- Apply the following formula for each r_i:

$$N(r_i) = \sum_{j=1}^{n}(a_j \times S_j) \qquad (4)$$

With n: the number of pixels in common between **W_space** and sections.

$$a_j = 1 + \ln(p_j) \qquad (5)$$

With p_j: the total number of pixels of the segment j.
    S_j: the number of common pixels between section j and **W_space**.
    - If N(ri) < N(rj) with i ≠ j then rj is located in the first plan compared to ri and vice versa.

5- Semantic descriptors: To automate the labeling of semantic concepts in each analyzed region, it is necessary to establish the signal-symbol correspondence. Such a link called labeling consists in partnering with different regions of images one or more concepts or keywords. For this calculation of semantic descriptors, two steps are required:

- Labeling regions - visual labels.
- Visual tagging labels - semantic concepts.

We present briefly each of these procedures:

- Labeling regions - visual labels: it allows to associate visual labels to segmented regions of image based on low-level descriptors. The visual similarity of the regions requires the use of a learning process. During this process few images issued from the segmentation are given and their use determines the label corresponding to particular regions and the learning process can therefore initiate. Once the learning process completed, the labeling of new regions is done by searching the nearest point (represented in our method by the center of gravity) to visual labels.
In Fig.3, on the left is an image segmented into regions; each one is identified by its center of gravity and on the right the space of corresponding visual labels.
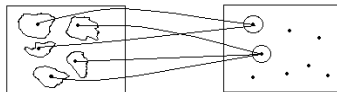
Fig. 3. Labeling regions - labels visual

- Labeling labels - Visual semantic concepts: it can correct the remaining stumbling block not yet remedied during the previous step. The description of the ground truth represents a solid support to overcome the deficiencies caused earlier thanks to this correspondence "label-visual concepts" that we can build from different linguistic structures inherent in the ground-truth (proximity, simultaneity appearance ...). In Fig.4, we can associate automatically to a visual label one or more concepts.
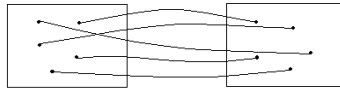
Fig. 4. Labeling of labels visual space (visual vocabulary) and space of concepts (vocabulary of concepts)

To elucidate the main idea of the approach, we consider the example of Fig.5 and present an image to the left and on the right its segmentation.
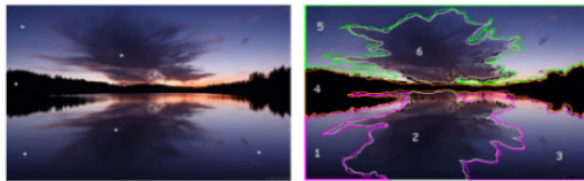
Fig. 5. Source image and segmented image (RGB, threshold = 5)

We note in particular that the sea is represented by regions 1, 2 and 3. By applying the process of regions passage to visual labels, we have this assignment:
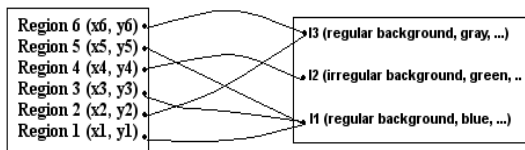
Fig. 6. Labeling space regions - visual space labels

We see that regions 1, 2 and 5 have very similar numerical characteristics under K-Means (classification algorithm chosen in our learning system); likewise for the regions 2 and 6.

In the second passage, the specifications of the ground-truth [1] annihilate the remaining unresolved ambiguities in the previous phase. For example, the specifications ". . . blue sea over the entire surface below the image." can then update the prototype station by removing the link between region 2 and l3 to replace it with the link region 2 and l1 and therefore merge the three regions 1, 2 and 3 substitute for a single region (based on the criterion of proximity and connectivity between regions); likewise for Region 5 (Blue Sky across the high width of the image).
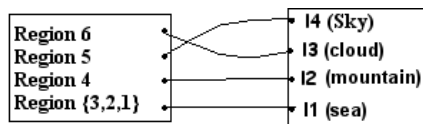
Fig.7. Labeling labels visual space - space of semantic concepts

The region merging procedure that we developed in our system is described by an iterative process that allows for each image region to consider the neighboring regions and test whether they are close in the sense k-means, if they are so we merge them and we update the graph and repeat this process up until there will be no more regions that are close (stability of the graph).

In [2] we can find details of this approach to annotation. The restructuring of the graph is a challenging task and requires a lot of time especially when the number of regions increases ; however, we contemplate in the future to define operators specific to the management of our model (e.g. merging, bursting, reification, etc...), they can take into account the fact that the images or objects to be compared are generally over- or under-segmented ; moreover, they can provide better support to the graph above pairing problems necessary for establishing the similarity between objects or images.

6- Calculation of the resulting descriptor: The performance of automatic generation of descriptors are based on two measures commonly used in the classification, these two measures are the precision and recall and they are often used because they reflect the views of the user: if the precision is low, the user will be dissatisfied with the fact that there is a discrepancy between his views on the descriptors perceived on the images and the descriptors provided by our system, so they do not interest him .For the automatic definition of descriptors used for the quantification of graph from data extracted from the image to model, we have two approaches:

- Each descriptor is calculated separately. The type of classification that is suitable for this solution is the late classification where each descriptor generates its own classification and the final result will be the fusion of these different obtained classifications.
- Group the characteristics issued from different descriptors before classification and as the nature and dimension of feature vectors vary from one descriptor to another, normalizing the values of each vector is performed to obtain the concatenation of these characteristics. This descriptor judged as global is the convenient type of classification for this type of solution that is the early classification.

We will determine in the experimental part from obtained results, which approach is better between the two ones.

5 Experiments

1- Principle: The purpose of this study is to determine from a sample of images (test basis) the appropriateness of different descriptors defined in our model. For this we considered two scenarios:

– consider separately each descriptor.
– consider the merger of descriptors.

Our model contains five types of descriptors corresponding respectively to the color, texture, shape, rank-level and semantic concepts. All these attributes are calculated locally in each region.

The sample used for the protocol is a set of 200 images drawn randomly from the database of images from the site `http://www.niffylux.com`. These images are sorted by category (landscape, building...); the relevance assessment of these features uses two test bases. One of the tests is performed in an empirical way by an expert with the selection of a relevance threshold set in advance, and a second base as a working basis for our model. The expert, with his cognition and knowledge, analyzes and quantifies each image according to the descriptors defined in our model. The instantiation of the model is manually done in this case by the expert. The first base will therefore serve us as a reference-result of an estimation done by the expert, and the second base will contain all descriptors calculated by our system using the algorithms defined in the previous section. The evaluation results obtained by our system are based on data found in the first base.

2- Variables and Parameters: The quantification of the graph structure is based on *two variables* which are: The number of regions that compose the analyzed image and the type of descriptors considered and an unique *parameter* which is the number of categories of semantic concepts (mountain, city, car ...).

To study the quantification of the proposed model, we leave on the basis of a screening of the images over the number of regions they contain that is to say, we take images with two regions then those with three regions and so on until 11 regions, and the images containing more than eleven regions, it means a total of 11 cases to study.

3- Method of evaluation: To evaluate our results, we will support the classical evaluation measures from data mining. These measures are defined by the precision, recall and F-measure. We are therefore interested in calculating the score of the latter which is a balanced measure of precision and recall.

We note the precision P, R the recall, Nt the total number of correct answers provided, Nc the number of responses correct provided by our model and Ni the number of incorrect answers provided by our model

$$P = Nc / (Nc + Ni) \quad ; \quad R = Nc / Nt \quad ; \quad \text{F-measure} = (2 \times P \times R) / (P + R)$$

4- Results: In Table1, we calculated for each class of regions, precision calculations for the different descriptors (color, texture, etc. ...).

For example for the class that contains four regions, the different F-measures for the five preceding descriptors are respectively 0.9607, 0.6039, 0.8824, 0.7194 and 0.6543. Calculation of these indicators is obtained from the arithmetic mean of the F-measure calculated for all tested images.

In Table2, we have merged all the descriptors discussed above into a single standardized global descriptor acquired by a weighted combination of descriptors. The weighting is done according to the degree of descriptors importance in relation with annotations to be done on the images and especially on an attributed favor to the descriptors of large dimensions.

5- Interpretation: Based on values of Table1, we can see that whatever the number of regions and the number of concepts taken, the color descriptor produces results close to the results expected by the expert despite the tiny quadratic quantification error; this is reflected in the choice of the algorithm used to guarantee the

conservation of color as it is perceived. For the evaluation of the other descriptors (texture, region-level and semantic concepts) on the contrary, we observe that the F-measure is low from the images having 5 regions, we can therefore say that the measure of relevance can be judged acceptable from this threshold and beyond that number the accuracy is far from acceptable. Table2 presents insufficient accuracy for the global descriptor combining the different descriptors; moreover the results of this evaluation are not homogeneous. These results provide yet less acceptable performance because of the descriptors combination method using an insufficient weighting to establish that order of importance between descriptors.

Table 1. Calculation of F-measures for a base of 200 images based on 5 descriptors and 11 classes of regions (number of concepts =10)

Descriptors / Number of regions	Color	Texture	Shape	Rank-plan	Semantic concepts
2	0.9425	0.6487	0.8857	0.7538	0.6491
3	0.9537	0.6061	0.8461	0.7347	0.6407
4	0.9607	0.6039	0.8824	0.7194	0.6543
5	0.9089	0.6010	0.9068	0.6209	0.6467
6	0.9664	0.5281	0.8509	0.6028	0.6080
7	0.9823	0.5046	0.8922	0.5443	0.5475
8	0.9251	0.5190	0.8163	0.5069	0.5192
9	0.9105	0.5075	0.8471	0.4786	0.4310
10	0.9342	0.4824	0.8628	0.4755	0.4074
11	0.9018	0.4833	0.8109	0.4240	0.3981
>11	0.9099	0.4073	0.8072	0.3237	0.3024

Table 2. Calculation of F-measures for a base of 200 images based on a single standardized global descriptor obtained from a combination of 5 previously defined descriptors and 11 classes of regions (number of concepts = 10)

Descriptor / Number of regions	global descriptor
2	0.4166
3	0.4420
4	0.3248
5	0.3709
6	0.4928
7	0.3089
8	0.4211
9	0.4981
10	0.3556
11	0.3607
>11	0.2941

In sum, we can conclude that the instantiation of the model is less efficient when the number of regions coming from the segmentation increases as we note that the accuracy of the descriptors is significantly low compared to other measures for n> 11 (n: the number of regions) on the one hand, and on the other hand the merger of descriptors does not give acceptable results (see reasons above). Therefore, we can confirm that the more the image contains regions, the more the computing performance of descriptors (texture, map-region, semantic concepts, global descriptor) degrade due to this rigid parameter due to the heterogeneity of the image indicating the presence of several objects of different natures in an image.

We have done the same evaluations changing at each time the parameter concerning the number of concepts and we have noticed that the higher this parameter is, the higher precision measurements concerning the descriptors are decreased, hence the best performance is obtained from a threshold equal to 6 (6 semantic concepts).

6   Conclusion and Outlook

We introduced in this paper a new approach to indexing images through a graph based on a representation combining several descriptors. We proposed an original model that will be defined as a support for image indexing and retrieval. Image modeling using graphs captures any spatial-topological structures and admits a 2D/3D representation. From the formal point of view, our model fits into the approaches based on labeled graphs model, and extend a number of subsequent work (definition of the model's dynamic). On a practical level, the use of regions and associated concepts allows a gain in generality when describing images, a generality that is beneficial when the use of the system differs from the training environment. This has great chances to occur in practice as we consider collections where one or more images can be used to represent a scene. The proposed model is able to serve as an indexing structure to respond to any form of queries.

As a perspective to this work, we plan to evaluate the similarity of two graphs, this point involves having a similarity measure that may be able to quantify and qualify the difference between two graphs and taking into account the labels associated to components of graphs; that is to say, the similarity of two vertices or two arcs must be based on labels that share and make the approach more tolerant to differences in the sense where two graphs are not identical. For this, it remains to add to the model all the necessary operations for the handling of its concepts, which allows to define a flexibility in the management of skeletons of configurations related to the images. Finally the definition of algebra for our model is necessary and even indispensable for the study of matching graphs, thus the similarity of images.

References

1. Benafia, A., Maameri, M., Saadaoui, S.: Un modèle A.O.V étendu pour les spécifications de la vérité-terrain (to appear)
2. Benafia, A., Maameri, M., Saadaoui, S.: Un système d'apprentissage pour l'annotation automatique des régions d'images issues de la segmentation (to appear)

3. Mechhour, M.: EMIR2: an extended model for image representation and retrieval. In: Database and Expert Systems Applications Conf., London, pp. 395–404 (1995)
4. Hsieh, S., Hsu, C.: Graph-based representation for similarity retrieval of symbolic images. Journal Data –Knowledge-Engineering 65(3) (2008)
5. LMugnier, M., Chein, M.: Représenter des connaissances et raisonner avec des graphes. Revue d'intelligence artificielle 10(1), 7–56 (1996)
6. Mulhem, P., Debanne, E.: A framework for Mixed Symbolic-based and Feature-based Query by Example Image Retreival. International Journal for Information Technology 12(1), 74–98 (2006)
7. Ounis, I., Pasca, M.: The relief retrieval system. In: KDEX 1997(1997)
8. Pham, T., Mulhem, P., Maisonnasse, L.: 3Relations explicites entre différentes représentations d'image dans un modèle de graphe visuel (2009)
9. Priese, L., Rehrmann, V.: A fast hybrid color segmentation method. In: Jn, S.J., Poppl, J., Handels, H. (eds.) Mustererkennunj Symposium Lubeck, Springer, Heidelberg (1993)
10. Smith, J.R., Chang, S.F.: VisualSEEk: A fully automated content-based image query system. In: Proceedings of the Fourth ACM international Conference on Multimedia, pp. 87–98 (1996)
11. Steinman, S.B., Steinman, B.A., Garzia, R.P.: Foundations of Binocular Vision: A Clinical perspective. McGraw – Hill Medical, New York (2000); ISBN 0-8385-2670-5
12. Tappen, M.F., Russell, B.C., Freeman, W.T.: Efficient Graphical Models for Processing. In: IEEE Conference on Computer Vision and Pattern Recognition (2004)
13. Yuan, J., Li, J., Zhang, B.: Exploiting spatial context constraints for automatic image region annotation. In: Proceedings of the 15th international Conference on Multimedia, pp. 25–29 (2007)
14. Chaudhuria, D., Samalb, A.: A simple method for fitting of bounding rectangle to closed regions, publication in the review Elsevier Ltd on behalf of Pattern Recognition Society (2007)

Advice Extraction from Web for Providing Prior Information Concerning Outdoor Activities

Shunsuke Kozawa, Masayuki Okamoto, Shinichi Nagano,
Kenta Cho, and Shigeki Matsubara

[1] Graduate School of Information Science, Nagoya University,
Furo-cho, Chikusa-ku, 464-8601, Japan
kozawa@el.itc.nagoya-u.ac.jp, matubara@nagoya-u.jp
[2] Corporate R&D Center, Toshiba Corporation, 1 Komukai Toshiba-cho, Saiwai-ku,
Kawasaki-shi, 212-8582, Japan
masayuki4.okamoto@tohsiba.co.jp, shinichi3.nagano@toshiba.co.jp,
kenta.cho@toshiba.co.jp

Abstract. Conventional context-aware recommendation systems do not provide information before user action, although they provide information considering users' ongoing activity. However, users want to know prior information such as how to go to their destination or get necessary items when they plan to do outdoor activities such as climbing and sightseeing. It takes time to collect the prior information since it is not so easy to appropriately find them. This paper proposes a method for extracting prior advices from the web. The method first identifies whether a given sentence is an advice or not. Then the method identifies whether the sentence is a prior advice or not if the sentence is identified as advice. In this paper, we will show availability of the proposed method through our experimentation. We also developed a system for providing prior information using the proposed method.

1 Introduction

With the development of mobile devices such as smart phones and tablet PCs, we often access the Internet and get knowledge and information in outside. Therefore, context-aware systems have been researched in recent years [1]. In particular, information recommendation systems considering users' contexts have been developed [2,8,11,13]. However, they did not fully consider contents of information provided to users since their main research interests were to recognize users' contexts and they provided users with prepared information or information based only on users' position.

Users want to know prior information such as how to go to their destination or get necessary items, when they plan to do outdoor activities such as climbing and sightseeing. For instance, when they plan to climb Mt. Fuji, they want to know the means of transportation to Mt. Fuji and the needed climbing gear before leaving. In outdoor activities, users' contexts are divided into two main categories: before or during activities. The existing information recommendation systems considering users' contexts do not provide information before user action, although they provide information considering users' ongoing activity.

Prior information on outdoor activities are written in blogs by many people and stored on the web. We could provide users with valuable information if we use the web as a collective information. However, it takes time to collect the prior information since it is not so easy to appropriately find them using existing web search engines and most web pages containing the prior information include information useful during actual activities. In conventional researches for text mining, sentences containing users' experiences and troubles were extracted[4,7,9,10]. However, the extracted information could not be identified as prior information since they did not consider situations for using the extracted information.

In this paper, we propose a method for extracting prior advices from the web to provide prior information before user action. We first identify whether a given sentence is an advice or not. Then, we identify whether the sentence is a prior advice or not if the sentence is identified as an advice. We also developed a system for providing prior information using the proposed method.

2 Characteristics Analysis of Advices

2.1 The Definition of Advices

Advices are sentences containing information which is worthy to provide users before or during their activities. In the following sections, we assume that the target language is Japanese. Table 1 shows advices about climbing Mt. Fuji. The third column in Table 1 represents whether the sentence is an advice or not. If the sentence is an advice, the value in the column is 1, otherwise, the value is 0. There exist appropriate situations for using the advices. However, the situations are various and depend on activities. In this paper, we assume that situations in which the advices is useful can be categorized into two types; before and during activities. This is because we consider situations which are common to variety of activities. We discriminate between them by

Table 1. Examples of advice sentences

id	sentences	advice or not	prior or not
1	上着は防寒用にウインドブレーカがお 薦め。 (I recommend a windbreaker for protection against the cold.)	1	1
2	登山にあたっては、落石の恐れがありますので、十分ご注意 ください。 (Please note that there is fear of falling rocks when climbing.)	1	0
3	ほぼ毎年富士山を訪れているが、2009 年も 8 月 1 日に富士に登った。 (I visit Mt. Fuji nearly every year and climbed Fuji on August 1 in 2009.)	0	-
4	ゆっくり歩いて、長時間休まないのが、楽 に登るコツですね。 (The knacks of easily climbing Mt. Fuji are to walk slowly and not to take a rest for a long time.)	1	0
5	私がお 薦め するルートがあります。 (There is routes I recommend.)	0	-
6	ただし、お盆休み近辺は マイカー 規制で入れません。 (However, you can't go to Mt. Fuji by driving your own car due to the regulation on private cars during the Bon Festival.)	1	1
7	万一、天候の急変により危険を感じたら最寄の山小屋に避難しましょう。 (If by any chance you feel a sense of danger by the sudden change in the weather, you should take shelter to near mountain lodge.)	1	0
8	体調に余裕を持って高度に慣れながら 登って いきましょう。 (Let's climb Mt. Fuji while keeping on the safe side and becoming accustomed to altitude.)	1	0

judging whether a given advice is needed before taking action. The fourth column in Table 1 represents whether the advice is needed before activities or not. If the advice is needed preliminarily, the value in the column is 1. Otherwise, the value is 0. For instance, the first example is a prior advice and the second one is an action advice.

2.2   Construction of Development Data

We constructed development data to analyze characteristics of advices. Firstly, we searched the web by using "富士山 (Mt. Fuji) & 登山 (climbing)" and "穂高岳 (Mt. Hotaka) & 登山 (climbing)" as queries and collected top 50 web pages from search results for each query. We used Yahoo! Web API [1] to search the web. Next, we extracted texts from the web pages and split them into sentences. Finally, we manually judged whether each sentence is an advice or not and then manually judged whether the sentence is a prior advice or not if the sentence is an advice. Note that we judged them by referring to the previous and next sentences of the sentence. The size of the development data is shown in Table 2.

Table 2. Size of the development data

location	action	# of sentences	# of advices	# of prior advices
Mt. Fuji	climbing	2,581	899	638
Mt. Hotaka	climbing	4,144	360	132
total		6,725	1,259	770

2.3   Characteristics of Advices

We manually investigated 6725 sentences in the development data to capture characteristics of advices. We carefully analyzed the data to capture general characteristics which do not represent a particular domain although the development data represents climbing. Consequently, we found the following five characteristics:

 A. function word at sentence end
   We found that advices are often described in a polite way. While most of sentences described in past tense and interrogative sentences are not advices. In Japanese, tense, aspect and modality of a sentence are often represented by function words at the end of the sentence. Therefore, we capture these writing styles by focusing on function words at the end of a sentence as shown in the underlined portions in the examples 2 and 3 in Table 1.
 B. evaluation expressions
   We found that advices often contain expressions having a semantic polarity. The example 4 is an advice containing positive or negative evaluation expressions and the underlined portions represent evaluation expressions.

C. clue expressions

We manually generated 470 expressions which were often contained in advices and classified them into 35 classes based on their meaning. Some of the expressions are shown in Table 3. The example 1 is an advice containing the clue expressions in the underlined portions.

D. sentence end information

We found that appearance position of evaluation and clue expressions in a sentence is important. For example, a clue expression "薦め (recommend)" is appeared at the end of the sentence in the example 1 which is an advice. While the example 5 is not an advice even though it contains the same clue expression. As seen in these examples, the advices often contain evaluation and clue expressions at the end of a sentence since a content word at the end of a sentence often plays important role in representing the sentence meaning in Japanese.

E. context information

It is often the case that the previous and next sentences of an advice are also advices since advices are often collectively described in a web page. Thus, the evaluation and clue expressions frequently appear in the previous and next sentences of an advice.

Table 3. Clue expressions which represent characteristics of advices

class	clue expression
recommend	薦め，方が良い (recommend, have better, rather ... than, etc.)
warning	気を付ける，注意 (care, note, attention, etc.)
prepare	用意，準備，持参 (prepare, ready, equip, etc.)
inhibition	禁止，禁物，厳禁，避ける (inhibit, prohibit, forbid, avoid, etc.)
necessary	必要，必須，必需 (necessary, essential, require, etc.)
schedule	計画，工程 (schedule, route, plan, etc.)
crowd	渋滞，混雑，満員 (traffic jam, crowded, full, etc.)
business	営業，開店，閉店，閉鎖 (open, closed, in business, etc)
occasion	場合，際，とき，状況 (occasion, when, in case, scene, etc.)
emergency	万が一，いざというとき (if by any chance, when the chips are down, etc.)

2.4   Characteristics of Advices Suitable for Situations

We manually investigated 1259 advice sentences in the development data to capture characteristics of prior and action advices, and found the following four characteristics:

A. clue expressions

We manually generated 457 expressions which often appeared in prior advices and classified them into 17 classes based on their meaning. We also manually generated 109 expressions which were often contained in action advices and classified them into nine classes based on their meaning. Some of the expressions are shown in Table 4. The class names attached with * in the first column in Table 4 represent clue expressions for action advices. The examples 6 and 7 are advices containing the clue expressions in the underlined portions.

Table 4. Clue expressions which represent situations of advices

class	clue expression
traffic	車，電車，駅，国道 (car, train, station, national road, etc.)
prepare	用意，準備，持参 (prepare, ready, equip, etc.)
preliminary	事前，予め，前もって，未然に (preliminary, on ahead, before happens, etc.)
weather	天気，気温，雨，暑い (weather, rain, temperature, hot, etc.)
impossible	不可，できない (impossible, can not, unable)
careful*	慎重，きちんと，しっかり (carefully, neatly, accurately, etc.)
knack*	コツ，ポイント，仕方 (knack, point, know-how, how to, etc.)
emergency*	万が一，いざというとき (if by any chance, when the chips are down, etc.)
possible*	可能，できる (can, possible, able, etc.)

B. action verbs

The action advices often contain action verbs which represent bodily movement. The example 8 contains an action verb "登る (climb)".

C. sentence end information

We found that prior advices often contain clue expressions for prior advices at the end of sentence and action advices often contain clue expressions for action advices and action verbs at the end of sentence.

D. context information

The advices are frequently described collectively. Therefore, the clue expressions for prior advices are often contained in the previous and next sentences of a prior advice. The clue expressions for action advices and action verbs are often contained in the previous and next sentences of an action advice in the same way.

3   Prior Advice Acquisition

Figure 1 shows the flow for extracting prior advices. Firstly, we search the web by using a location name and an action name as a query and get HTML texts whose title includes the location name. Yahoo! Web API is used for searching the web. Secondly, we preprocess the HTML texts and extract sentences from them. Thirdly, each sentence is identified whether it is an advice or not by a classification learning. Finally, each extracted advice is identified whether it is a prior advice or not.
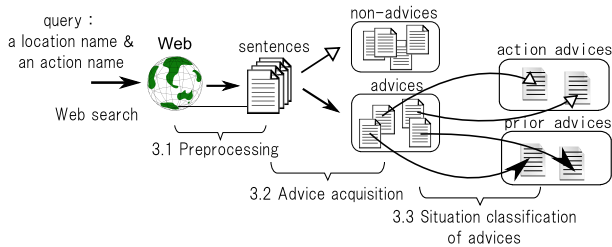
Fig. 1. Flow of prior advice acquisition

3.1   Preprocessing

Sentences are extracted from HTML texts by the following method. The texts enclosed by the HTML tags are extracted if "id" or "class" attributes in HTML tags contain any of the following strings.

contentCentryCmain

Otherwise, the texts enclosed by the body tags are extracted. Note that the texts enclosed by the HTML tags are eliminated from the targets if "id" or "class" attributes in HTML tags contain any of the following strings.

head, foot, menu, copy, list, comment

Then each sentence extracted from the texts is segmented into morphological elements using MeCab[6] and the sentences which meet the following requirements are extracted. The stopword list is used to eliminate advices for browsing web pages.

- The number of morphemes in the sentence is more than five.
- The sentence contains any of verbs, adverbs, adjectives or auxiliary verbs.
- The sentence does not contain any of the stopwords (ex. submit, account, browser, Adobe, JavaScript, spam, comment).

3.2   Advice Acquisition

We identify whether a given sentence is an advice or not by using a classification learning. The characteristics of advices described in Section 2.3 are used as features of the classification learning. The features for acquiring advices are shown as follows. The features through a to e correspond to the characteristics through A to E in Section 2.3, respectively.

 a. An auxiliary verb at the end of the sentence is any of the following auxiliary verbs.

   'ぬ, ん, ない', 'べし', 'だ', 'たい', 'た', 'う', 'です', 'ござる', 'ます', 'らしい', etc.
 b. The frequencies of evaluative expressions (13,590 expressions for four classes)
   The evaluative expressions in the dictionaries [3,5] are used.
 c. The frequencies of clue expressions (470 expressions for 35 classes)
 d. Whether the content word (noun, verb, adjective and adverb) at the end of the sentence is evaluative and clue expression or not.
   Note that we checked its previous content word if the content word is one of the following words

   '下さる', '思う', 'する', 'いる', 'くれる', 'できる', 'おく', 'れる', 'られる', 'の', 'こと', 'もの'
 e. Through b to d for the previous and next two sentences of the target sentence
 f. The morpheme length
 g. The ratio of the number of morphemes to the morpheme length for each part-of-speech

3.3   Situation Classification of Advices

We identify whether a given advice is a prior advice or not by using a classification learning. The characteristics of advices described in Section 2.4 are used as features of the classification learning. The features for classifying advices are shown as follows. The features through *a* to *d* correspond to the characteristics through *A* to *D* in Section 2.4, respectively.

- *a.* The frequencies of clue expressions for prior (457 expressions for 17 classes) and clue expressions for action (109 expressions for nine classes)
- *b.* The frequencies of action verbs (447 verbs for two classes)
  We used 76 verbs which represent movement and 371 verbs which represent body movement in a thesaurus of predicate-argument structure [12].
- c. Whether the content word (noun, verb, adjective and adverb) at the end of the sentence is the clue expressions and the action verbs or not.
  Note that we checked its previous content word for some of the content word in the same way as the feature *d* in Section 3.2.
- d. Through *a* to *c* for the previous and next two sentences of the target sentence

4   Experiment

We carried out experiments to show the performance of the proposed method and to validate our observed features.

4.1   Evaluation Data

We constructed evaluation data. Firstly, we searched the web by using "高尾山 (Mt. Takao) & 登山 (climbing)" as a query and extracted top 20 web pages whose title included the location name "高尾山 (Mt. Takao)". We used Yahoo! Web API to search the web. Secondly, sentences were extracted from the web pages by the preprocessing method described in Section 3.1. Finally, we manually judged whether each sentence was an advice or not and then judged whether the advice was a prior advice or not if the sentence was an advice. The size of the evaluation data is shown in Table 5.

4.2   Experiment for Acquiring Advices

We carried out experiments by training 6725 sentences in the development data as training data and testing 1335 sentences in the evaluation data as testing data. As features

Table 5. Size of evaluate data

# of sentneces	# of advices	# of prior advices
1,335	172	107

Table 6. Experimental results for acquiring advices

feature	precision	recall	f-measure
word uni-gram (baseline)	49.4%	26.2%	34.2
proposed method (features through a to e)	61.2%	35.7%	45.1
proposed method - function word at sentence end (feature a)	56.4%	31.5%	40.5
proposed method - evaluate expressions (feature b)	56.5%	36.3%	44.2
proposed method - clue expressions (feature c)	-	0	0
proposed method - sentence end information (feature d)	43.4%	25.6%	32.2
proposed method - context information (feature e)	53.7%	34.5%	42.0

Table 7. Experimental results for classifying situation of advices

feature	precision	recall	f-measure
word uni-gram (baseline)	68.7%	76.7%	72.5
proposed method (feature through a to d)	75.0%	81.3%	78.0
proposed method - clue expressions (feature a)	62.0%	94.4%	74.8
proposed method - action verbs (feature b)	73.1%	73.8%	73.5
proposed method - sentence end information (feature c)	72.9%	80.4%	76.4
proposed method - context information (feature d)	66.9%	75.7%	71.1

for a classification learning, the features described in Section 3.2 were used. We used Support Vector Machines (SVM) as a machine learning model and trained SVM with a linear kernel using LibSVM[2]. We used the precision (ratio of the number of successfully acquired advices to the number of automatically acquired advices), the recall (ratio of the number of successfully acquired advices to the number of advices in the evaluation data) and the f-measure as metrics. In the baseline method, we used word uni-gram as features.

Experimental results are shown in Table 6. The results when each feature was eliminated from the proposed method are shown in through forth to eighth rows in Table 6. In comparison with the baseline method, the proposed method increased in both the precision and the recall. The f-measure decreased when each feature was eliminated from the proposed method. This results show that our features obtained by the analysis described in Section 2.3 are valid.

4.3   Experiment for Classifying Situation of Advices

We carried out experiments by training 1259 advice sentences in the development data as training data and testing 172 advice sentences in the evaluation data as testing data. As features for a classification learning, the features described in Section 3.3 were used. We used SVM as a machine learning model and trained SVM with a linear kernel using LibSVM. We used the precision (ratio of the number of successfully acquired prior advices to the number of automatically acquired prior advices), the recall (ratio of the number of successfully acquired prior advices to the number of prior advices in the evaluate data) and the f-measure as metrics. In the baseline method, we used word unigram as features.

Table 7 shows the experimental results. Both the precision and recall obtained by using the proposed method showed more increase than the baseline method. The

---

[2] http://www.csie.ntu.edu.tw/~cjlin/libsvm/

f-measure decreased when each feature was eliminated from the proposed method. This results show that our features obtained by the analysis described in Section 2.4 are avalible for classifying situations of adivces.

5   Conclusion

In this paper, we proposed a method for extracting prior advices from the web to provide prior information before user action. The experimental results show the availability of our method. We also developed the system for providing prior advices.

For future works, we will identify whether a given advice is related to the target location and the target action or not. In addition, we would like to consider more detailed situations.

References

1. Baldauf, M., Dustdar, S., Rosenberg, F.: A survey on context-aware systems. International Journal of Ad Hoc and Ubiquitous Computing 2(4), 263–277 (2007)
2. Cheverst, K., Davies, N., Mitchell, K., Friday, A.: Experiences of developing and deploying a context-aware tourist guide: the GUIDE project. In: Proceedings of the 6th Annual International Conference on Mobile Computing and Networking, pp. 20–31. ACM, New York (2000)
3. Higashiyama, M., Inui, K., Matsumoto, Y.: Acquiring noun polarity knowledge using selectional preferences. In: Proceedings of the 14th Annual Meeting of the Association for Natural Language Processing, pp. 584–587 (2008)
4. Inui, K., Abe, S., Morita, H., Eguchi, M., Sumida, A., Sao, C., Hara, K., Murakami, K., Matsuyoshi, S.: Experience mining: Building a large-scale database of personal experiences and opinions from web documents. In: Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence, pp. 314–321 (2008)
5. Kobayashi, N., Inui, K., Matsumoto, Y., Tateishi, K., Fukushima, T.: Collecting evaluative expressions for opinion extraction. In: Proceedings of the 2nd International Joint Conference on Natural Language Processing, pp. 584–589 (2004)
6. Kudo, T., Yamamoto, K., Matsumoto, Y.: Applying conditional random fields to japanese morphological analysis. In: Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, pp. 230–237 (2004)
7. Kurashima, T., Fujimura, K., Okuda, H.: Discovering association rules on experiences from large-scale blog entries. In: Boughanem, M., Berrut, C., Mothe, J., Soule-Dupuy, C. (eds.) ECIR 2009. LNCS, vol. 5478, pp. 546–553. Springer, Heidelberg (2009)
8. Oku, K., Nakajima, S., Miyazaki, J., Uemura, S., Kato, H.: A recommendation method considering users' time series contexts. In: Proceedings of the 3rd International Conference on Ubiquitous Information Management and Communication, pp. 465–470. ACM, New York (2009)
9. Park, K.C., Jeong, Y., Myaeng, S.H.: Detecting experiences from weblogs. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pp. 1464–1472. Association for Computational Linguistics (2010)
10. Saeger, S.D., Torisawa, K., Kazama, J.: Looking for trouble. In: Proceedings of the 22nd International Conference on Computational Linguistics, pp. 185–192. Association for Computational Linguistics (2008)

11. van Setten, M., Pokraev, S., Koolwaaij, J.: Context-aware recommendations in the mobile tourist application COMPASS. In: International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, pp. 235–244 (2004)
12. Takeuchi, K., Inui, K., Takeuchi, N., Fujita, A.: A thesaurus of predicate-argument structure for japanese verbs to deal with granularity of verb meanings. In: Proceedings of the 8th Workshop on Asian Language Resources, pp. 1–8 (2010)
13. Zheng, V.W., Zheng, Y., Xie, X., Yang, Q.: Collaborative location and activity recommendations with gps history data. In: Proceedings of the 19th International Conference on World Wide Web, pp. 1029–1038. ACM, New York (2010)

Automatic Composition of Presentation Slides, Based on Semantic Relationships among Slide Components

Toyohide Watanabe, Yusuke Ishiguro, and Koichi Hanaue

Department of Systems and Social Informatics,
Graduate School of Information Science, Nagoya University
Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan
watanabe@is.nagoya-u.ac.jp

1 Introduction

Today, presentation slides are often used at the congresses, lectures, discussions/meetings, etc. as useful/powerful support tools for speakers. However, currently usable presentation tools such as Microsoft PowerPoint [1], and Apple Keynote [2], may not be always convenient for speakers, who must talk along a sequence of slides preset in composition phase. Namely, composition phase and presentation phase are completely separated in the conventional tools: in the composition phase the speaker edits his slides individually with the related topics and sets the sequence under his presentation story; and in the presentation phase he must talk along the preset sequence, but cannot rearrange the sequence easily or use other slides corresponding to the reactions of audiences without any overheads. The traditional presentation support tool is very rigid but not flexible.

Such a framework disturbs speakers from timely assigning the correlated relationships to predecessor/successor slides on his way.　Tufte pointed out that the slides which were composed by the presentation tools do not have consistent relationships among contents because the key parts of corresponding contents in each slide are not always clear [3, 4]. Various researches, which compose automatically the presentation slides or support the preparation process of slides effectively, have been reported. Yasumura et al. proposed a method to generate the corresponding slides in HTML from a paper in LaTex style-file [5]. Also, Miyamoto et al. proposed a method to compose the slides from ones in LaTex style-files [6]. This method not only extracts some information about the title, figures and so on, based on the commands embedded in LaTex style-file, but also identifies the connective relationship among sentences by using conjunctions. Moreover, Kurohashi et al. reported an interesting means to first extract knowledge segments called knowledge cards from a paper, and then generate automatically appropriate slides from the knowledge cards [7]. These methods make it possible to generate the enumeration type of slide structures, but impossible to decode or allocate items so as to represent the contents effectively in 2-dimensional slide sheets.

In this paper, we address a slide composition method to make up various kinds of components with respect to the visual understandability and the structural/semantic slide structure. Our idea is to compose slides from the presentation scenario, predefined by another specification tool [8]. In order to make the transformation between

two different components successful, we introduce semantic relationships to distinguish the related features of presentation scenario, and determine the allocation procedure on the basis of the related features.

2   Approach

2.1   Document Structure

Usually, document structure is defined with logical structure and geometric structure [9]. The logical structure is semantic relationships among document elements such as the title, texts, sub-texts, figures, tables, footnotes and so on. The elements have their own physical features: the size, length, location, font style and size in texts, etc. The semantic relationships are assigned among two or more elements: is-a, part-of, grouping, etc. The geometric structure is related to the physical positions of individual elements from a location-oriented point of view on the sheet: neighboring, left-right, upper-lower, overlap, etc. Thus, it is important to transform the elements from logical structure to geometric structure. Figure 1 shows this transformation mechanism. The transformation is dependent on the semantic relationships among the elements in the logical structure and the physical properties of elements, and is performed interpretatively from the upper elements in the logical structure, which is specified commonly by a tree structure.

   We introduce the idea to transform the properties and relationships from the logical structure to the geometric structure. Here, our description in the logical structure is corresponded to the presentation scenario which can be interactively composed with other tools [8]. Figure 2 shows an example of scenarios displayed on window. Individual main items are linked among related items, and circles surround directly
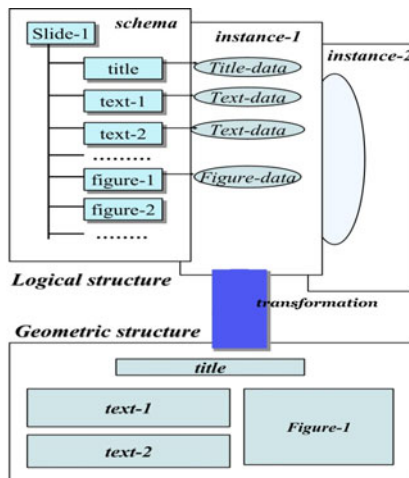

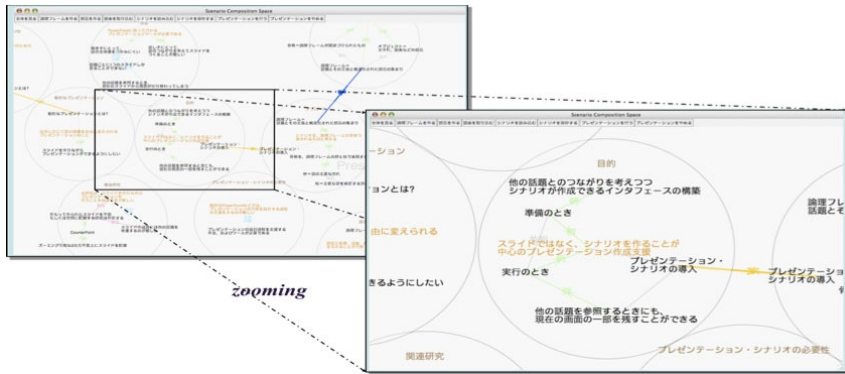**Fig. 1.** Transformation from logical structure to geometric structure

Fig. 2. Presentation scenario

Table 1. Semantic relationship

category	sub-category	group
and	cause/ result	brother
	condition/ consequence	
but	opposition	
	conciliation	
particulars	outline/ details	child
	exemplification	
others	complement/ comment	
	parallel	brother
	others	

Table 2. Condition and means in modification

modification plan	condition	means
insert title	when there is no title or corresponding tagged term	by dialog
insert image	when there is no image	by dialog
insert key-point	when there is no key-point	by dialog
decrease components	when number of texts is more than "m"	by dialog
simplify text string	when number of characters in text is more than "n"	high-light of corresponding term
use name-form	when sentence is used	high-light of corresponding term

related items. A set of items surrounded by the circle are candidates for slide components in one sheet. On the presentation scenario, the semantic relationships between individual items are denoted meaningfully by colored arrows, and can be interpretatively set into the locations in the geometric structure. Table 1 shows the semantic relationship to be attached with links.

2.2 Processing Flow

Our basic mechanism is the mapping based on the presentation scenario, and the processing flow is mainly composed of two different steps: editing and allocation. The logical structure (scenario) and properties of items are set at the editing step, and the geometric structure (template slide) is composed in the allocation step.   In the editing step, the following rules are usable for components to be edited:

1) The amount of components in a slide is not too much;
2) The linguistic representation in a text component is simple;
3) Figures are included.

If these rules cannot be well applied, the messages for modification request and new modification plan are indicated. On the other hand, in the allocation step two composition rules are mainly effective:

1) Easy to grasp the component groups meaningfully;
2) Necessary to represent explanation contents in figure-like form.

In order to grasp the component groups meaningfully, it is reasonable that the components to be assigned to the same groups should be located in neighboring positions and that the components which belong to different groups are assigned comparatively to independent positions. Also, in order to represent explanation contents in figure-like form, it is effective to specify the relationships among components by directed arrows or make use of templates which can represent visually the relationships among components, such as SmartArt in PowerPoint. Our approach is to first group components based on the semantic relationship, and then compose slides by using appropriate templates which are consistent to semantic relationships. In semantic relationships, components which have inclusive relationships such as "outline/details", "complement/comment", etc. can be collected into the same group, while the components which have comparative relationships such as "opposition", "cause/result", etc. can be collected into different groups.

3   Slide Editing

3.1   Semantic Relationship

The slide component c is defined as the following five-terms tuple:

$$c = (\ id,\ content,\ type,\ role,\ size\ ) \tag{1}$$
$$type \in \{\ text,\ figure\ \}$$
$$role \in \{\ title,\ point,\ normal\ \}$$

Here, *id* is an identifier of component and *content* is an explanation instance like character strings in *text*, and figures in *figure*. Also, *type* is the kind of components, and *role* is the role for topic, related to c. For example, *role* is *title* of topic, abstract of topic and normal component. *size* is available for *figure*: in this case, *size* keeps horizontal length and vertical length. The semantic structure *SSt* for a topic t, which is specified by the corresponding semantic relationship, is:

$$SSt = (\ Ct,\ Rt\ ) \tag{2}$$

Here, *Ct* is a set of slide components in t, and *SSt* is represented as a directed graph which regards the component in *Ct* as the node. *Rt* is a set of directed edges, and corresponds to a set of relationships among components in *Ct*.

$$Rt = \{\ (\ type,\ c_{src},\ c_{dst}\ ) \mid c_{src},\ c_{dst} \in Ct\ \} \tag{3}$$

Here, *type* is a semantic relationship between components c_{src} and c_{dst}, and (*type*, c_{src}, c_{dst}) represents a directed edge from c_{src} to c_{dst} with label *type*.

3.2 Editing Operation

Basic editing operations are:

1) Add, update and delete components;
2) Assign new attributes to components;
3) Assign semantic relationship to combine two or more components.

In the editing phase, when the currently composed slides are not always consistent to conventional rules, contained often in the guideline [4] to judge whether these rules should be applicable, the system prompts individual modification plans and keeps the corresponding conditions and means applicable to edited slides. Table 2 shows such conditions and means. In this case, the modification plan is fired when one of these conditions is particularly matched with an arbitrary semantic relationship among all components.

4 Slide Composition

Our slide composition process is organized as illustrated in Figure 3. The first step is to group correlated components based on semantic relationships. The second step is to compose the presentation form of slide with respect to the grouped components. The final step is to select a template suitable to this presentation form and then allocate the corresponding components into a slide sheet, according to the selected template.
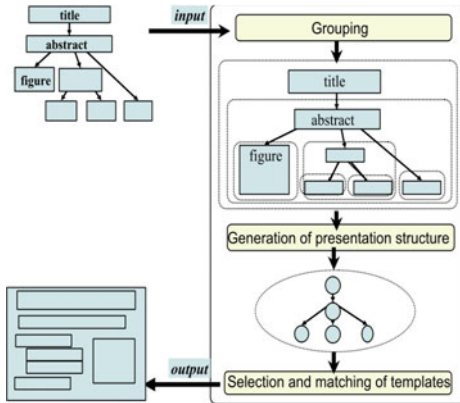
Fig. 3. Slide composition process   **Fig. 4.** Interface for modification
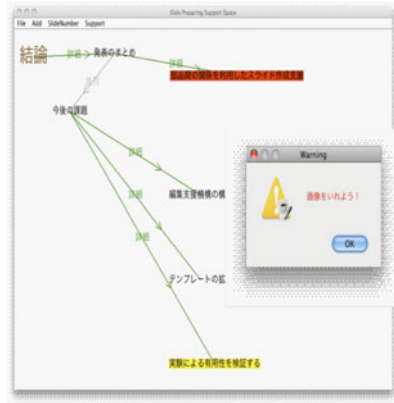
4.1 Grouping of Slide Components

The group is a set of strongly related slide components. This grouping procedure distinguishes two different types of components: one is to extract a collection of components contained meaningfully by a component c; and the other is to identify a set of components, which are equivalent to c. The former group is called a child-group for c, and the latter is called a brother-group for c. Table 1 shows such two groups for

semantic relationships. For a topic t, a group of various components in the semantic structure SSt is called the presentation structure for t and is expressed as ESt. ESt is a tree structure whose root is a title, whose nodes are independent components such as components of child-group or brother-group, and whose edges are semantic relationships between components.

$$ESt = \{ (c, G_c^C, G_c^N) \mid c \in Ct \} \tag{4}$$

Here, G_c^C is the child-group for c, G_c^N is the brother-group for c.

$$G_c^C = \{ (type, G_{c'}) \mid (type, c, c') \in Rt \} \tag{5}$$
$$G_c^N = \{ (type, G_{c'}) \mid (type, c, c') \in Rt \}$$

Here, $type$ is a semantic relationship to be defined between c and c'. The procedure generates ESt from the semantic structure for t, based on component group. **Algorithm-1** is a procedure to generate ESt from SSt for t.

【**Algorithm-1:** grouping of slide components】
Make-groups(SSt)
$ESt \leftarrow$ Make-groups(title(SSt))
return ESt

Make-group(c)
$G_c^C \leftarrow \varphi$
$G_c^N \leftarrow \varphi$
for each ($type$, c, c') **in** Rt **do**
  if $type \in$ {outline/details, exemplification, complement/comment} **then**
    $G_{c'}^C \leftarrow$ Make-group(c')
    $G_c^C \leftarrow G_c^C \cup \{(type, G_{c'}^C)\}$
  else if $type \in$ {cause/result, condition/consequence, opposition, conciliation,
                parallel, others} **then**
    $G_{c'}^N \leftarrow$ Make-group(c')
    $G_c^N \leftarrow G_c^N \cup \{(type, G_{c'}^N)\}$
  end if
end for
return (c, G_c^C, G_c^N)
■

4.2   Template-Based Slide Composition

Our template is typically a slide frame which is collected from practically used slides or designed abstractly as a semantic unit. Its form specification indicates in detail the semantic relationships among components in comparison with template sheets, supplied in PowerPoint. The group is interpreted as a node in the tree structure, and the connection among groups and the semantic relationship on a group collection are corresponded as edge and its label, respectively. Thus, the appropriate template can be

selected by comparing the presentation structure of template with subsets of semantic structure among components in a slide sheet.

Templates suitable to the structure among slide components are selected after we have compared the presentation structure to be effective for allocation with a tree structure on the corresponding template. We looked upon the slide in which we could allocate all components completely as a certain slide to be generated. After we have selected appropriate templates, the corresponding components are allocated into the pre-assigned areas of template when the applicable conditions are satisfied. In this case, the allocated areas are composed of area which is useful to allocate groups, and area which keeps in advance for future usage. Our applicable conditions mainly consist of two different criteria: the condition for structure indicates whether the given group and its related un-allocated group are consistent to the presentation structure of template; and the condition for area points out whether the components can be located sufficiently in the assigned area.

5   Prototype System

Our prototype system was implemented in JAVA.   Also, the slide displayed in our allocation interface is usable on Apache POI [10], usable in Microsoft Office.

5.1   Component Editing Interface

Interface of editing slide components was implemented, using Piccolo [11] as a zooming interface API. Also, the lexical analyzer, which was used to detect modification points, is Sen [12] in JAVA. The menu "add" indicates to add a new component such as image, text, etc. into the existing template and also the menu "support" checks whether the existing slide was organized suitably under our conditions or not; if not suitable, the slide plan to be appropriately modified is indicated. An example is shown in Figure 4. In this example, since the slide does not include even an image component, the system advises to use image components so as to make the total balance stable and keep the high visualization by a message "画像をいれよう (use image components)". Additionally, the modification message, in which the corresponding text is colored by "yellow", is promoted since the system regarded that the description style is different from others.

Moreover, operations for deletion, update and tagging are available. The tagging assigns an attribute such as "title", "abstract", "important viewpoint" and so on to the newly selected components, when any tags are not attached to the attributes of components. The update operation modifies the contents of components. In the text update, the component is updated by exchanging string contents in dialog. On the other hand, in the image update the size is adjusted in special window, as shown in Figure 5.

5.2   Display Interface

This window interface enforces to choose the most preferable slide form several candidates, generated by the system. The interface is shown in Figure 6. The candidates are arranged in the lower side of display window. The appropriate slide can be chosen
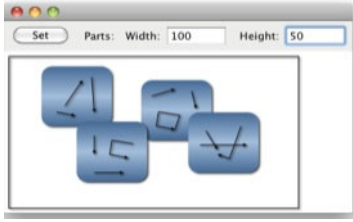
Fig. 5. Arrangement interface of figure
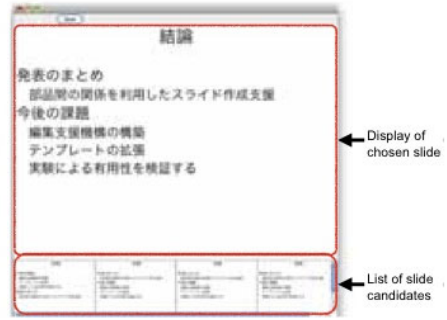
Fig. 6. Interface for slide selection
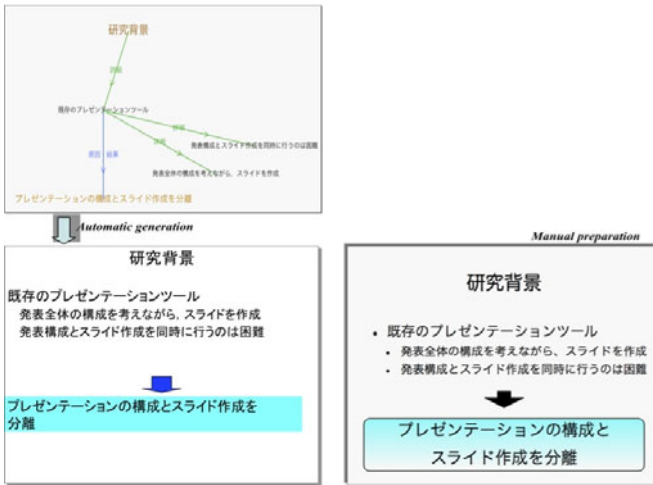
Fig. 7. Example of slide composition (1)

in this case and the chosen slide is displayed in the main area.   Moreover, the chosen slide is output in PowerPoint form.

Here, we show examples in Figure 7 and Figure 8. In individual figures, the upper illustration is our presentation scenario, the left form is a slide which is automatically generated by our prototype system with the interpretation of the corresponding presentation scenario (in the upper illustration). The right form is a slide made manually by the original speaker. Though system-made slides are not always well structured in points of space allocation in comparison with hand-made slides, the geometric structure is well evaluated.
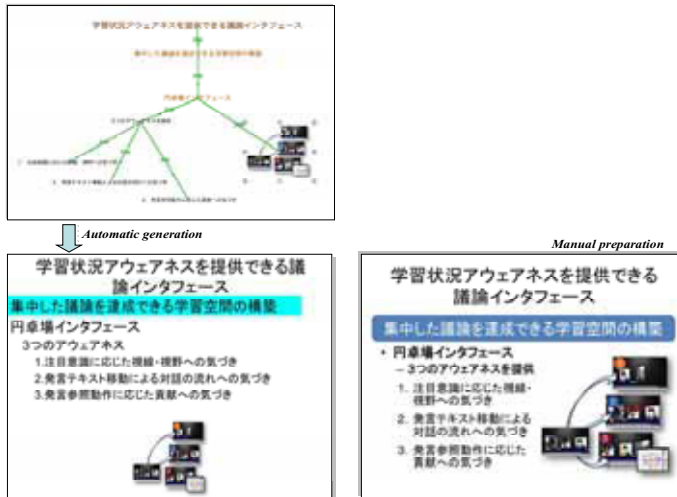
Fig. 8. Example of slide composition (2)

6 Conclusion

In this paper, we proposed an experimental method to compose automatically presentation slides, based on the transformation mechanism from the logical structure to the geometric structure. Although this concept of logical structure and geometric structure with a view to interpreting document organization has been successfully investigated as research fields of document image understanding, in our case this transformation takes an important role to compose the presentation slides: the logical structure is the presentation scenario; and the geometric structure is the slide template. Our idea was to assign the semantic relationships to the related slide components in order to attain the transformation powerfully. Some slides, composed automatically by our prototype system, are successfully generated in our experimental evaluation. Of course, it is necessary to improve the ability in order to compose more excellent slides.

References

1. Microsoft PowerPoint, http://office.microsoft.com/powerpoint/
2. Apple Keynote, http://www.apple.com/iwork/keynote/
3. Tufte, E.R.: The Cognitive Style of PowerPoint. Graphics Press (2004)
4. Reynolds, G.: Presentation Zen. New Riders Press, Indianapolis (2007)
5. Yasumura, Y., Takeichi, M., Nitta, K.: A Support System for Making Presentation Slides. Trans.of JSAI 18(4), 212–220 (2003) (in Japanese)
6. Miyamoto, M., Sakai, H., Masuyama, S.: Research on Automatic Generation of Presentation Slides from a LaTex Manuscript of a Paper. Trans.of JSFTII 18(5), 752–760 (2006) (in Japanese)
7. Shibata, T., Kurohashi, S.: Automatic Slide Generation Based on Discourse Structure Analysis. Journal of Natural Language Processing 13(3), 91–111 (2006)

8. Hanaue, K., Watanabe, T.: Externalization Support of Key Phrase Channel in Presentation Preparation. International Journal of Intelligent Decision Technologies 3, 85–92 (2009)
9. Watanabe, T., Sugie, N.: Structure Recognition Methods for Various Types of Documents. International Journal of Machine Vision and Applications 6, 163–176 (1993)
10. The Apache POI Project, http://poi.apache.org/
11. Bederson, B.B., Grosjean, J., Meyer, J.: Toolkit Design for Interactive Structured Graphics. IEEE Trans.on Software Engineering 30(8), 535–546 (2004)
12. Sen, http://sen.dev.java.net/

Sustainable Obsolescence Management – A Conceptual Unified Framework to Form Basis of an Interactive Intelligent Multimedia System

T.E.[*] Butt, J.C. Cooper, and K.G. Jones

Sustainable Built Environments Research Group (SBERG),
University of Greenwich, Avery Hill Campus, Bexley Road,
Eltham, London. PostCode: SE9 2PQ. England, UK. Tel.: +44(0)7817 139170;
t_e_butt@hotmail.com

Abstract. The environment that surrounds us can be categorised into two parts, the natural environment and the built environment. Which ever environment type is to be maintained and / or enhanced for good, its sustainability has to be managed against obsolescence. From the built environment perspective, well more than half of whatever has been built and is being built, is going to be around for many decades to come. Thus, managing the sustainability of the existing built environment against obsolescence is of paramount importance. Keeping the focus on the built environment category, this paper establishes that sustainability and obsolescence are inversely related concepts. Furthermore, like sustainability, obsolescence is a multifaceted entity, thereby necessitating multimedia engagement of people/stakeholders, programs, technologies, equipment and other resources more than usual. Discussing the importance of the existing built environment, this paper presents a tool in the form of a conceptual but unified framework of sustainable obsolescence management. This framework can be used as fundamentals for further research in future to develop an intelligent multimedia architecture. Thus, the research work presented in this paper is at an initial stage of development of an intelligent multimedia system in a holistic format.

Keywords: Sustainability; sustainable development; climate change; multimedia, multiagent system; multicriteria; framework; obsolescence; built environment.

1 Introduction

The total environment that we live in can be divided into two main groups which are the natural environment and the built environment. In a mathematical equation this can be expressed as:

$$E_T = NE + BE \qquad \text{Eq. 1A}$$
$$\text{Or}$$
$$BE = E_T - NE \qquad \text{Eq. 1B}$$

Where:

E_T = Total Environment
NE = Natural Environment
BE = Built Environment

The term built environment means human-made surroundings that provide a setting for human activity, ranging in scale from personal shelter to neighbourhoods and large scale-scale civic surroundings. Thus, whatever is human-made is the built environment. The built environment consists of two main parts which are buildings and infrastructures. The built environment density in an urban environment is more than in a rural environment. The biophysical properties of the urban environment are distinctive with a large building mass ($350kg.m^{-2}$ in dense residential areas) and associated heat storage capacity, reduced greenspace cover (with its evaporative cooling and rainwater interception and infiltration functions) and extensive surface sealing (around 70% in high density settlement and city centres) which promotes rapid runoff of precipitation (Handley, 2010). Climate change amplifies this distinctive behaviour by strengthening the urban heat island (Gill et. al. 2004). So it can be safely presumed as a general rule that the greater the density of a built environment, the greater the potential of the obsolescence, irrespective of drivers and reasons. For instance, London is one of the most urbanised parts of the UK built environment in terms of a range of elements such as geographical size, value, economy, human population, diversity, ecology and heritage. Furthermore, London is the capital of the UK and located near the North Sea, stretching around an estuary, with the River Thames running through it, thereby further adding significance and sensitivity to the city in a hydrological context e.g. increased potential of pluvial, fluvial, tidal and coastal floods. In view of these wide-ranging elements together, the overall London share in the total obsolescence to take place in the total UK built environment over time, is most probably to be larger than anywhere else in the UK, and probably one of the largest shares throughout the world. (Butt et. al., 2010a; 2010b).

Any constituent (such as a building or infrastructure) of a built environment grows to become obsolete or suffers increasing obsolescence over time. Moreover, what is being built now shall predominantly be around as a substantial part of our built environment for decades to come, which are bound to suffer various degrees of obsolescence in different ways (Butt et. al., 2010a; 2010b). In order to render our built environment more sustainable, obsolescence needs to be combated. There is a host of factors which play a role either alone or collectively to cause obsolescence. Examples of these factors are not only conventional such as aging, wear and tear, etc. but also, rather contemporary factors including energy consumption efficiency, environmental pressures such as reduction of carbon / greenhouse gases emissions, legislation / regulations, change of use, clean and waste water management, water quality and resources, land use, land contamination / soil quality, air emissions, changing occupier / end user demands, waste management, ecological concerns, health & safety, and climate change (Butt et. al. 2010a; 2010b). This host of aforesaid factors form part of sustainable development or sustainability concept. As evidence it can be seen that all such factors (directly and / or indirectly to various degrees) constitute National Sustainability Strategy and Headline Indicators of Sustainable Development

in the UK as a whole, and even at state level i.e. Welsh and Scottish Headline Indicators (DEFRA, 2009; Munday and Roberts, 2006; Al Waer and Sibley, 2005; London SDC, 2005; Plows et. al., 2003; Scottish Parliament, 2002). Hence, there is a strong link between obsolescence and sustainability.

In addition to the aforesaid list of factors that cause obsolescence and are building blocks of sustainable development philosophy, a new driver which is being increasingly realised is climate change. By 2050s the UK is expected to experience: increase in average summer mean temperatures (predicted to rise by up to 3.5°C) and frequency of heat-waves / very hot days; and increases in winter precipitation (of up to 20%) and possibly more frequent severe storms (Hulme et. al., 2002). Also, in 2050s approximately 70% of UK buildings will have been built before 2010, which due to climate change factors (examples of which are indicated above) is already suffering and will further increasingly suffer from various types of obsolescence (Butt, et. al., 2010a; 2010b). Thus, if sustainable built environment is to accommodate climate change and the investment in these buildings (which was approximately £129 billions in 2007 in the UK alone (UK Status online, 2007)) is to be protected, action needs to be taken now to assess and reduce likely obsolescence of the existing UK built environment; and plan adaptation and mitigation interventions, that continue to support the quality of life and well-being of UK citizens. Failure to act now will mean that the costs of tackling climate change associated obsolescence in future will be much higher (CBI, 2007). The situation with other countries around the globe is not dissimilar, although there may be some variation in nature and quantity of climate change, and the way climate change impacts manifest themselves in relation to the resources and governance of a given country.

In order to render a given built asset more sustainable, implications of obsolescence and sustainability concepts need to be understood in relation to each other, rather than only as individually independent concepts. Sustainability or sustainable development is a multi-faceted philosophy with three main dimensions at the core it (described below in Section 2.0). Obsolescence is also multi-faceted but in opposite direction to that of sustainability. Keeping this in view, this paper outlines a conceptual unified framework of sustainable obsolescence management for built environments. There has been found no evidence of such a unified framework in the reported literature to date. Thus, the framework presented in this paper is an initial step of the research undertaken for the development of a holistic multimedia intelligent system for managing sustainability of built environments against obsolescence of any type, shape and size.

2   Definitions

2.1   Sustainability / Sustainable Development

The word sustainability is derived from the Latin (*tenere*, to hold; sus, up). Dictionaries provide a number of meanings but main ones being to 'maintain', 'support', or 'endure'. (Onions, 1964; Dictionary.com, 2010a). In built environment context, these are exactly the meanings which apply, i.e. built environment needs to

be maintained, supported, and made endurable to obsolescence against a whole host of factors that cause obsolescence (indicated in the Introduction Section above).

The Sustainability philosophy comprises three main dimensions which are Social (S) or Ethics (E), Environment (E) and Economics (E). These can be abbreviated as SEE or EEE. Although, sustainability as an individual concept, in its own right, has many definitions, meanings and implications, has been most diverse and still one of the most rapidly growing concepts; this research study is focusing on obsolescence versus sustainability only from the perspective of built environment.

2.2 Obsolescence

In English language the word 'obsolescence' means the process of becoming obsolete; falling into disuse or becoming out of date; being in the process of passing out of use or usefulness. In other words, obsolescence is the state of being which occurs when a person, object, or service is no longer wanted even though it may still be in good working order (Word Net, 2010; Butt et. al., 2010a; 2010b; Dictionary.com; 2010b; Hornby and Cowie, 1989).

In the context of built environment, obsolescence can be defined as depreciation in value and / or usefulness of a built asset (or its individual units or components) due to an impairment of desirability and / or function caused by new inventions, current changes in design, change in technology, improved process of production, change in use or end-user demands, other social factors like instability in politics of a country or tightening of environmental legislation that make a property or built asset less desirable and valuable for a continued use. (Cooper, 2004; Montgomery Law, 2010; Leeper Appraisal Services, 2010; Richmond Virginia Real Estate, 2003; Nky Condo Rentals, 2010: SMA Financing, 2009).

3   Sustainability versus Obsolescence – Built Environment Context

With reference to Sections 1.0 and 2.0, factors which cause obsolescence are in fact the ones which cause 'unsustainability'. These, factors need to be managed in order to render a given built environment sustainable. The more these factors are mismanaged, the more the obsolescence and correspondingly the less the sustainability. Thus, obsolescence and sustainability are inversely associated (further explained below with examples). This new insight is presented with a conceptual graph below. However, for a specific obsolescence factor the relationship may not necessarily be always linear it but exponential with varying degrees. However, this paper is drawing knowledge only in conceptual context, therefore, for simplicity a linear conceptual graph is selected. Some examples are described below to demonstrate that sustainability and obsolescence are opposite to each other along all the three dimensions of sustainable development philosophy i.e. social, environmental and economic.
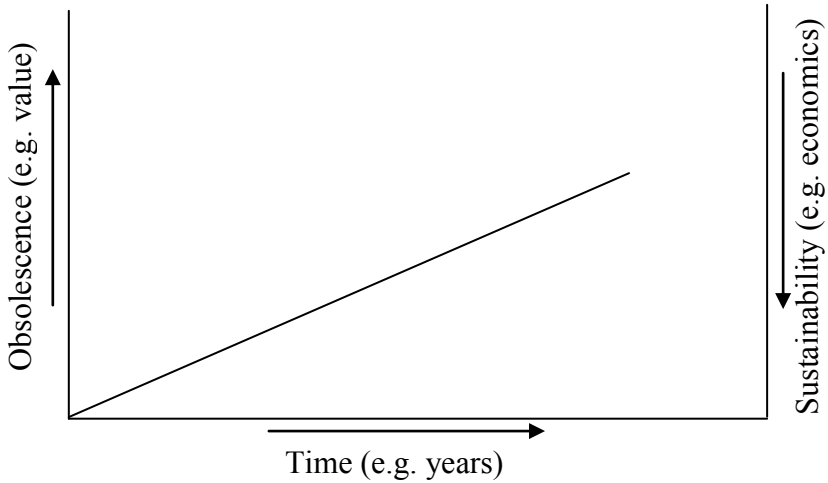
Fig. 1. A conceptual graph of inverse relationship between sustainability and obsolescence

3.1 Social

Consider a built environment scenario as follows: There is a food processing factory which also includes cooking and other processes which involves heating, thus, fuel burning activities. To dissipate emissions from burning a small chimney is used for there are no developments around. Due to lack of land availability, suppose the relevant authority approves construction application of building residential and / or commercial properties in the area. Now, as the human population increases in the area over time, the same chimney size and diffusion efficiency may not be enough for the locality. The people in the area may start objecting to chimney's emissions and ashes not being diffused in the air high enough. Thus, the current system would begin to become obsolete and correspondingly sustainability will be reducing in social / ethical context. The same situation also applies to noise pollution, if any. This example shows that social obsolescence is opposite of social sustainability of a given built environment.

3.2 Environmental

Consider a built environment setting which contains residential, commercial and light industrial developments. All of these produce domestic / municipal, commercial and industrial wastes, respectively, but there is no sustainable waste management strategy and practice in place. This means, the community is inclined to consume virgin materials and for waste materials not being reused and / or recycled, the ecological footprint of the community is even higher. Due to growing environmental protection issues, the built environment's current systems and practices regarding waste management (i.e. the current waste management infrastructure) will begin to increasingly become obsolete. This increase in degree of environmental obsolescence simply means corresponding reduction in environmental sustainability, which

indicates that the two entities, i.e. sustainability and obsolescence, are inversely linked on environmental grounds.

3.3   Economic

Consider a densely built environment in a place where water supplies are metered. Thus, water costs are in proportion to the amount consumed. Substantial amount of precipitation occurs each year in the area. Still rain water instead of being harnessed for reuse and / or recycle is simply wasted to public sewers. For water consumption, water is used only from the main supply. Due to inflation and increase in water rates, over time, this method of water consumption may become less and less sustainable on economic grounds. This, would also mean that economic obsolescence will begin to be induced in the current (unsustainable) water management system and escalate over time. This is how sustainability and obsolescence are inversely related in economic context of sustainable development philosophy.

4   Holistic Sustainable Obsolescence Management

In this section of the paper, a unified framework of a holistic intelligent system is developed in the form of a tool which interacts between the three fundamental dimensions of sustainability (i.e. social, economic and environment), where people, programs, technologies, equipment and other resources can be engaged in a multimedia environment. The framework is currently at conceptual stage of development in the research project undertaken at the University of Greenwich and depicted in Figure 2. The framework provides a foundation which takes account of various aspects of the three principal dimensions of sustainability particularly at the cost-benefit analysis stage, where all such aspects interact to assist evidence-based decision making thereby drawing a best compromise between various aspects of social, environmental and economic dimensions in order to manage obsolescence sustainably.

   Although the whole framework has multifaceted features and multimedia engagement from various stakeholders, but particularly at the cost-benefit analysis stage of the framework this becomes very evident where inputs from diverse technical to non-technical stakeholders are put together to analyse as a multimedia interaction to yield a most sustainable solution. However, since computing technology has enhanced traditional information technologies many folds, therefore as a future research potential, once this framework is converted into a detailed computer-aided intelligent system, then particularly the CBA stage will provide a computer-aided multimedia architecture where the framework would be more readily useable by built asset managers and other relevant decision makers both in maintenance as well as refurbishment.

   All the items in the framework are shadowed under technical, non-technical, physical and / or non-physical aspects. The contributing proportions of these four aspects for items in the framework will vary from scenario to scenario of built environment, depending upon a number of characteristics such as nature, size, scope, and type of the component under consideration. Examples of technical aspects are

heating systems; limitations of technologies (such as non-energy saver i.e. conventional lighting bulbs); etc. Whereas, the behaviour of occupants and maintenance staff of a commercial building; energy using patterns of dwellers of a house; etc. are examples of non-technical aspects. Examples of physical aspects are fabric of buildings; facade of buildings; furniture; etc. Whereas for non-physical the following can be taken as examples: energy; ventilation; maintenance or refurbishment schedule; etc. Among various items in the holistic structure (Figure 2), some would be physical, some non-physical, some technical, some non-technical, and some could be various combinations of any four of these aspects. This will depend on various characteristics specific to a given built environment scenario under consideration. For instance, natural lighting is non-physical but may have aspects that are associated with physical design of the building or non-physical phenomenon such as summer sun or winter sun. Similarly, environmental legislation (e.g. the Climate Change Act 2008; 2010 (DECC, 2010a; 2010b; HM, 2010)) regarding carbon cuts is a non-technical entity but may have technical aspects integrated when carbon cut technologies (e.g. carbon capture and storage) are employed to a fossil-fuel power generation plant. Also, the heating system of a building is a physical commodity, but energy consumption efficiency in terms of users' behaviour is non-physical aspect and the employed heating technology is a technical matter. Management systems (such as maintenance schedule; environmental management system e.g. ISO 14000; quality management system e.g. ISO 9000, etc.) are non-physical matters but may have technical as well as non-technical aspects associated with them.

Obsolescence management can be described as a process of analysis, evaluation, control and organisation of obsolescence for a given built environment scenario. Figure 2 depicts the conceptual framework as a basis of a holistic multimedia system of sustainable obsolescence management, where people/stakeholders, technologies, maintenance and refurbishment programs, equipment and other resources can sequentially and categorically interact to help control obsolescence of existing built assets. The framework consists of two main parts i.e. Obsolescence Assessment (OA) and Obsolescence Reduction (OR). The output of OA is input to OR and this is how the former provides basis for the latter. Therefore, the more robust and stronger the foundation yielded from OA, the more effective the OR is likely to be.

4.1 Obsolescence Assessment (OA)

OA consists of two sub-parts i.e. baseline study and, Identification and categorisation. In the baseline study section of the framework an obsolescence assessor is to gather relevant information by carrying out a desk study of various documents such as drawings, engineering design, historic data on various aspects e.g. flooding, etc. This information can be reinforced by paying investigative visits to the site and gathering anecdotes. The baseline study can also play a vital role of screening before scoping is carried out in a later stage of the framework. Based on the information collected, the assessor can identify and categorise various items under a set of two headings i.e. built environment constituents (BEC) and obsolescence nature (Figure 2).
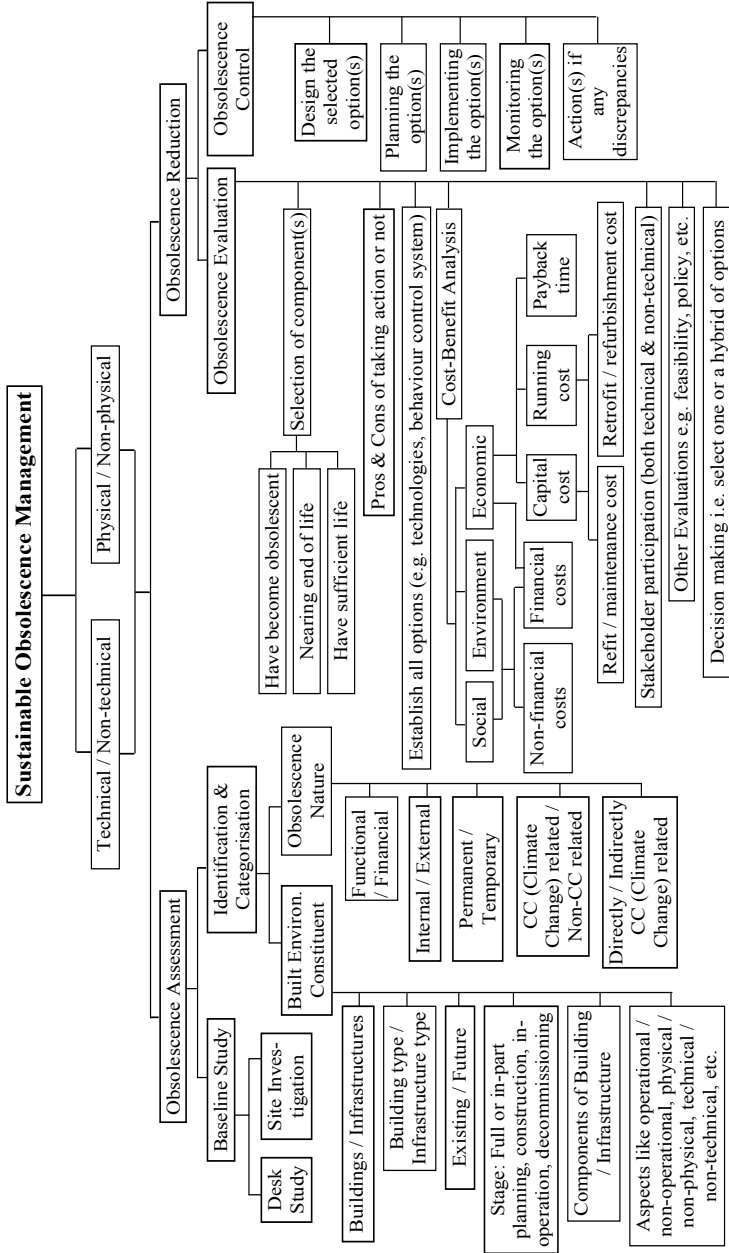
Fig. 2. Holistic conceptual framework of sustainable obsolescence assessment and management for the built environment

4.1.1 Built Environment Constituents (BEC)

In the BEC module, scope or boundaries can be established of a given built environment. For instance, whether it is a building or infrastructure; what is the type of building (e.g. commercial, domestic, industrial, etc.); what is the type of infrastructure e.g. transport (and if transport then road network or buses in specific; railway networks or trains themselves); energy generation or distribution network; etc. At this point, it can also be established whether it is an existing or a future development. The stage of a development can also be further broken down along the following steps: planning, construction, in-operation and / or decommissioning. Sometimes it may be the extension of an already existing building which is being planned, constructed or decommissioned. Therefore, it is better to identify whether planning, construction, in-operation and / or decommissioning are in full or in part. After this, the assessor can break down the given built environment under consideration into all constituting components. These components can then be characterised along facets such as operational, non-operational, physical, non-physical, technical, non-technical, socio-technical, managerial, non-managerial, fabric, non-fabric, etc.

4.1.2 Obsolescence Nature

The next main stage in the framework is identification and categorisation of obsolescence nature for the specified components (Figure 2). The nature of the obsolescence can be characterised as follows: *Financial obsolescence* means loss in value where as *functional obsolescence* is loss of usefulness, effectiveness, efficiency or productivity. The financial obsolescence is also termed as *social or economic obsolescence*, and functional obsolescence as *technical obsolescence*. (Cooper, 2004; Montgomery Law, 2010; Leeper Appraisal Services, 2010; Richmond Virginia Real Estate, 2003; Nky Condo Rentals, 2010: SMA Financing, 2009). Irrespective of whether obsolescence is in value or function or both, *internal obsolescence* in a building component or built asset is due to factors that exist within the component or built asset. For instance, general wear and tear, fatigue, corrosion, oxidation, evaporation, rusting, leaking of gas / water or any other fluid like coolant, breaking, age, etc. Where as *external obsolescence* is temporary or permanent impairment in value or usefulness of a built asset due to factors outside the system such as change in existing or advent of a new environmental legislation; social forces / pressure groups; arrival of new technology; improvement or enhancement of knowledge; fluctuation in demand; inflation of currency; etc. (Landmark Properties, 2009; Salt Lake County, 2004; ESD Appraisal Services, 2010; Drew Mortgage, 2006). *Permanent obsolescence* is irreversible, for instance materials in buildings that contain asbestos have become permanently obsolete due to its adverse health impacts. On the contrary, factors which are not permanent such as temporary civil unrest in a society, loss of power for days, flooding, etc. can cause a *temporary obsolescence*.

Similarly, Irrespective of whether an obsolescence is internal or external and financial or functional, if a given obsolescence is due to impacts of climate change (e.g. more intense and more frequent rainfall, stronger and more frequent hurricanes, heat-wave, flooding, etc.) is referred to as *climate change induced obsolescence* by the authors (Butt et. al., 2010a; 2010b). If obsolescence is independent of climate change impacts, then it is called *non-climate change related obsolescence*. However,

if obsolescence is climate change related, then it could be either directly climate change related or indirectly climate change related. Obsolescence that is caused by direct impact of climate change factors is termed as *directly induced climate change obsolescence*. For instance, current air conditioning systems in our built environment may not be as effective due to global warming / heat-waves which are a resultant of climate change. Thus global warming / heat-waves may bring about obsolescence in a given building's air conditioning system as a direct impact. These heat-waves can also have direct adverse affects on the structure or fabric of buildings. Obsolescence that results from the impact of climate change factors in an indirect manner is referred to as *indirectly induced climate change obsolescence*. For example, irrespective of the degree, one of the reasons of climate change acceleration is anthropogenic activities such as greenhouse gas (GHG) emissions which include carbon dioxide. This has contributed in shaping environmental legislation such as European Union (EU) Directive of Energy Performance of Buildings (2002/91/EC) (EU, 2002; EC, 2010); EU Climate and Energy objectives; and legally binding carbon reduction targets set up by Climate Change Act 2008 (DECC, 2010a; 2010b). Such environmental legislations have begun to cause indirectly induced climate change obsolescence in existing buildings.

4.2   Obsolescence Reduction (OR)

As stated earlier, the second main part of the sustainable obsolescence management is OR. Although the whole of the obsolescence management framework may be iterated a number of times depending on various characteristics of a given built environment scenario, this part of the holistic framework certainly needs repeating a number of times until most sustainable and yet realistically possible solution or set of solutions have been derived and implemented. The Obsolescence Assessment (OA) is predominantly around gathering and categorisation of data and information of the given built environment, which does not need much iteration. However, as for Or, the main reason for repeating the OR more frequently is that various modules in this part are mutually dependent on each other for mutual information transfer. This is due to the fact that information processed in various OR modules have to be delivered and received between the modules backwards and forwards a number of times. For instance, the information between cost-benefit analysis and stakeholder participation modules has to be used backwards and forwards due to various sustainability aspects as well as variation in interests among different stake holders (Figure 2). This iteration aspect becomes clearer in the discussion below where various modules of the OR part are described in more details. This part has been divided into two sub-parts i.e. Obsolescence Evaluation (OE) and Obsolescence Control (OC). Details on them are described below:

4.2.1   Obsolescence Evaluation (OE)

The first unit in the OE section is 'selection of component(s)' module. Based on the information which would have been collated earlier in the OA part of the OM, in this module the components of the built environment scenario, which are the point of interest in terms of obsolescence assessment and management, can be identified and categorised. In order to assist prioritisation and selection of the components, this

module categorises the identified components into three groups based on the rationale around various sustainability aspects. These three groups are:

1. The components which have become obsolescent;
2. The components which are nearing end of life; and
3. The components which have sufficient life.

There is a module allocated for establishing positive and negative impacts of both taking action and not taking action to fix the obsolescence problems, particularly those of from the first two groups above. This can help to further prioritise on which components need more and quick attention as opposed to others. Following this, there is another module in the framework where all possible options to control obsolescence can be identified and their characteristics (both advantages and limitations) can be established. These options could be technical (e.g. some innovative technology); non-technical (such as a new managerial system to control behaviour of energy consumption); or even combinations of technical and non-technical facets with varying proportions. This information on various options can later also feed into the 'cost-benefit analysis' module, which is divided into three sub-modules to address the three principal dimensions of sustainable development philosophy. These three sub-modules are: Social, Environment and Economic. Each of the social and environment sub-modules cover sustainability aspects in two categories, which are financial costs and non-financial costs. For the social sub-module, examples of financial costs are fine that a company may face due to not complying with some legislation requirements such as health and safety regulations; compensation which might have to be paid to the relevant stakeholder e.g. an employee who suffers a health problem or an accident at work; compensation might have to be paid to an external customer too; etc. Where as adverse impact on company image, quality of service or product of the company, poor social corporate responsibility, are examples of non-financial aspects. Similarly for the environment sub-module, lets consider a case in which some spillage of a toxic substance takes place due to some improper or obsolete component. This can cause financial costs such as the cost to fix the environmental damage and compensation to the victims of the environmental damage. Whereas the bad publicity and extra-ordinarily high pressures from the government bodies (e.g. the Environment Agency) as well as various voluntary environmental pressure groups are examples of non-financial environmental costs . For the economic sub-module there are three categories which are:

1. capital cost,
2. running cost, and
3. payback time.

In the first two categories above, costs of refitting (i.e. maintenance) and / or retrofitting (i.e. refurbishment) of the selected components are to be analysed. The payback time will also play a vital role in decision making. The financial costs of environmental and social aspects can also be tapped into the economics sub-module to draw a bigger picture of total costs. Thus, economic sub-module is shown connected

with financial costs of the social and environmental sub-modules (Figure 1). The cost-benefit analysis can be reinforced by consulting diverse spectrum of (internal and external) stakeholders ranging from technical to non-technical. Each and every stakeholder needs not to be consulted for each and every obsolescence scenario but only appropriate and relevant stakeholders depending on characteristics of the scenario. Thus, in the 'stakeholder participation' module of the framework, appropriate stakeholders should also be identified prior to consultation. Information from the 'other evaluations' module regarding e.g. feasibility report, the company's policy and mission statement, etc., can also be tapped into the cost-benefit analysis module to render it more holistic. Eventually, in the decision making module, a decision is made in terms selection of an option or a set of options to reduce impacts of the obsolescence in the built environment scenario.

4.2.2 Obsolescence Control (OC)

In the OC section of the sustainable obsolescence management framework, the selected obsolescence control option(s) is/are designed and planned. If any unexpected implications, these can be reconsulted with the appropriate stakeholders and rechecked via the cost-benefit analysis module. If any problems, another option or set of options can be selected and then designed and planned again. Such iterations can continue till a satisfactory option or set of options has/have been designed and planned, following which the option(s) can be implemented. While implementing monitoring needs to take place for if there are any discrepancies. Frequency of monitoring can also be incorporated into the system at the design and planning stages earlier. If any discrepancies observed, corrective actions need to be taken to control the implementation process. Such corrective actions against discrepancies can also be set during the design and planning stage.

5   Concluding Remarks

This paper establishes link between sustainability and obsolescence. The elements which are associated with sustainability, the same elements are associated with obsolescence but in the opposite direction. Based on this fact, the paper establishes that like sustainability, obsolescence is also multifaceted and sustainable management of obsolescence means an equally holistic multimedia architecture is needed where all various resources from people / stakeholders to information and technologies have a multimedia interaction along the three fundamental dimensions of sustainability philosophy i.e. social, economic and environmental aspects. This paper presents such a holistic framework which forms fundamentals of an intelligent multimedia interaction model, where all sustainability aspects are taken into account particularly at the cost-benefit analysis stage to create an environment for evidence-based decision-making.

The multimedia conceptual framework is developed in a holistic fashion for the sustainable obsolescence management to cover all appropriate modules and sub-modules from the start to the end. Obsolescence types and obsolescence drivers also identified and categorised to ease interaction of various modules and sub-modules of the framework. The framework is specific to the built environment and is able to deal with built environment scenarios whether fully or partly at planning, construction, in-

operation and / or decommissioning stage. The physical, non-physical, technical and non-technical aspects are also included. This renders the framework useful for diverse range of stakeholders from experts and technical to non-experts and non-technical, respectively. This research work is a step towards making obsolescence management possible in a holistic and sustainable manner. This research work can also streamline current practices of obsolescence management which are available in a non-integrated and peace-meal fashion, and to a limited extent. This framework can attract debate and interests from both practitioners and researchers for further study and research, and later be converted into a computer-aided intelligent multimedia system. However, the framework, in its current shape can still be effectively used as a decision making tool to select an option or set of options to assess and manage obsolescence in a sustainable manner based on cost-benefit analysis. Thereby, assisting in rendering our existing built environment (which will be around for many decades to come) more sustainable against various obsolescence drivers from conventional to as modern factors as climate change.

References

1. Al Waer, H., Sibley, M.: Sustainability Indicators: Complexity of measurements and new trends. In: 11th Annual International Sustainable Development Research Conference, Helsinki, Finland, pp. 6–8 (June 2005)
2. Butt, T.E., Giddings, B., Cooper, J.C., Umeadi, B.B.N., Jones, K.G.: Advent of Climate Change and Energy Related Obsolescence in the Built Environment. In: International Conference on Sustainability in Energy and Buildings, Brighton, UK, May 6-7 (2010a)
3. Butt, T.E., Umeadi, B.B.N., Jones, K.G.: Sustainable Development and Climate Change Induced Obsolescence in the Built Environment. In: International Sustainable Development Research Conference, Hong Kong, China, May 30 – June 1 (2010b)
4. CBI (Confederation of British Industry), Climate Change: Everyone's business, CBI (2007)
5. Cooper, T.: Inadequate Life? Evidence of Consumer Attitudes to Product Obosolescence. Journal of Consumer Policy 27, 421–449 (2004)
6. DECC (Department of Energy and Climate Change), Climate Change Act (2008), `http://www.decc.gov.uk/en/content/cms/legislation/cc_act_08/cc_act_08.aspx` (Viewed April 2010b)
7. DECC (Department of Energy and Climate Change), Legislation, `http://www.decc.gov.uk/en/content/cms/legislation/legislation.aspx` (Viewed April 2010a)
8. DEFRA (Department for Environment, Food and Rural Affairs), Sustainable development indicators in your pocket 2009, An update of the UK Government Strategy indicators, Crown Copyright (2009)
9. Dictionary.com, `http://dictionary.rererence.com/browse/obsolescence` (viewed June 2010b)
10. Dictionary.com, `http://dictionary.rererence.com/browse/sustain` (viewed June 2010a)
11. EC [European Commission – Energy (ManagEnergy)], COM 2002/91/EC: Directive on the Energy Performance of Buildings, `http://www.managenergy.net/products/R210.htm` ManagEnergy (last modified March 09, 2010)

12. EU (European Union), Directive 2009/91/EC of the European Parliament and of the Council of 16 December 2002 on the energy performance of buildings. OJ L1/65 (4-1-2003) (2002)
13. Gill, S., Pauleit, S., Ennos, A.R., Lindley, S.J., Handley, J.F., Gwilliam, J., Ueberjahn-Tritta, A.: Literature review: Impacts of climate change on urban environment. CURE, University of Manchester (available on line) (2004)
14. Handley, John F.: Adaptation strategies for climate change in the urban environment (ASCCUE), Narrative report for GR/S19233/01, `http://www.sed.manchester.ac.uk/research/cure/downloads/ asccue-epsrc-report.pdf` (viewed March 2010)
15. HM Government, Climate Change: Taking Action (Delivering the Low Carbon Transition Plan and preparing for changing climate), Crown Copyright (2010)
16. Hornby, A.S., Cowie, A.P.: Oxford Advanced Learner's Dictionary of Current English. Oxford University Press, Oxford (1989)
17. Hulme et. al. Climate change scenarios for the United Kingdom: The UKCIP 2002 Scientific Report, Tyndall Centre for Climate Change Research, School of Environmental Sciences, University of East Anglia, p. 120 (2002)
18. Leeper Appraisal Services, California Appraisal / Appraisers, `http://www.leeperappraisal.com/appraiser_jargon.htm` (viewed March 2010)
19. London SDC (London Sustainable Development Commission), London Sustainable Development Commission, 2005 report on London's Quality of Life indicators, Greater London Authority (May 2005)
20. Montgomery Law, Family Law Glossary, `http://www.montylaw.com/family-law-glossaryo.php` (viewed March 2010)
21. Max, M., Annette, R.: Developing Approaches to Measuring and Monitoring Sustainable Development in Wales: A Review. Regional Studies 40(5), 535–554 (2006)
22. Nky Condo Rentals – Rodney Gillum (2010), `http://www.nkycondorentals.com/index.cfm/fuseaction/ terms.list/letter/O/contentid/511AF257-5436-4595-81BBAA7704C1AC40` (viewed March 2010)
23. Onions, Charles, T. (eds.): The Shorter Oxford English Dictionary, p. 2095. Clarendon Press, Oxford (1964)
24. Plows, A.J., Jones, J.M., Foley, S.A., Butt, T.E., Pomeroy, I.R.: Awareness of Best Practice for Sustainable Development: The feasibility and use of a broad-based sustainability tool by organisations in Wales. In: The 2003 Int. Sustainable Development Research Conference, Nottingham, UK, March 24 - 25, pp. 353–361 (2003)
25. Richmond Virginia Real Estate, Real Estate Dictionary (2003), `http://www.therichmondsite.com/Blogs/Dictionary_O.html` (viewed March 2010)
26. Scottish Parliament (The Information Centre – SPCI), Sustainable Development (Updated), SPCI Briefing 02/47, Scottish Parliament (May 7, 2002)
27. SMA Financing, Real Estate Glossary (2009), `http://www.smafinancing.com/glossary.htm` (viewed March 2010)
28. UK Status online, Gross Fixed Capital Formation at Chained Volume Measure (2007)
29. Word Net - A Lexical Database for English, `http://www.wordnet.princeton.edu` (viewed March 2010)

Automatic Text Formatting for Social Media Based on Linefeed and Comma Insertion

Masaki Murata[1], Tomohiro Ohno[2], and Shigeki Matsubara[1]

[1] Graduate School of Information Science, Nagoya University,
Furo-cho, Chikusa-ku, 464-8601, Japan
murata@el.itc.nagoya-u.ac.jp, matubara@nagoya-u.jp
[2] Graduate School of International Development, Nagoya University,
Furo-cho, Chikusa-ku, 464-8601, Japan
ohno@nagoya-u.jp

Abstract. By appearance of social media, people are coming to be able to transmit information easily on a personal level. However, because users of social media generally spend little time on describing information, low-quality texts are transmitted and it blocks the spread of information. On transmitted texts in social media, commas and linefeeds are inserted incorrectly, and it becomes a factor of low-quality texts. This paper proposes a method for automatically formatting Japanese texts in social media. Our method formats texts by inserting commas and linefeeds appropriately. In our method, the positions where commas and linefeeds should be inserted are decided based on machine learning using morphological information, dependency relation and clause boundary information. An experiment using Japanese spoken language texts has shown the effectiveness of our method.

1 Introduction

Recently, social media such as Facebook, Twitter and blog service are appearing. Facebook has over five hundred million users, and Twitter has over one hundred million users. By using social media, those users can spread information easily to people all over the world through interaction on a personal level. However, users generally spend little time on describing information in social media because they want easy transmission of information, unlike those in newspaper or TV broadcast. Therefore, the transmitted texts tend to be low-quality. Low-quality texts might prevent information from spreading in social media.

This paper proposes a method for automatically formatting Japanese texts in social media. Our method formats texts by inserting commas and linefeeds at proper positions. On transmitted texts in social media, too many or too few commas are often inserted or proper linefeed insertion may not be done, and it becomes a factor of low-quality texts. However, commas and linefeeds play important roles in the readability of texts.

In our method, first, commas are inserted into a Japanese text, and then, linefeeds are inserted into the text into which commas were inserted. The comma insertion method

decides whether or not to insert a comma at each *bunsetsu*[1] boundary by machine learning. The linefeed insertion is also executed by the similar technique.

We conducted an experiment on comma and linefeed insertion using Japanese spoken language data and confirmed the effectiveness of our method.

2 Text Formatting by Comma and Linefeed Insertion

In social media, while users can transmit information easily, users spend little time on describing information. Therefore, the transmitted texts tend to be low-quality. Automatic text formatting is desired so that texts become easy to read.

As related works, there have been researches on automatic text summarization [1], conversion of spoken language into written language [2] and conversion of spoken documents into web documents [3]. In these works, text formatting is realized by editing strings in the texts. However, it is important not only to be written in simple words or simple structure, but also to be displayed properly so that texts are easy to read.

In our work, we focus on comma and linefeed insertion as the method for text formatting. For Japanese, which is a non-segmented language, it is necessary to segment texts properly by commas and linefeeds so that the texts are easy to read.

2.1 Text Formatting by Comma Insertion

There are several usages of commas such as making clear sentence structures or marking word boundaries, and the positions where commas are inserted depend on these usages. It is important to insert commas at proper positions in accordance with the usages of commas. However, while there is a tendency about the positions where commas should be inserted, there is no clear standard for these positions. Therefore, positions where commas are inserted could be different in individuals. In social media, texts into which commas were inserted incorrectly are created. Because such texts become hard to read, it is required to format these texts by inserting commas at proper positions.

Newspaper articles are one of the texts into which commas were inserted properly. Examples of a text in newspaper articles are shown below:

- 世界の成長センターとなったアジアで、急浮上する中国の存在は、希望にあ ふれていると同時に、困難な課題も山積している。(In Asia, which becomes a center of the growth of the world, the strong presence of China not only brings hope but also causes difficult problems.)
- むしろ地球規模の環境、人口、食糧など広範に国連の果たさなければなら ない役割は大きい。 (The United Nations should play a lot of roles in a broad range of fields, such as the global environment, population, and food.)

It is possible to improve the readability of web texts by inserting commas at positions where commas are inserted in newspaper articles.

---

[1] *Bunsetsu* is a linguistic unit in Japanese that roughly corresponds to a basic phrase in English. A bunsetsu consists of one independent word and zero or more ancillary words.

2.2   Text Formatting by Linefeed Insertion

In order to generate readable texts, it is important how to display texts. In the text in Figure 1, how to display is not considered, and linefeeds are forcibly inserted according to the width. Therefore, the text is not easy to read. In fact, such the texts are written in newspaper articles because newspaper have space constraints. On the other hand, the space constraints are not so hard on the web. Therefore, to display texts into which linefeeds are inserted appropriately is effective.

By inserting linefeeds at semantic boundaries, the texts become easy to read. Figure 2 shows the text into which linefeeds are inserted at proper positions. Linefeeds are inserted at the semantic boundary 関係があるんですけれども (are connected)" and right after the subject of the sentence 状態というのは (the situation)".
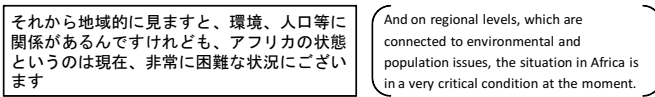
それから地域的に見ますと、環境、人口等に
関係があるんですけれども、アフリカの状態
というのは現在、非常に困難な状況にござい
ます

(And on regional levels, which are connected to environmental and population issues, the situation in Africa is in a very critical condition at the moment.)

Fig. 1. Text into which linefeeds are inserted forcibly

それから地域的に見ますと、
環境、人口等に関係があるんですけれども、
アフリカの状態というのは
現在、非常に困難な状況にございます

(And on regional levels,)
which are connected to environmental and population issues,
(the situation in Africa)
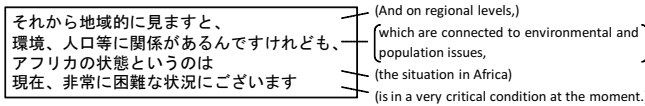(is in a very critical condition at the moment.)

Fig. 2. Text into which linefeeds are inserted properly

3   Text Formatting Method

Our method formats texts by inserting commas and linefeeds appropriately. In social media, texts into which commas and linefeeds are inserted incorrectly can be created. Therefore, our method ignores these inserted commas and linefeeds, and inserts commas and linefeeds into texts which have no commas and linefeeds.

In our method, we use the comma insertion method [4] and the linefeed insertion method [5]. Figure 3 shows an example of text formatting by our method. First, our method inserts commas into a Japanese sentence, and then, inserts linefeeds so that the number of characters in each line can be less than or equal to the maximum number of characters per line.

3.1   Comma Insertion Method

In comma insertion method, a sentence, on which morphological analysis, bunsetsu segmentation, clause boundary analysis and dependency[2] analysis have been performed, is

---

[2] A dependency in Japanese is a modification relation in which a modifier bunsetsu depends on a modified bunsetsu. That is, the modifier bunsetsu and the modified bunsetsu work as modifier and modifyee, respectively.
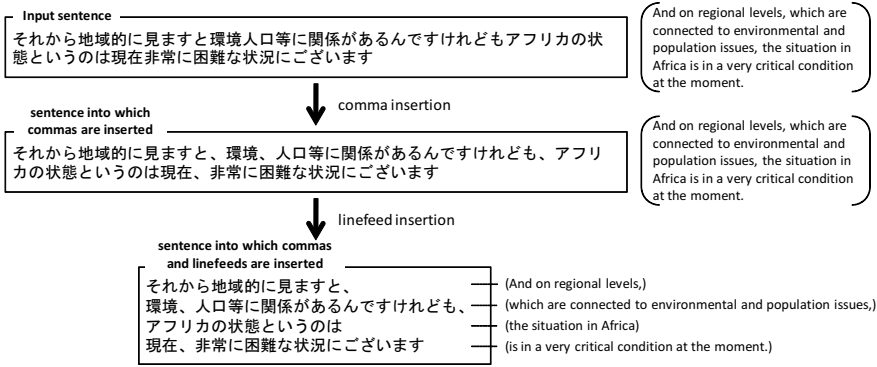
Input sentence
それから地域的に見ますと環境人口等に関係があるんですけれどもアフリカの状態というのは現在非常に困難な状況にございます

(And on regional levels, which are connected to environmental and population issues, the situation in Africa is in a very critical condition at the moment.)

comma insertion

sentence into which commas are inserted
それから地域的に見ますと、環境、人口等に関係があるんですけれども、アフリカの状態というのは現在、非常に困難な状況にございます

(And on regional levels, which are connected to environmental and population issues, the situation in Africa is in a very critical condition at the moment.)

linefeed insertion

sentence into which commas and linefeeds are inserted
それから地域的に見ますと、   (And on regional levels,)
環境、人口等に関係があるんですけれども、 (which are connected to environmental and population issues,)
アフリカの状態というのは   (the situation in Africa)
現在、非常に困難な状況にございます   (is in a very critical condition at the moment.)

Fig. 3. Example of text formatting by our method

considered the input. Our method decides whether or not to insert a comma at each bunsetsu boundary in an input sentence. Comma insertion method identifies the most appropriate combination among all combinations of positions where a comma can be inserted, by using the probabilistic model.

In this paper, input sentences which consist of n bunsetsus are represented by $B = b_1 \cdots b_n$, and the results of comma insertion by $R = r_1 \cdots r_n$. Here, r_i is 1 if a comma is inserted right after bunsetsu b_i, and 0 otherwise. Also, $r_n = 1$. We indicate the j-th sequence of bunsetsus created by dividing an input sentence into m sequences as $L_j = b_1^j \cdots b_{n_j}^j (1 \leq j \leq m)$, and then, $r_k^j = 0$ if $1 \leq k < n_j$, and $r_k^j = 1$ if $k = n_j$.

3.1.1 Probabilistic Model for Comma Insertion

When an input sentence B is provided, our method identifies the comma insertion R that maximizes the conditional probability $P(R|B)$. Assuming that whether or not to insert a comma right after a bunsetsu is independent of other commas except the one appearing immediately before that bunsetsu, $P(R|B)$ can be calculated as follows:

$$
\begin{aligned}
P(R|B) \quad &(1)\\
= {}& P(r_1^1 = 0, \cdots, r_{n_1-1}^1 = 0, r_{n_1}^1 = 1, \cdots, r_1^m = 0, \cdots, r_{n_m-1}^m = 0, r_{n_m}^m = 1|B)\\
\cong {}& P(r_1^1 = 0|B) \times \cdots \times P(r_{n_1-1}^1 = 0|r_{n_1-2}^1 = 0, \cdots, r_1^1 = 0, B)\\
& \times P(r_{n_1}^1 = 1|r_{n_1-1}^1 = 0, \cdots, r_1^1 = 0, B) \times \cdots\\
& \times P(r_1^m = 0|r_{n_{m-1}}^{m-1} = 1, B) \times \cdots \times P(r_{n_m-1}^m = 0|r_{n_m-2}^m = 0, \cdots, r_1^m = 0, r_{n_{m-1}}^{m-1} = 1, B)\\
& \times P(r_{n_m}^m = 1|r_{n_m-1}^m = 0, \cdots, r_1^m = 0, r_{n_{m-1}}^{m-1} = 1, B)
\end{aligned}
$$

where $P(r_k^j = 1|r_{k-1}^j = 0, \cdots, r_1^j = 0, r_{n_{j-1}}^{j-1} = 1, B)$ is the probability that a comma is inserted right after a bunsetsu b_k^j when the sequence of bunsetsus B is provided and the position of j-th comma is identified. Similarly, $P(r_k^j = 0|r_{k-1}^j = 0, \cdots, r_1^j = 0, r_{n_{j-1}}^{j-1} = 1, B)$ is the probability that a comma is not inserted right after a bunsetsu b_k^j. These probabilities are estimated by the maximum entropy method. The result R which maximizes

the conditional probability $P(R|B)$ is regarded as the most appropriate result of comma insertion, and calculated by dynamic programming.

Table 1. Features used for the maximum entropy method (comma insertion)

morphological information	the rightmost independent morpheme, i.e. head word, (part-of-speech and inflected form) and rightmost morpheme (part-of-speech) of a bunsetsu b_k^j
	the rightmost morpheme (a surface form) of b_k^j if the rightmost morpheme is a particle
	the first morpheme (part-of-speech) of b_{k+1}^j
commas inserted between clauses	whether or not a clause boundary exists right after b_k^j
	type of a clause boundary right after b_k^j if there exists a clause boundary
commas indicating clear dependency relations	whether or not b_k^j depends on the next bunsetsu
	whether or not b_k^j depends on a bunsetsu located after the final bunsetsu of the clause including the next bunsetsu of b_k^j
	whether or not b_k^j is depended on by the bunsetsu located right before it
	whether or not the dependency structure of a sequence of bunsetsus between b_k^j and b_1^j is closed
commas avoiding reading mistakes and reading difficulty	whether or not both the rightmost morpheme of b_k^j and first morpheme of b_{k+1}^j are *kanji* characters
	whether or not both the rightmost morpheme of b_k^j and first morpheme of b_{k+1}^j are *katakana* characters
commas indicating the subject	whether or not there exists a clause boundary "topicalized element-*wa*" right after b_k^j and b_k^j depends on the next bunsetsu
	whether or not there exists a clause boundary "topicalized element-*wa*" right after b_k^j and the string of characters right before b_k^j is "では (*dewa*)"
	the number of characters in a phrase indicating the subject[3] if there exists a clause boundary "topicalized element-*wa*" right after b_k^j
	whether or not a clause boundary "topicalized element-*wa*" exists right after b_k^j and a bunsetsu whose rightmost morpheme is a verb depends on the modified bunsetsu of b_k^j
commas inserted after a conjunction or adverb at the beginning of a sentence	whether or not b_k^j appears at the beginning of a sentence and its rightmost morpheme is a conjunction
	whether or not b_k^j appears at the beginning of a sentence and its rightmost morpheme is an adverb
commas inserted between parallel words or phrases	whether or not both the rightmost morphemes of b_k^j and b_{k+1}^j are nouns
	whether or not a predicate at the sentence end is depended on by b_k^j whose rightmost independent morpheme is a verb and by any of the bunsetsus which are located after b_k^j and of which the rightmost independent morpheme is a verb
number of characters from b_1^j to b_k^j	one of the following 4 categories if the number of characters from b_1^j to b_k^j is found there ([num = 1], [2 ≤ num ≤ 3], [4 ≤ num ≤ 21], [22 ≤ num])

---

[3] Phrases indicating the subject is a sequence of bunsetsus consisting of b_k^j and all bunsetsus that are connected to b_k^j when we trace their dependency relationship in modifier-to-modifyee direction.

3.1.2 Features on Maximum Entropy Method

To estimate $P(r_k^j = 1 | r_{k-1}^j = 0, \cdots, r_1^j = 0, r_{n_{j-1}}^{j-1} = 1, B)$ and $P(r_k^j = 0 | r_{k-1}^j = 0, \cdots, r_1^j = 0, r_{n_{j-1}}^{j-1} = 1, B)$ by the maximum entropy method, we used the features in Table 1. To decide the features, we grouped the usages of commas, and investigated the appearance tendency for each category by using Japanese newspaper articles. For more details, please refer to the literature [4].

3.2 Linefeed Insertion Method

We adopt the same method as comma insertion for linefeed insertion. A sentence, on which morphological analysis, bunsetsu segmentation, clause boundary analysis and dependency analysis have been performed, is considered the input. Linefeed insertion method decides whether or not to insert a linefeed at each bunsetsu boundary in an input sentence. Under the construction that the number of characters in each line has to be less than or equal to the maximum number of characters per line, linefeed insertion method identifies the most appropriate combination among all combinations of

Table 2. Features used for the maximum entropy method (linefeed insertion)

Morphological information	the rightmost independent morpheme i.e. head word, (part-of-speech and inflected form) and rightmost morpheme (part-of-speech) of a bunsetsu b_k^j
Clause boundary information	whether or not a clause boundary exists right after b_k^j
	type of a clause boundary right after b_k^j if there exists a clause boundary
Dependency information	whether or not b_k^j depends on the next bunsetsu
	whether or not b_k^j depends on the final bunsetsu of a clause
	whether or not b_k^j depends on a bunsetsu to which the number of characters from the start of the line is less than or equal to the maximum number of characters
	whether or not b_k^j is depended on by the final bunsetsu of an adnominal clause
	whether or not b_k^j is depended on by the bunsetsu located right before it
	whether or not the dependency structure of a sequence of bunsetsus between b_k^j and b_1^j, which is the first bunsetsu of the line, is closed
	whether or not there exists a bunsetsu which depends on the modified bunsetsu of b_k^j, among bunsetsus which are located after b_k^j and to which the number of characters from the start of the line is less than or equal to the maximum number of characters
Line length	one of the following 3 categories if the number of characters from the start of the line to b_k^j is found there ([num $\leq$ 2], [3 $\leq$ num $\leq$ 6], [7 $\leq$ num])
Pause	whether or not a pause exists right after b_k^j
Leftmost morpheme of a bunsetsu	whether or not the basic form or part-of-speech of the leftmost morpheme of the next bunsetsu of b_k^j is one of the following morphemes (Basic form: "思う (think)", "問題 (problem)," "する (do)," "なる (become)," "必要 (necessary)" Part-of-speech: noun-non independent-general, noun-nai adjective stem, noun-non independent-adverbial)
Comma	whether or not a comma exists right after b_k^j

the positions into which linefeeds can be inserted, by using the probabilistic model. To estimate $P(r_k^j = 1 | r_{k-1}^j = 0, \cdots, r_1^j = 0, r_{n_{j-1}}^{j-1} = 1, M)$ and $P(r_k^j = 0 | r_{k-1}^j = 0, \cdots, r_1^j = 0, r_{n_{j-1}}^{j-1} = 1, M)$ by the maximum entropy method, we used the features in Table 2. The features were decided based on an analysis of linefeed data. For more details, please refer to the literature [5].

4 Experiment

We conducted an experiment on inserting commas and linefeeds. We assume that our method can be used to format various types of texts. In the experiment, we evaluate the effectiveness of our method by using Japanese spoken language data.

4.1 Outline of Experiment

As the experimental data, we used the Japanese spoken language texts in the SIDB [6]. All the data were annotated with information on morphological analysis, dependency analysis and clause boundary detection by hand. The correct data of comma and linefeed insertion were created by inserting commas and linefeeds properly into this data by hand. Table 3 shows the size of the correct data. We performed a 16-fold cross-validation experiment by using this data. However, since we used 221 sentences among 1,935 sentences as the analysis data in our research [5], we evaluated the experimental result for the other 1,714 sentences (20,707 bunsetsus). Here, we used the maximum entropy method tool [7]. As for options, the repeat count on learning algorithm is set to 2000, and other options are set to default. In the experiment, we defined the maximum number of characters per line as 20.

Table 3. Size of correct data

sentence	bunsetsu	comma	linefeed
1,935	23,598	4,833	5,841

In the evaluation, we obtained the recall and precision. The recall and precision of comma insertion are respectively defined as follows.

$$recall = \frac{\#\ of\ correctly\ inserted\ commas}{\#\ of\ commas\ in\ the\ correct\ data}$$

$$precision = \frac{\#\ of\ correctly\ inserted\ commas}{\#\ of\ automatically\ inserted\ commas}$$

The recall and precision of linefeed insertion are defined similarly.

4.2 Experimental Result

Table 4 shows the experimental result. Both F-measures were higher than 70, which showed an effectiveness of our method.

Table 4. Experimental result

	recall	precision	F-measure
commas	71.65% (2,899/4,046)	81.07% (2,899/3,576)	76.07
linefeeds	76.91% (3,878/5,042)	69.19% (3,878/5,605)	72.85

姉の結婚式にまつわりまして、 ———— (A As for my sister's wedding,)
まずパンフレットの作成、 ———— (such as making a brochure for the wedding ceremony,)
それから、私は母親の看病、 ———— (nursing my mother,)
そして、会社に行っておりましたので、 ———— (and, as a salaried worker,)
仕事の方をしたりとか、 ———— (going to work)
毎日大変な雑用とかをしておりました ———— (I was taking charge of doing various things every day.)

トースター用のマイクロコード、 ———— (The micro code for toasters)
つまりソフトウェアですが、 ———— (means software for toasters,)
それと衛星のフライトダイナミックス、 ———— (and the flight dynamics for satellites)
航空力学のソフトですね、 ———— (means software for aeronautic dynamics.)
これを作ろうというのに ———— (In order to manufacture both of them,)
同じプロセス、メソッド、組織、ツール、 (the same process method, the same organization and
それらを使って両方作ってしまおうとは (the same tools
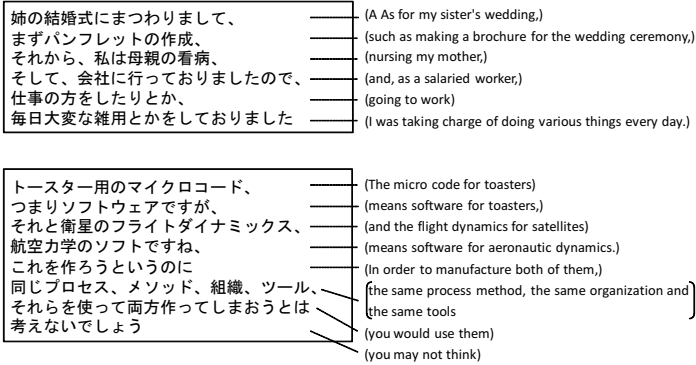考えないでしょう (you would use them)
(you may not think)

Fig. 4. Examples of comma and linefeed insertion by our method

Figure 4 shows the results of comma and linefeed insertion into the following two sentences by our method:

- 姉の結婚式にまつわりましてまずパンフレットの作成それから私は母親の看病そして会社に行っておりましたので仕事の方をしたりとか毎日大変な雑用とかをしておりました (As for my sister's wedding, I was taking charge of doing various things every day, such as making a brochure for the wedding ceremony, nursing my mother, and going to work as a salaried worker.)
- トースター用のマイクロコードつまりソフトウェアですがそれと衛星のフライトダイナミックス航空力学のソフトですねこれを作ろうというのに同じプロセスメソッド組織ツールそれらを使って両方作ってしまおうとは考えないでしょう (The micro code for toasters means software for toasters and the flight dynamics for satellites means software for aeronautic dynamics. In order to manufacture both of them, you may not think you would use the same process method, the same organization and the same tools.)

As shown in Figure 4, readable texts have been generated by our method.

4.3   Discussion

In this subsection, we focus on the relation between positions where commas were inserted and positions where linefeeds were inserted in the evaluation data.

Among positions where a comma existed (4,046 positions) and positions where a linefeed existed (5,042 positions) in the evaluation data, there existed 2,519 positions where a comma and a linefeed were existed. Table 5 shows the recall of comma insertion

Table 5. Recalls at positions where a comma and a linefeed existed in the evaluation data

commas	79.24% (1,996/2,519)
linefeeds	87.89% (2,214/2,519)

and the recall of linefeed insertion at these positions. Each recall was higher than the recall shown in Table 4. This is because positions where a comma and a linefeed are inserted may be semantic boundaries, and, our features can capture these positions well. Also, there existed 1,527 positions where only a comma existed, and the recall at these positions was 59.14% (903/1,527). And, there existed 2,467 positions where only a linefeed existed, and the recall at these positions was 65.95% (1,664/2,523). Each recall was lower than the recall shown in Table 4.

Among positions where commas existed in the evaluation data, there existed 462 positions where our method inserted linefeeds incorrectly. In case that linefeeds are inserted at positions where commas existed in the evaluation data, short lines will be generated unnecessarily and the texts will become hard to read. On the other hand, among positions where linefeeds existed in the evaluation data, there existed 273 positions where our method inserted commas incorrectly. Commas inserted too much harm readability of the texts. Because our method inserts commas and linefeeds sequentially, our method cannot consider linefeeds when our method decides whether or not to insert a comma. To realize more flexible comma and linefeed insertion, to develop a method which inserts commas and linefeeds at the same time is desired.

5   Conclusion

This paper proposed a method for formatting Japanese texts in social media. Our method formats texts by inserting commas and linefeeds at proper positions. In our method, commas and linefeeds are inserted into a sentence by machine learning, based on the information such as morpheme, dependencies and clause boundaries. An experiment by using Japanese spoken language data showed F-measure of comma insertion was 76.07 and F-measure of linefeed insertion was 72.85, and we confirmed the effectiveness of our method.

Our method identifies positions where commas are inserted first, and then, inserts linefeeds. Therefore, our method cannot judge a bunsetsu boundary as a position where a comma and a linefeed are inserted, or where a comma is not inserted but a linefeed is inserted at the same time. However, humans would not insert commas and linefeeds sequentially. In the future, we will develop more flexible method which inserts commas and linefeeds at the same time.

References

1. Knight, K., Marcu, D.: Summarization beyond sentence extraction: a probabilistic approach to sentence compression. Artificial Intelligence 139(1) (2002)
2. Shitaoka, K., Nanjo, H., Kawahara, T.: Automatic Transformation of Lecture Tr anscription into Document Style Using Statistical Framework. In: Proceedings of 8th International Conference on Spoken Language Processing, pp. 2169–2172 (2004)
3. Ito, M., Ohno, T., Matsubara, S.: Text-Style Conversio n of Speech Transcript intoWeb Document for Lecture Archive. Journal of Advanced Computational Intelligence and Intelligent Informatics 13(4), 499–505 (2009)
4. Murata, M., Ohno, T., Matsubara, S.: Automatic Comma Insertion for Japanese Text Generation. In: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, pp. 892–901 (2010)
5. Ohno, T., Murata, M., Matsubara, S.: Linefeed Insertion into Japanese Spoken Monologue for Captioning. In: Proceedings of Joint conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, pp. 531–539 (2009)
6. Matsubara, S., Takagi, A., Kawaguchi, N., Inagaki, Y.: Bilingual Spoken Monologue Corpus for Simultaneous Machine Interpretation Research. In: Proceedings of Language Resources and Evaluation Conference, pp. 153–159 (2002)
7. Le, Z.: Maximum entropy modeling toolkit for python and c++ (2008), http://homepages.inf.ed.ac.uk/s0450736/maxenttoolkit.html (Online accessed March 1, 2008)

Emotion Recognition from Body Movements and Gestures

Ioanna-Ourania Stathopoulou and George A. Tsihrintzis

University of Piraeus, Department of Informatics,
80 Karaoli and Demetriou St, Piraeus 18534, Greece
{iostath,geoatsi}@unipi.gr

Abstract. Towards building new, friendlier human-computer interaction and multimedia interactive services systems, new computer techniques are applied which aim at revealing information about the affective state, cognitive activity, personality, intention and psychological state of a person. In this paper we conduct a survey on systems which attempt to recognize human emotion from body movements and/or hand/arm gestures. Our survey shows that, although there are a lot of researches on human tracking, human motion analysis and gesture analysis, there are not many researches regarding emotion recognition from body movements and gestures. Moreover, there is a lack of databases depicting human movements and/or hand gestures in different states of emotions.

1 Introduction

With the progress of computer vision and multimedia technologies, there has been much interest in motion understanding. Human motion analysis is currently one of the most active research fields in computer vision. The aim is to detect, track and identify people, and more generally, to interpret human behaviors, from image sequences involving humans. A significant part of this task consists of capturing large scale body movements, such as movements of the head, arms, torso, and legs. For this reason and because of the similarities (i.e., both involve articulated structures and non-rigid motions, same techniques are used towards both tasks, etc.) gesture analysis can be considered as a part of human motion analysis. D. M. Gavrila [1] identified this research field as *"looking at people"* and presented some of the most important areas that this research can be applied on with promising results, which as summarized in Table 1.

Although there are a lot of researches on human tracking, human motion analysis and gesture analysis, there are only quite a few researches regarding emotion recognition from body movements and gestures. The majority of these approaches have appeared in the last decade and aim at facilitating human computer interaction by adding the ability to the computers to recognize the users' emotional states.

In this paper we conduct a survey on these approaches, which can be divided in two categories: studies in emotion recognition from body movements and/or hand/arm gestures which were part of a multimodal emotion recognition system

Table 1. Applications of "Looking at People" [1]

General domain	Specific area
Virtual reality	Interactive virtual worlds
	Games
	Virtual studios
	Character animation
	Teleconferencing (e.g., film, advertising, home-use)
"Smart" surveillance systems	Access control
	Parking lots
	Supermarkets, department stores
	Vending machines, ATMs
	Traffic
Advanced user interfaces	Social interfaces
	Sign-language translation
	Gesture driven control
	Signaling in high-noise environments (airports, factories)
Motion analysis	Content-based indexing of sports video footage
	Personalized training in golf, tennis, etc.
	Choreography of dance and ballet
	Clinical studies of orthopedic patients
Model-based coding	Very low bit-rate video compression

and, (2) independent studies emotion recognition from body movements and/or hand/arm gestures. The studies are compared in terms of the emotion classes, classification rate, use or non use of markers and the type of the movements that are being classified (stylized or non-stylized). Specifically, in Chapter 2 we describe the studies which try to recognize the emotion from body movements or hand/arm gestures independently, whereas in Chapter 3 we present the multimodal emotion recognition systems which use body movements and/or hand/arm gestures. Finally, we conclude and point to future work.

2   Emotion Recognition Systems from Body Movements and/or Hand/Arm Gestures

These systems try to recognize the emotional state of a person from the movements or the posture of the body as well as the movements of the hands or arms [2,3,4,5,6,7,8]. Specifically, Bianchi-Berthouze and Kleinsmith [2] addressed the issue of adding the capability to robots to recognize the affective state of the human by interpreting their gestural cues. To achieve that, they used an associative neural network called *categorizing and learning module* (CALM) [9], which incorporates brain-like structural and functional constraints such as modularity and organization with excitatory and inhibitory connections. To evaluate their

proposed framework, they collected emotional gestures with a VICON motion capture system. Twelve subjects of different gender, race and age were asked to perform as actors for the experiment. Dressed in a suit with 34 markers located on the joints and segments of his/her body, each subject performed an in-place freely decided emotional gesture. The total set consisted of 138 gestures depicting the following emotions: 'anger', 'happiness' and 'sadness'. The CALM network achieved a performance of 95,7% in classifying the emotions.

Bernhardt and Robinson [3] developed a computational model for describing and detecting affective content in everyday body movements. In their approach, they divided complex motions into a set of automatically derived motion primitives and, then, analyzed each primitive in terms of dynamic features which were shown to encode affective information. They also developed an algorithm to derive unbiased motion features adapted to personal movement idiosyncrasies, for persons already in the database. In order to classify the emotional states, they developed a SVM-based classifier and tested by applying Leave-One-Subject-Out cross-validation (LOSO-CV) tests. The system achieved 50% and 81% classification rates for biased and unbiased features, respectively.

Camurri et al. [4,5] explored the classification of expressive gesture in human full-body movement and in particular in dance performances. Their system consisted of four layers. In the first layer, they analyzed the video and movement (techniques for background subtraction, motion detection, motion tracking etc.). In the second layer, they computed the low-level features by applying computer vision techniques on the incoming images and computing statistical measures. In the third layer, they computed mid-level features and maps, e.g. gesture segmentation and representation of gestures as trajectories in semantic spaces (e.g., Laban's Effort space, energy-articulation space). In the final (fourth) layer, they classified the movement in four emotional classes, namely : 'anger', 'fear', 'joy' and 'grief'. For the classification step, they first used statistical techniques by computing a one-way ANOVA for each motion cue [4]. Later, they developed decision tree models for better results [5]. They developed five decision tree models, which achieved different performance among the various emotional states. They compared the system performance towards the classification accuracy of human spectators. Human spectators managed to classify the emotion with 56% accuracy, while the system achieved a 40% accuracy.

More recently, Castellano et al. [6] presented an approach for the recognition of acted emotional states from the body movements and gesture expressivity. Their approach was based on the fact that distinct emotions are often associated with different qualities of body movement. In order to classify the emotion, they used non-propositional movement qualities (e.g. amplitude, speed and fluidity of movement) to infer emotions, rather than trying to recognize different gesture shapes expressing specific emotions [10]. Their data set included 240 gestures which were collected when they asked ten participants (six male and four female) to project eight emotional states ('anger', 'despair', 'interest', 'pleasure', 'sadness', 'irritation', 'joy' and 'pride') equally distributed in the valence arousal space. To extract the silhouette and the hands of the subjects, they used the

EyesWeb platform [11]. They computed five different expressive motion cues: quantity of motion and contraction index of the body, velocity, acceleration and fluidity of the hands barycentre, using the EyesWeb Expressive Gesture Processing Library. They analyzed the emotional behaviour based on both direct classification of time series and a model that provides indicators describing the dynamics of expressive motion cues. For the first task, they built a nearest neighbour based on DTW distance (1NN-DTW) classifier. They evaluated their system over two perspectives: (1) train and evaluation of performance using only the personal set of gestures (i.e. training one classifier per subject), and (2) train and evaluation of performance using the universal set (i.e. an inter-subject enabled classifier using the universal set of gestures). Both classifiers were evaluated in the classification task of four (namely, 'anger', 'joy', 'pleasure' and 'sadness') and all the eight emotional states. Their system was not able to discriminate successfully between the eight emotions, but achieved an accuracy of 61% for the four emotions.

Burgoon et al. [7] considered contextual cues, as well as analyzing cues from multiple body regions to lay the foundation for automatic identification of emotion from video. Their approach did not aim at classifying the body movements in discrete emotion classes, rather than identifying the person's emotional state in terms of positive/negative valence (pleasant/unpleasant) and high/low arousal/intensity. They first tracked the head and hands once an individual had been identified. In order to achieve this, they used a blob analysis developed by the Computational Bio-medicine Imaging and Modeling Center at Rutgers University [12], which is based on color analysis, eigenspace-based shape segmentation and Kalman filters to track head and hand positions throughout the video segment. General metrics, such as position, size, and angles were computed for each blob and utilized when generating meaningful features. For blob analysis, Time Delay Neural Networks (TDNN) and Recurrent Neural Networks (RNN) were tested in the task of classifying individual gestures and RNN were found more adequate.

Kapur et al. [8] classified full-body skeletal movements data obtained with a technology based on the VICON motion capturing system to classify four emotional states, namely: 'sadness', 'joy', 'anger' and 'fear'. VICON uses a series of 6 cameras to capture lightweight markers placed on various points of the body in 3-D space and digitizes movement into x, y, and z displacement data. They videotaped the movements of five subjects depicting the four emotions that they were asked to portray. They developed and tested five different classifiers: (1) a logistic regression, (2) a Naive-Bayes with a single multidimensional Gaussian distribution modeling each class, (3) a Decision Tree classifier based on the C4.5 algorithm, (4) a multi-layer perceptron back-propagation artificial neural network, and (5) a support vector machine trained using Sequential Minimal Optimization (SMO). Experimental results with different classifiers showed that automatic classification of this data ranged from 84% to 92% depending on how it was calculated. They also compared the system performance towards

the classification accuracy of human spectators, who managed to classify the emotion with 93% accuracy.

All these approaches are summarized in Table 3, based on the requirement set in Table 2.

Table 2. Requirements/Test for the emotion recognition systems from body movements and gestures

1	Type of movements (Can be 'stylized', when the subjects the emotion instead of feeling it, or, alternatively, 'non-stylized' or other)
2	Type of classifier
3	Use of markers
4	Number of emotion classes
5	Names of emotion classes
6	Classification rate

Table 3. Review of the emotion recognition systems from body movements and gestures, based on the requirements in Table 2

Reference	1	2	3	4	5	6
Bianchi-Berthouze [2]	non-stylized	CALM ANN	Yes	3	'angry', 'happy', 'sad'	95,7%
Bernhardt [3]	non-stylized	SVM	No	4	'neutral', 'angry', 'happy', 'sad'	50%, 81%
Camurri [4, 5]	dancing moves	ANOVA, Decision trees	No	4	'angry', 'fear', 'joy', 'grief'	40%
Castellano [6]	stylized	INN-DTW	No	4	'anger', 'joy', 'pleasure', 'sadness'	61%
Burgoon [7]	non-stylized	TDNN, RNN	No	-	No emotional classes	n.a.
Kapur [8]	stylized	Perceptron, SVM	Yes	4	'anger', 'sad', 'joy', 'fear'	84% - 92%

3   Multimodal Emotion Recognition Systems

These systems incorporate two or more input data in order to recognize the emotional state. The modalities usually include body movements and/or hand gestures with facial expressions [13,14].

Balomenos et al. [13] developed an intelligent rule-based system for emotion classification into the six basic emotions, by using facial and hand movement information. Facial analysis included a number of processing steps which attempted to detect or track the face, to locate characteristic facial regions such

as eyes, mouth and nose on it, to extract and follow the movement of facial features, such as characteristic points in these regions, or model facial gestures using anatomic information about the faces. Hand gesture analysis included tracking the person's hands by creating moving skin color areas which were tracked between subsequent frames. They tracked the centroid of those skin masks in order to estimate the user's movements. To facilitate the tracking process, they used a-priori knowledge related to the expected characteristics of the input image: the head is in the middle area of upper half of the frame and the hand segments near the respective lower corners. They trained a Hidden Markov Model (HMM) Classifier to classify among the following gesture classes: hand clapping - high frequency, hand clapping - low frequency, lift of the hand - low speed, lift of the hand - high speed, hands over the head - gesture, hands over the head - posture and italianate gestures. For the multimodal emotion classification task, the correlated these classes to the six emotional states. Specifically:

- *'Joy':* hand clapping - high frequency
- *'Sadness':* hands over the head - posture
- *'Anger':* lift of the hand - high speed, italianate gestures
- *'Fear':* hands over the head - gesture, italianate gestures
- *'Disgust':* lift of the hand - low speed, hand clapping-low frequency
- *'Surprise':* hands over the head - gesture

The two subsystems were combined as a weighted sum, where the weights were 0.75 and 0.25 for the facial expression recognition modality and the affective gesture recognition modality, respectively.

Gunes et al. [14] also fused facial expression and body gesture information for bimodal emotion recognition. Their system was tested for the six following emotional classes: 'anxiety', 'anger', 'disgust', 'fear', 'happiness' and 'uncertainty'. They also built their own bimodal database (FABO) that consisted of recordings of facial expressions alone and combined face and body expressions [15]. To form the training and testing set, they processed 54 sequences of frames in total, 27 for face and 27 for body from four subjects and selected the more expressive. Half of the frame were used for training and the other half for testing purposes. Their facial analysis subsystem was based on the FACS coding system, specifically, they identified the following for each emotion class:

- *'Anxiety':* Lip bite; stretching of the mouth; eyes turn up/down/left/right; lip wipe
- *'Happiness':* Corners of lips are drawn back and up; mouth may or may not be parted with teeth exposed or not; cheeks are raised; lower eyelid shows wrinkles below it, and may be raised but not tense; wrinkles around the outer corners of the eyes
- *'Anger':* Brows lowered and drawn together; lines appear between brows; lower lid is tensed and may or may not be raised; upper lid is tense and may or may not be lowered due to brows action; lips are either pressed firmly together with corners straight or down or open

- *'Fear':* Brows raised and drawn together; forehead wrinkles drawn to the center; upper eyelid is raised and lower eyelid is drawn up; mouth is open; lips are slightly tense or stretched and drawn back
- *'Disgust':* Upper lip is raised; lower lip is raised and pushed up to upper lip or it is lowered; nose is wrinkled; cheeks are raised; brows are lowered; tongue out
- *'Uncertainty':* Lid drop; inner brow raised; outer brow raised; chin raised; jaw sideways; corners of the lips are drawn downwards

They trained a BayesNet classifier which achieved an average performance of 76.40%.

For their body analysis subsystem, they identified the following for each emotion class:

- *'Anxiety':* Hands close to the table surface; fingers moving; fingers tapping on the table
- *'Happiness':* Body extended; hands kept high; hands made into fists and kept high
- *'Anger':* Body extended; hands on the waist; hands made into fists and kept low, close to the table surface
- *'Fear':* Body contracted; body backing; hands high up, trying to cover bodily parts
- *'Disgust':* Body backing; left/right hand touching the neck or face
- *'Uncertainty':* Shoulder shrug; palms up

They also used a BayesNet classifier which achieved an average performance of 89.90%. For the fusion of the two modalities, the applied two different methods:

- *Feature-level fusion:* They fused all the features from the two modalities into a bigger vector matrix. Best-first search method was used with ten-fold cross validation to obtain a decisive reduction in the features' number and fed to a BayesNet classifier. The average classification rate was 94.03% for subjects already in the database.
- *Decision-level fusion:* In this case, the final classification was based on the fusion of the outputs the different modalities. They tested various approaches, including the sum rule, product rule and the use of using weights. Sum rule provided the best fusion result with average recognition accuracy of 91.1%.

Finally, el Kaliouby and Robinson [16] proposed a vision-based computational model to infer acted mental states from head movements and facial expressions. Their system was evaluated on videos that were posed by lay people in a relatively uncontrolled recording environment for six mental statesagreeing, concentrating, disagreeing, interested, thinking and unsure. They used a use Dynamic Bayesian Networks (DBNs) to model the unfolding of mental states over time and each display was represented as a Hidden Markov Model (HMM) of a sequence of head/facial actions, recognized non-intrusively, in real time. Head actions were described by the magnitude and direction of 3 Euler angles, while facial actions

were extracted using motion, shape and colour analysis of the lips, mouth and eyebrows. The writers had trained and tested their system on videos from the Mind-Reading DVD (MR) [17], a guide to emotions developed for Autism Spectrum Disorders and achieve an average accuracy of 77.4%. They further tested the generalization of their system by collecting videos at the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2004). The classification rate was highest for disagreeing (85.7%) and lowest for thinking (26.7%).

4   Conclusions and Future Work

Generally, in the area of affective computing, there is a lack of researches on emotion recognition from body movements and hand/arm gestures, compared to the number of researches on human motion analysis and/or hand gesture analysis. Although there has been some psychological studies regarding emotion and nonverbal communication in expressive movements [18,19,20], these studies were based on acted, stylized basic emotions. But, in affective computing the spontaneity of the emotions is very important and must be taken into account. Moreover, there is a lack of a widely accepted model which can represent human movements with regards to the experienced emotion, for example, like Facial Action Coding System (FACS) [21] which is considered the basis in the development of facial expression analysis systems. Maybe this is why that the majority of the approaches in emotion recognition from body movements and hand/arm gestures try to develop their own model by observing the persons' movements when they experiencing an emotion. Despite all the difficulties, the results are very promising, especially for the development of multimodal affective systems, and further research is necessary.

References

1. Gavrila, D.M.: The visual analysis of human movement: a survey. Computer Vision and Image Understanding 73, 82–98 (1999)
2. Bianchi-Berthouze, N., Kleinsmith, A.: A categorical approach to affective gesture recognition. Connection Science 15, 259–269 (2003)
3. Bernhardt, D., Robinson, P.: Detecting emotions from everyday body movements. In: Presenccia PhD Symposium, Barcelona, Spain (2007)
4. Camurri, A., Lagerlof, I., Volpe, G.: Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques. International Journal of Human-Computer Studies 59, 213–225 (2003); Applications of Affective Computing in Human-Computer Interaction
5. Camurri, A., Mazzarino, B., Ricchetti, M., Timmers, R., Volpe, G.: Multimodal analysis of expressive gesture in music and dance performances. In: Camurri, A., Volpe, G. (eds.) GW 2003. LNCS (LNAI), vol. 2915, pp. 20–39. Springer, Heidelberg (2004)
6. Castellano, G., Villalba, S.D., Camurri, A.: Recognising human emotions from body movement and gesture dynamics. In: Paiva, A.C.R., Prada, R., Picard, R.W. (eds.) ACII 2007. LNCS, vol. 4738, pp. 71–82. Springer, Heidelberg (2007)

7. Burgoon, J.K., Meservy, L.J.M., Kruse, T.O., Nunamaker, J., Augmenting, J.F.: human identification of emotional states in video. In: Proceedings of the International Conference on Intelligent Data Analysis (2005)

8. Kapur, A., Kapur, A., Virji-Babul, N., Tzanetakis, G., Driessen, P.F.: Gesture-based affective computing on motion capture data. In: Tao, J., Tan, T., Picard, R.W. (eds.) ACII 2005. LNCS, vol. 3784, pp. 1–7. Springer, Heidelberg (2005)

9. Murre, J.M.J., Phaf, R.H., Wolters, G.: Original contribution: Calm: Categorizing and learning module. Neural Netw 5, 55–82 (1992)

10. Pollick, F.E., Paterson, H.M., Bruderlin, A., Sanford, A.J.: Perceiving affect from arm movement. Cognition 82, 51–61 (2001)

11. Camurri, A., Coletta, P., Massari, A., Mazzarino, B., Peri, M., Ricchetti, M., Ricci, A.G.: Toward real-time multimodal processing: Eyesweb 4.0. In: AISB 2004 Convention: Motion, Emotion and Cognition (2004)

12. Lu, S., Tsechpenakis, G., Metaxas, D.N.: Blob analysis of the head and hands: A method for deception detection. In: Hawaii International Conference on System Sciences and Emotional State Identification, Big Island (2005)

13. Balomenos, T., Raouzaiou, A., Ioannou, S., Drosopoulos, A., Karpouzis, K., Kollias, S.D.: Emotion Analysis in Man-Machine Interaction Systems. In: Bengio, S., Bourlard, H. (eds.) MLMI 2004. LNCS, vol. 3361, pp. 318–328. Springer, Heidelberg (2005)

14. Gunes, H., Piccardi, M.: Bi-modal emotion recognition from expressive face and body gestures. J. Netw. Comput. Appl. 30, 1334–1345 (2007)

15. Gunes, H., Piccardi, M.: A bimodal face and body gesture database for automatic analysis of human nonverbal affective behavior. In: ICPR 2006: Proceedings of the 18th International Conference on Pattern Recognition, pp. 1148–1153. IEEE Computer Society, Washington, DC, USA (2006)

16. el Kaliouby, R., Robinson, P.: Generalization of a vision-based computational model of mind-reading. In: Tao, J., Tan, T., Picard, R.W. (eds.) ACII 2005. LNCS, vol. 3784, pp. 582–589. Springer, Heidelberg (2005)

17. Baron-Cohen, S., Golan, O., Wheelwright, S., Hill, J.J.: Mind Reading: The Interactive Guide to Emotions. Jessica Kingsley Publishers, London (2004)

18. De Meijer, M.: The contribution of general features of body movement to the attribution of emotions. Journal of Nonverbal Behaviour 13, 247–268 (1989)

19. Wallbott, H.G.: Bodily expression of emotion. European Journal of Social Psychology 28, 879–896 (1998)

20. Shaarani, A., Romano, D.: Basic emotions from body movements. In: Proc of CCID 2006:The First International Symposium on Culture, Creativity and Interaction Design. HCI 2006 Workshops, The 20th BCS HCI Group Conference. University of London, Queen Mary (2006)

21. Ekman, P., Friesen, W.: Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, Palo Alto (1978)

Using Two Stage Classification for Improved Tropical Wood Species Recognition System

Anis Salwa Mohd Khairuddin[2], Marzuki Khalid[1,*], and Rubiyah Yusof[1]

[1] Center for Artificial Intelligence and Robotics, Universiti Teknologi Malaysia,
Kuala Lumpur, Malaysia
[2] Department of Electrical Engineering, University of Malaya,
Kuala Lumpur, Malaysia
marzuki.khalid@gmail.com, rubiyah@ic.utm.my, anissalwa@um.edu.my

Abstract. An automated wood recognition system is designed based on five stages: data acquisition, pre-processing images, feature extraction, pre classification and classification. The proposed system is able to identify 52 types of wood species based on wood features extracted using Basic Grey Level Aura Matrix (BGLAM) technique and statistical properties of pores distribution (SPPD) technique. The features obtained from both feature extractors are fused together and will determine the classification between the various wood species. In order to enhance the class separability, a pre-classification stage is developed which includes clustering and dimension reduction. K-means clustering is introduced to cluster the 52 wood species. As for dimension reduction, we proposed linear discriminant analysis (LDA) to solve linear data and kernel discriminant analysis/ generalized singular value decomposition (KDA/GSVD) to solve nonlinearly structured data. For final classification, K-Nearest Neighbour (KNN) classifier is implemented to classify the wood species.

Keywords: wood recognition, LDA, KDA/GSVD, K-means cluster, kNN classifier.

1 Introduction

Currently, very few certified officers are involved in the traditional wood identification process. The process of training up experienced officers in performing the job is difficult since there are approximately more than 1000 wood species in Malaysia. Moreover, the possibility of biasness in human recognition system has to be considered. Besides that, it is impractical and cost effective for a human to analyze and identify large amount of timber species. Hence automatic wood species recognition system is needed to overcome the errors caused by traditional wood identification system which based solely on human expertise.

Khalid et. al [11] has developed an automatic Tropical Wood Species Recognition System using macroscopic features of timber based on image processing. In the work, the surface texture of timber is captured, and the features are extracted from the images

---

* Corresponding author.

using Grey Level Co-occurrence Matrix (GLCM) and back propagation Neural Network (BPNN) is used to train and classify the timber species. Yusof et. al [12] adopted fusion of two feature sets using multi feature extractor technique. Two feature extraction methods were used in this system: co-occurrence matrix approach (GLCM) and Gabor filters. These two extraction methods produced more variation of features and also improved the accuracy rate of the system. Gabor filters are used in the pre-processing stage of the wood texture image to multiply the number of features for a single image, thus providing more information for feature extractor to capture. K. Wang et.al [13] proposed a wood surface texture recognition method based on feature level data fusion which uses three types of texture analysis methods. The three texture analysis methods used are gray co-occurrence matrix (GLCM), Gauss-Markov random field (GMRF) and wavelet multi-resolution fractal dimension. The fusion method is based on Simulated Annealing Algorithm with memory and the recognition rate of nearest neighbor classifier, fuses the extracted parameters in the feature level. In a later development, Khairuddin et.al [14] used feature selection based on genetic algorithm to improve the accuracy of wood species recognition, in particular to reduce the redundant features which are not considered as discriminatory enough for accurate classification.

Due to the large variations of features within inter and intra among the species of the tropical wood, and the problems relating to variations in the wood samples, there is a pressing need to improve the methodology of the wood recognition system. The conventional method of using the feature extraction followed by classification in the recognition of wood species is no longer adequate for a larger sample of wood species. Therefore, in this paper, we proposed a two step classifications for the wood recognition system. The first step is called the pre classification stage utilizes the clustering and dimension reduction techniques. K-means clustering is introduced to cluster kernel in order to enhance the class-discriminatory information in the lower-dimensional space. LDA is applied to reduce dimension of linear data while KDA/GSVD is used to reduce dimension of nonlinear data. Finally, at the second classification stage, we used k-NN (K nearest neighbour) classifier to classify the wood species. The experimental results presented in section 3 shows that the implementation of clustering and dimension reduction improves the classification accuracy of the system.

This paper is organized as follows: in Section 2, we presented the methodology of the proposed system. In section 3, experimental results are discussed followed by brief conclusion in section 4.

2   Proposed Methodologies

The methodology of the proposed system consists of image acquisition, image pre-processing, feature extraction, clustering in the pre classification   stage and classification. Figure 1 shows the overview of the proposed system.

2.1   Image Acquisition

The wood samples for this research are obtained from the Forest Research Institute of Malaysia (FRIM). There are 52 wood species in cubic form (approximately 1 inch by

1 inch in size) where 5 cubes are provided for each species. The images of the wood surface are captured by using a specially designed portable camera with 10 times magnification. The size of the each image is 768x576 pixels. In total, there are 100 images taken from each wood species where 90 images are for training and the other 10 images are for testing.
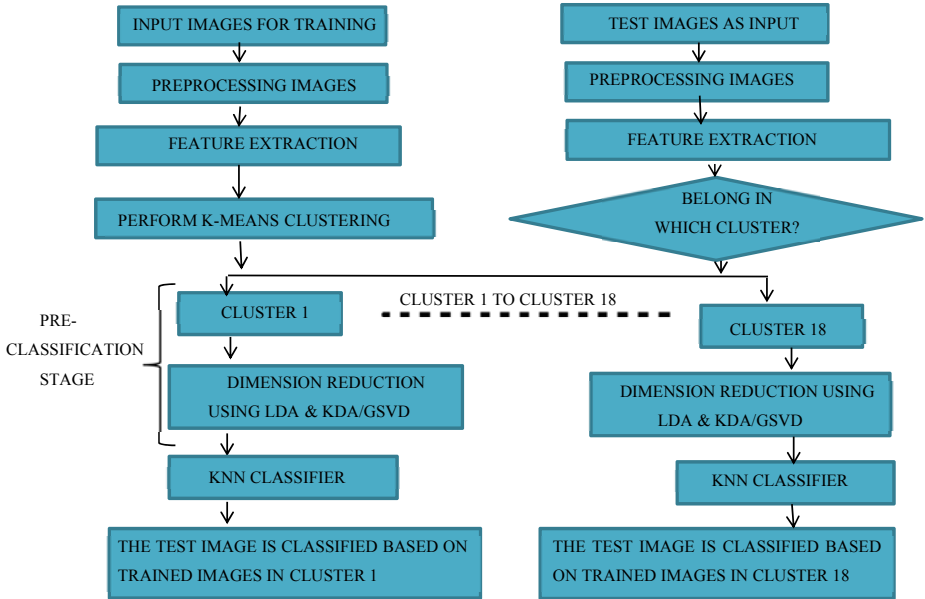

Fig. 1. Overview of the proposed system

2.2 Image Pre-processing

The wood images are pre-processed using homomorphic filters in order to enhance the image presentation. Homomorphic filtering sharpens features and flattens lighting variations in an image. Hence, illumination and reflectance on the image can be removed. Figure 2 below shows the pre-processed images of wood species:
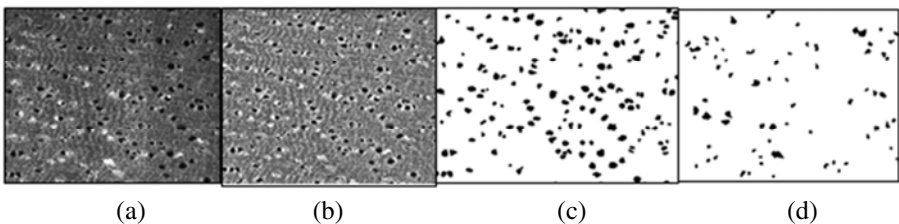
Fig. 2. (a) Original image of 'bitis', (b) Homomorphic image of 'bitis',(c) Black pores image of 'bitis', (d) White pores image of 'bitis'

2.3   Feature Extraction

In this paper, we use the same features that have been used by Khairuddin et.al [14]. In their work, the wood features were extracted based on the fusion of Basic Grey Level Aura Matrix (BGLAM) & Statistical Properties of Pores Distribution (SPPD). The total number of features extracted from each wood sample is 157 features (136 features from BGLAM and 21 features from SPPD).

An image can be uniquely represented by its BGLAMs. In essence, the BGLAMs of an image characterize the co-occurrence probability distributions of gray levels at all possible displacements configurations. It is proven that BGLAMs can give the necessary and sufficient information to differentiate between images. In addition, BGLAM features are directly calculated from the images thus providing a more accurate representation of an image. The framework of BGLAM feature extractor is explained in detail in [10].

SPPD extracts statistical properties of pores distribution on the image sample. These features will allow only distinct pores to be acknowledged as characteristics of a species. From the images shown in figure 2, these features are extracted from each wood sample:

1. Mean size of the pores and the corresponding standard deviation (2 features)
2. Mean distance between pores and the corresponding standard deviation (2 features)
3. Number of small, medium and large pores (3 features)
4. Number of pairs and solitary pores (2 features)
5. Number of pores per square millimeters (1 feature)

By using SPPD feature extrtactor, 10 features are obtained from black pores image and 10 features are obtained from white pores image. Another 1 feature is the mean grey level feature which is obtained from the original image. Hence, total features extracted from each image by using SPPD technique is 21 features.

2.4   Pre Classification stage

Pre-classification stage includes clustering using k-means clustering technique and dimension reduction using LDA and KDA/GSVD techniques. The goal of pre-classification stage is to enhance the class-discriminatory information in a lower-dimensional space.

2.4.1   Clustering

Based on the features extracted from each wood sample, the wood database is clustered into 18 clusters using k-means clustering. The number of cluster is chosen based on results presented in Table 1 in section 3. Each cluster will have its own wood database and classification stage.

Given a set of training wood samples n (x_1, x_2, ..., x_n), where n represents the number of training wood samples in the database. Each wood sample, x is a real vector of wood features where $x = (x_1,...,x_{157})$ since there are 157 features extracted

from each wood sample. K-means clustering aims to partition n wood samples into k clusters ($k \leq n$)

$$S = \{S_1, S_2, ..., S_k\} \tag{1}$$

so as to minimize the within-cluster sum of squares as in equation 1:

$$\underset{S}{\arg min} \sum_{i=1}^{k} \sum_{x_n \in S_i} \|x_n - \mu_i\|^2 \quad , \mu_i \text{ is the mean of points in } S_i. \tag{2}$$

The algorithm of the k-means clustering is shown below:

1) Determine number of cluster, k.
2) Determine initial set of k means $\mu_1, ..., \mu_k$ as cluster centers using random permutation.
3) Assign each wood sample to the nearest cluster center based on minimum-distance classifier. That is, we can say that x_n is in cluster k if $\| x_n - \mu_i \|$ is the minimum of all the i distances based on equation 2.

$$S_i = \{x_n : \|x_n - \mu_i\| \leq \|x_n - \mu_{i^*}\| \quad , \qquad \text{for all } i^* = 1,...,k\} \tag{3}$$

4) Based on the clustering results from step 3, we will update the cluster centers, μ_i by calculating new means to be the centroid of the wood samples in the cluster i. Where x_i represents the wood samples belong to cluster i. The centroid is updated based on equation 4.

$$\mu_i = \frac{1}{|S_i|} \sum_{x_i \in S_i} x_i \tag{4}$$

5) Repeat step 3 and 4 until the centroid for every clusters no longer move which means convergence has been reached.
6) Then, variance can be used to measure how far a set of wood samples are spread out from each other within the same cluster. Variance of each cluster, v_i is calculated based on equation 5.

$$v_i = \frac{\sum(x_i - \mu_i)^2}{|S_i|} \tag{5}$$

7) Average variance, v_{avg} is calculated in order to determine the right number of cluster to be used in the pre-classification stage. Smaller value of v_{avg} will optimize the clustering.

$$v_{avg} = \frac{\sum_{i=1}^{k} v_i}{k} \tag{6}$$

2.4.2 Dimension Reduction Using LDA and KDA/GSVD

The objective of LDA is to perform dimensionality reduction while preserving as much of the class discriminatory information as possible. But this method fails for a nonlinear problem. In order to make LDA applicable to nonlinearly structured data, kernel-based method has been applied which is known as kernel discriminant analysis (KDA). In KDA, Radial basis function (RBF) kernel function was employed into LDA to handle nonlinearity in a computationally attractive manner. However, due to the nonlinear map by the kernel function, the dimension of the feature space often becomes much larger than that of the original data space, and as a result, the scatter matrices become singular, which is referred to as "small sample size" (SSS) problem. Hence, KDA/GSVD has been proposed, which is a generalization of LDA based on the generalized singular value decomposition (GSVD). It overcomes the SSS problem.

Linear Discriminant Analysis (LDA)

The concept of LDA is to find vectors in the underlying space which best discriminate among classes. This idea is achieved by maximizing the within class variance. In the proposed system, there are $c = 52$ distinct classes in all the training samples, M_i is the number of training samples in class i, thus, the between-class scatter matrix S_b and the within-class scatter matrix S_w can be defined by:

$$S_b = \sum_{i=1}^{c} M_i (x_i - \mu)(x_i - \mu)^T \tag{7}$$

$$S_w = \sum_{i=1}^{c} \sum_{x_k \in X_i} (x_k - \mu_i)(x_k - \mu_i)^T c \tag{8}$$

Here, x_i represents the set of wood images belonging to class i, μ_i is the mean of samples in class i, μ is the mean of all samples and x_k is the kth image of that class. LDA computes a transformation that maximizes the between-class scatter while minimizing the within-class scatter:

$$\text{Maximize } \frac{det(S_b)}{det(S_w)} \tag{9}$$

Nonlinear Discriminant Analysis Based on Kernel Functions KDA and GSVD

In this section, we present a nonlinear extension of LDA based on kernel functions and the GSVD. The main idea of the kernel method is that without knowing the nonlinear feature mapping or the mapped feature space explicitly, we can work on the feature space through kernel functions, as long as the problem formulation depends only on the inner products between data points. This is based on the fact that for any kernel function k satisfying Mercer's condition, there exists a mapping φ such that,

$$< \varphi(a), \varphi(b) >= k(a,b) \tag{10}$$

Where $<,>$ is an inner product in the feature space transformed by φ.

We apply the kernel method to perform LDA in the feature space instead of the original input space. Given a kernel function k, let φ be a mapping satisfying equation 10 and define $F \subset R^N$ to be the feature space from the mapping φ. A Gaussian is used as the RBF kernel based on equation 11.

$$k(A,B) = K_{rbf2D}(A,B) = exp(-\frac{d^2(A,B)}{\sigma^2}) \quad , \text{where } \sigma \in R \tag{11}$$

The RBF kernel returns an RBF kernel matrix from the input coordinates. The inputs of the RBF kernel are a matrix containing all wood features and the kernel width. The output of the kernel is dependent on the Euclidean distance of B from A (one of these will be the support vector and the other will be the testing data point. The support vector will be the center of the RBF and σ will determine the area of influence this support vector has over the data space.

The algorithm of KDA/GSVD is shown below [5]:

Given a data matrix $A = [a_1, ..., a_n] \in R^{mxn}$ with r classes and a kernel function K, it computes the r-1 dimensional representation of any input vector $z \in R^m$ by applying GSVD in the feature space defined by the feature mapping φ such that $K(a_i, a_j) = <\varphi(a_i), \varphi(a_j)>$

1) Compute between-class scatter matrix, K_b and within-class scatter matrix, K_w.
2) Apply GSVD to the pair K_b and K_w. And we have vector of singular matrix,

$$X^T K_b K_b^T X = \begin{bmatrix} \Gamma_b^T \Gamma_b & 0 \\ 0 & 0 \end{bmatrix} \tag{12}$$

$$X^T K_w K_w^T X = \begin{bmatrix} \Gamma_w^T \Gamma_w & 0 \\ 0 & 0 \end{bmatrix} \tag{13}$$

3) Assign the first r-1 columns of X to G.

$$G = [\alpha^{(1)}, ..., \alpha^{(r-1)}] = \begin{bmatrix} \alpha_{1,1} & \cdots & \alpha_{1,r-1} \\ \vdots & \vdots & \vdots \\ \alpha_{n,1} & \cdots & \alpha_{n,r-1} \end{bmatrix} \tag{14}$$

4) For any input vector, $z \in R^{mx1}$, a dimension reduced representation is computed as

$$G^T \begin{bmatrix} K(a_{1,z}) \\ \vdots \\ K(a_{n,z}) \end{bmatrix} \in R^{(r-1)x1} \tag{15}$$

3   Final Classification

KNN classifier is used to classify the wood species. The test data is classified by calculating the Euclidean distance, $d(x,y)$ to the nearest training data as in equation 16. Then, use simple majority of the category of nearest neighbours to determine the wood species of the test data.

$$d(x,y) = \sqrt{\sum_{i=1}^{157}(x_i - y_i)^2} \qquad (16)$$

Where x represents the training data and y represents the test data.

The efficiency of the proposed system is calculated using equation 17.

$$Efficiency = \frac{\text{number of correctly classified test images}}{\text{number of test images}} \times 100 \qquad (17)$$

4   Experimental Results

We performed several experiments to investigate the effectiveness of the proposed method. For clustering method, it is important to select the correct number of clusters for good recognition rate. Based on results shown in Table 1, the classification accuracy increases from 87% to 96.15% as the number of cluster increases from $k=6$ to $k=18$. The highest accuracy rate is achieved when $k=18$. Hence, the right number of cluster that we chose to cluster the wood database is 18 clusters.

Table 2 shows that the variance of each cluster varies for different number of k. The variance is computed in order to measure how far a set of wood samples are spread out from the cluster centroid within the same cluster. Average variance is calculated to compare the optimization of clustering for different number of clusters, k. The smallest average variance is achieved when k=18 with value $v_{avg} = 0.5049$ which results to highest classification accuracy shown in table 1.

The implementation of pre classification stage in the proposed recognition system has improved the classification accuracy from 84% to 96.15%. Besides that, the introduction of clustering technique has also shown to improve the proposed system with increment of classification accuracy from 89.5% to 96.15%.

Table 1. The wood classification accuracy based on different number of cluster

Num.of cluster, k	$k=6$	$k=11$	$k=15$	$k=18$
Classification accuracy	87.0 %	89.4%	94.0%	96.15%

Table 2. Comparison of variance within each cluster for different number of clusters

Number of cluster, k	Variance for each cluster, v_k	Average variance, v_{avg}
$k = 6$	[1.3569 2.2933 1.5373 2.0283 1.8009 2.7366]	1.9589
$k = 11$	[1.3300 1.0287 1.1149 1.1070 1.7326 1.5403 2.8631 2.5559 1.0920 1.0404 2.0154]	1.5837
$k = 15$	[1.3364 0.9942 1.3886 1.3101 1.4616 0.6380 1.0404 1.1671 1.3713 1.1902 1.3382 0.6547 1.1551 2.0420 1.4306]	1.2346
$k = 18$	[0.4760 0.3379 0.2738 0.5837 0.5940 0.2338 0.8384 0.2712 0.5067 0.5354 0.4832 0.3235 0.4135 1.1128 0.2485 0.6211 0.6738 0.5606]	0.5049

Table 3. The wood classification accuracy in different conditions

System condition	Classification accuracy
Proposed system without including the clustering technique	89.5 %
The complete proposed system	96.15%

5   Conclusion

This paper concludes that the implementation of k-means clustering and dimension reduction as the pre-classification stage has reduced the computing time and improved the performance of the wood recognition system. The k-means clustering enables the system to only compute the wood database in a respective cluster instead of computing the entire database to classify a wood species while dimension reduction enables the wood samples to be represented accurately in a lower-dimensional space.

References

[1] XiaoKai, Z., Xiang, L.: Image kernel for recognition. In: Proceedings of ICSP 2008. IEEE, Los Alamitos (2008)

[2] Ye, F., Shi, Z., Shi, Z.: A comparative study of PCA, LDA and Kernel LDA for image classification. In: 2009 International Symposium on ubiquitous virtual reality. IEEE, Los Alamitos (2009)

[3] Chen, W.-S., Yuen, P.C., Ji, Z.: Kernel subspace LDA with convolution kernel function for face recognition. In: 2010 International Conference on Wavelet Analysis and Pattern Recognition, Qingdao (2010)

[4] Jiang, Y., Chen, X., Guo, P., Lu, H.: An improved random sampling LDA for face recognition. In: 2008 Congress on Image and Signal Processing (2008)

[5] Park, C.H., Park, H.: Nonlinear discriminant analysis using kernel functions and the generalized singular value decomposition. SIAM Journal on Matrix Analysis and Applications (2005)

[6]   Howland, P., Park, H.: Generalizing Discriminant Analysis Using the Generalized Singular Value Decomposition. IEEE Transaction, PAM (2004)

[7]   Lu, J., Plataniotis, K.N., Venetsanopoulos, A.N.: Regularized discriminant analysis for the small sample size problem in face recognition. Pattern recognition Letters (2003)

[8]   Wang, L., Bo, L., Jiao, L.: Kernel uncorrelated discriminant analysis for radar target recognition. Springer, Heidelberg (2006)

[9]   Teknomo, Kardi. Discriminant Analysis Tutorial,
      `http://people.revoledu.com/kardi/tutorial/LDA/`

[10] Qin, X., Yang, Y.-H.: Similarity measure and learning with gray level aura matrices (GLAM)for texture image retrieval. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR 2004, vol. 1, pp. I-326 - I-333 (2004)

[11] Khalid, M., Lew, Y.L., Yusof, R., Nadaraj, M.: Design of An Intelligent Wood Species Recognition System. IJSSST 9(3) (September 2008)

[12] Yusof, R., Rosli, N.R., Khalid, M.: Using Gabor Filters as Image Multiplier for Tropical Wood Species Recognition System. In: 12th International Conference on Computer Modelling and Simulation, pp. 289–294 (2010)

[13] Wang, K., Bai, X.: Research on classification of wood surface texture based on feature level data fusion. In: 2nd IEEE Conference on Industrial Electronics and Applications, ICIEA (2007)

[14] Khairuddin, U., Yusof, R., Khalid, M., Cordova, F.: Optimized Feature Selection for Improved Tropical Wood Species Recognition System. ICIC Express letters, Part B: Applications, An International Journal of Research and Surveys 2(2), 441–446 (2011)

Text Summarization of Single Documents Based on Syntactic Sequences

Paul Villavicencio and Toyohide Watanabe

Department of Systems and Social Informatics
Graduate School of Information Science, Nagoya University
Nagoya 464-8603, Japan
{paul,watanabe}@watanabe.ss.is.nagoya-u.ac.jp

Abstract. In this paper we propose a summarization method for scientific articles from the viewpoint of the syntactic sequences. The objective is to generate an extractive summary by ranking sentences according to their informative content, on the basis of the idea that the writing styles of authors create syntactic patterns which may contain important information about topics explained in a research paper. We use two main document features in our summarizing algorithm: syntactic sequences and frequent terms per section. We present an evaluation of our proposed algorithm by comparing it with existing summarization methods.

Keywords: text summarization, text analysis, parts of speech.

1   Introduction

In the scientific area an important task is to search and select information from research papers. This task is usually time-consuming and difficult, since information is created daily. Automatic summarization tools assist this task and provide brief overviews of documents. Although different methods have been used in summarization, it is still difficult to obtain good results. One of the core steps in automatic document summarization is the selection of sentences which contain representative contents. Contents inside a scientific article are expressed by means of terms and syntactic structure. The terms are usually related to a certain thematic domain, and the syntactic structure is shown by the writer's knowledge and writing skills. Patterns in the syntactic structure may represent the authors' intentions to express important information. These patterns are assumed to be represented as frequent syntactic structures in a document and may include informative content. The utilization of these structure patterns in combination with other summarization techniques may provide better sentence selection.

Text in general is composed by the lexical part, referring to the meaning of the words, and by the syntactical part, corresponding to the arrangement of those words. For the extraction of information from texts, words are commonly used because they explicitly carry meaning, but the arrangement of words in sentences also represents meaning [4]. Scientific articles have particular use of

certain syntactical sequences due to their formal style, and these are in the form of phrases, collocations and expressions used by writers [5].

In this research we mainly focus on the document features of the scientific articles from the viewpoint of the text writing way, such as the uses of syntactic structures and document sections for the explanation of key topics in addition to looking upon words as content source. Our objective is to generate document summaries using these features. This paper is organized as follows. The next section describes some related works with respect to the use of the syntactic features of documents for summarization. Section 3 introduces our idea behind the selection of document features. Section 4 presents the summarization algorithm and Section 5 describes the experiment in comparison with other summarization methods. Finally, Section 6 concludes this paper.

2   Related Works

The arrangement of terms has been considered in summarization methods for example to locate collocations, such as lexical chains, although the concept of lexical chains differs from syntactic sequences in which the lexical chains are sequences of word meanings rather than the syntactic classification of those words [1,14]. Barzilay et al. use lexical chains for the sentence selection, frequent terms occurring in the lexical chains and sentence location in the text.

Most of the existing summarization methods, which use the syntactic classification of terms, employ them for expansion of term meanings, and this approach is usually combined with other features such as sentence position, term frequency, etc. Examples of these methods are typically proposed in the works [2,7]. Bawakid et al. use sentence selection based on the semantic similarity of sentences and also use parts of speech to find linguistic qualifiers which may indicate the importance of terms, as well as to expand term meanings with synonyms. No other syntactic information is used in their method. Liu et al. use the positions of words in the sentence along with parts of speech tagging to determine the most significant words in the text. The tagged terms are used to extend their definitions by employing external sources such as ontologies, and after term frequencies are obtained. They do not consider the syntactical features for the sentence ranking, although parts of speech are used.

More sophisticated uses of syntactic information are for rhetorical analysis of texts [12,16]. Teufel et al. uses document features such as position, section structure, term frequency, verb syntax, etc. among others to rank sentences based on their rhetorical status and relevance. Although a rhetoric analysis of documents can lead to approximate the discourse structure of the document better than simple syntactic analysis, this method requires manually annotated resources which are not always available.

3   Idea behind Selected Document Features

For the summarization process we have considered two main features of scientific articles. The first is based on the informative content of syntactic sequences and

the second is related to the document structure from a viewpoint which the distribution of key terms differ according to the document sections. We explain in more detail these features.

The original idea in the use of syntactic sequences was introduced by [6], which demonstrated that frequently occurring syntactic sequences contained within a corpus, captured the most informative content based on the least-effort consuming principle. Lioma referred to these syntactic sequences as parts of speech blocks or POS blocks, and employed them as part of an indexing method for information retrieval. In our research we consider the extraction of syntactic sequences from scientific articles based on a similar principle. Our main reason is that authors usually have their own writing styles, formed of common structures distributed throughout the text. These structures are used to convey or contextualize data which requires explanation, using understandable language.

POS blocks are obtained by replacing words with their syntactic information or parts of speech classification, but by retaining their sequential order in the sentence. Sequences can have variable length and their tokens may overlap. The extraction process is similar to n-grams extraction method where n tokens are defined as a sequence, and in the next iteration the starting position for the next sequence advances one token. Figure 1 shows POS blocks from a sentence (highlighted in yellow) with a length of four:

Sample sentence:

`[A visual representation of multi-layer design might clarify possibilities.]`

POS blocks from sample sentence:

A	visual	representation	of	multi-layer	design	might	clarify	possibilities
determiner	adjective	noun	preposition					
	adjective	noun	preposition	adjective				
		noun	preposition	adjective	noun			
			preposition	adjective	noun	modal		
				adjective	noun	modal	verb	
					noun	modal	verb	noun

Fig. 1. POS blocks of four tokens extracted from a sample sentence

The syntactic sequences observed frequently in scientific articles contain terms corresponding to domain-specific concepts or main topics explained in the article. The terms which form part of the syntactic sequences determine the amount of informative content. Table 1 shows fragments of text corresponding to two syntactic sequences, the underlined terms are part of the article's main topic.

The second feature of scientific articles is their use of sections to explain different topics. Researches have demonstrated that in certain sections of an article there are words that appear more frequently in relation to other sections [13].

Table 1. Terms appearing in frequent parts of speech sequences

determinant - noun - preposition - determinant - noun
the decision of the <u>HCRF</u> some <u>information</u> for the business the <u>segmentation</u> of the page the belief that the <u>structure</u> the text within the <u>vision</u>
noun - preposition - determinant - noun - noun
idea of the <u>HCRF algorithm</u> <u>content</u> of an <u>HTML</u> element text within the <u>vision</u> node procedure in the <u>SemiCRF model</u> loop in the <u>webpage understanding</u>

This is consistent with the idea that research papers use particular sections to explain topics. If these topics are explained using the authors' writing styles, then words from those topics may form part of frequent syntactic sequences. Combining these two concepts we can assume that frequent overlapping words from a section with frequent syntactic sequences, represents important information about the article's content. Therefore these document features may be used as a criteria for sentence selection in automatic summarization.

4   Summarization Algorithm

For the application of the summarization algorithm, the text is initially pre-processed by separating the text into sentences and assigning shallow parts of speech tags. The summarization algorithm consists of three main steps: First, divide the document according to sections and remove non-contextual text such as the reference list, Acknowledgements, etc.; Second, apply a keyword extraction method optimized for use on a single document based on word co-occurrence [8,15]; and Third, select the sentences which contain the k most frequent syntactic sequences. Algorithms 1, 2 and 3 describe the summarization process.

In the first step POS blocks with a specified length of n tagged terms are created. Each created sequence is stored if it has not already been stored before; otherwise a counter for that sequence is incremented. This process is repeated for all the terms in the text. Finally, the top k sequences with the highest counter value are selected.

In the second step, high scores are assigned to terms which occur frequently in each section but are also mentioned in other sections; this method is based on a chi square scoring function. Finally, the top m terms are selected, where m must be specified.

The last step is the combination of the previous two. From the selected sentences in step 1, sentences which contained the frequent terms per section

---

Algorithm 1. Step 1, frequent parts of speech sequence extraction.

Input: text, n, k
Output: top k frequent POS blocks
Preprocess Text
Tag terms with POS tags
Add term from text to buffer
while buffer not end **do**
    Take n POS tags
    for each tag **do**
        add tag to sequence
        if sequence exists **then**
            increment counter of existing sequence
        else
            create new sequence
Return top k sequences

Algorithm 2. Step 2, term ranking based on document sections.

Input: text, m
Output: top m ranked terms
for section in text **do**
    get all terms
    for terms in section **do**
        rank terms by section
Return top m terms

Algorithm 3. Step 3, sentence selection.

Input: top POS blocks, top ranked terms, t
Output: top t sentences with frequent POS blocks and ranked terms
for top POS blocks **do**
    find top ranked terms in POS blocks
    if sequence contains term **then**
        add sentence to result
Return top t sentences

(obtained in step two), are selected. The order of the result sentences is the same order in which they appear in the document.

5 Evaluation

In order to evaluate the syntactic sequence based summarization we initially processed a set of 20 scientific articles, removed the abstract and the bibliography, and then applied our summarizer to the documents. Our extractive summarization method selected in average 20% of the sentences from the documents. We applied five other extractive summarizers to the same text based on three types of summarization methods: baseline, frequency-based and multiple feature. The

two baseline summarizers are Baseline-Lead, which sequentially selects the first 20 % of sentences from the source text, and Baseline-Random, which randomly selects 20 % of the sentences in the source text. The frequency-based summarizer Open Text Summarizer (OTS) [11] uses frequently-used words to give a higher score to sentences. SweSum [3] and MEAD use multiple features to identify sentences [10]. SweSum is a summarizer for texts written in Swedish or English, and uses features such as sentence position and numerical data identification. MEAD is a single/multi-document summarizer which uses features such as position of sentence within the text, first sentence overlap, sentence length, and also a centroid method based on a cluster of related documents. The criteria for the selection of the summarizers was based on their availability.

We compared the resulting summarizations with the original abstract of the paper. The comparison was done using content-based scoring methods to measure lexical similarity. The six used methods were: Simple cosine, which calculates the cosine similarity with a simple binary count if the word exists in the document; Cosine, which uses the inverse document frequency weights and includes the actual count of words; B-Overlap and T-Overlap, which measure the bigram and trigram overlap respectively; Norm LCS, which measures the normalized longest common substring and finally Bleu, which uses closeness metrics based on n-grams. Bleu was originally developed to evaluate translated texts [9]. The similarity scores are shown in Table 2.

Table 2. Similarity scores for the different summarization methods

Summarizer	amp; S-cos	amp; Cos	amp; T-overlap	amp; B-overlap	amp; Norm LCS	amp; Bleu
10FreqPOS4Seq	amp; 0.27658	amp; 0.40542	amp; 0.14568	amp; 0.03714	amp; 0.16833	amp; 0.03274
10FreqPOS5Seq	amp; 0.28197	amp; 0.41151	amp; 0.14973	amp; 0.03588	amp; 0.16609	amp; 0.03201
20FreqPOS5Seq	amp; 0.27893	amp; 0.38946	amp; 0.14781	amp; 0.03900	amp; 0.16613	amp; 0.03533
10FreqPOS6Seq	amp; 0.27766	amp; 0.34397	amp; 0.15085	amp; 0.03839	amp; 0.16109	amp; 0.04187
20FreqPOS6Seq	amp; 0.28538	amp; 0.38082	amp; 0.15298	amp; 0.03824	amp; 0.16710	amp; 0.03751
20FreqPOS7Seq	amp; 0.26616	amp; 0.33294	amp; 0.14599	amp; 0.03513	amp; 0.15578	amp; 0.03357
10TermPerSect	amp; 0.27308	amp; 0.39008	amp; 0.14370	amp; 0.03735	amp; 0.16722	amp; 0.03434
Mead	amp; 0.31768	amp; 0.43504	amp; 0.16140	amp; 0.05234	amp; 0.16823	amp; 0.04769
Ots	amp; 0.24362	amp; 0.21079	amp; 0.12015	amp; 0.02737	amp; 0.12258	amp; 0.02184
SweSum	amp; 0.28369	amp; 0.45559	amp; 0.15350	amp; 0.04123	amp; 0.17585	amp; 0.03274
Baseline-Lead	amp; 0.33079	amp; 0.42060	amp; 0.18011	amp; 0.06063	amp; 0.19356	amp; 0.06444
Baseline-Random	amp; 0.27817	amp; 0.38885	amp; 0.14762	amp; 0.03531	amp; 0.16407	amp; 0.03658

For the first six summarizations shown in Table 2 the proposed algorithm was used. Our algorithm requires the parameter for the number of top k key words per section. In this experiment setup a value of 10 was used. Also values for the length of the POS blocks were specified from four to seven. Finally the number of top k frequent POS blocks was set to ten and twenty. From Table 2, the first six rows correspond to different combinations of the POS block length values and the top k frequency values. In the case of the first row, the top ten frequent sequences of length four was used; from the second row, the top ten frequent sequences of length five, and so on. The row labeled TermPerSect corresponds

to the summarization where only the term ranking score was used in order to compare the summarization based only on frequent terms per section without considering frequent POS blocks. For the different configurations of POS block lengths and top k selected sentence values, the scores are similar among the scoring methods with exception of the Cosine score where there is a variation. In all the other scores, the selection of top twenty frequent POS blocks of length six has a slight improvement over the other settings in our algorithm. Comparing our summarization method with the additional methods, our summarizer scored better than the OTS and Baseline Random, but had a low score with respect to Mead, SweSum and Baseline-Lead.

6   Conclusion

We have proposed a method of single document summarization applied to research papers considering mainly two document features: syntactic sequences and term frequency in document sections. One of the advantages of this method is that it does not require external sources. This method may also be applied for multi-language summarization. Although the scoring results of the experiment did not reveal much improvement over other existing methods, it can be further enhanced by using advance statistics in the ranking of syntactic sequences and in combination with other document features. It is also necessary to evaluate the summarization using other standardized summarization evaluation methods.

References

1. Barzilay, R., Elhadad, M.: Using Lexical Chains for Text Summarization. In: Proceedings of the Intelligent Scalable Text Summarization Workshop, Madrid, pp. 10–17 (1997)
2. Bawakid, A., Oussalah, M.: A semantic summarization system: University of Birmingham at TAC 2008. In: Proceedings of the First Text Analysis Conference, Maryland, USA (2008)
3. Hercules Dalianis. SweSum - A Text Summarizer for Swedish. Technical report, NADA, KTH, Stockholm (2000),
   http://people.dsv.su.se/hercules/papers/Textsumsummary.html
4. Hunston, S.: Starting with the small words Patterns, lexis and semantic sequences. International journal of corpus linguistics 13(3), 271–295 (2008)
5. Gledhill, C.J.: Collocations in science writing. Narr Verlag, Tübingen (2000)
6. Lioma, C., Ounis, I.: A syntactically-based query reformulation technique for information retrieval. Information Processing and Management: an International Journal 44(1), 143–162 (2008)
7. Liu, X., Webster, J.J., Kit, C.: An Extractive Text Summarizer Based on Significant Words. In: Proceedings of the 22nd International Conference on Computer Processing of Oriental Languages. Language Technology for the Knowledge-based Economy, pp. 168–178. Springer, Heidelberg (2009)
8. Matsuo, Y., Ishizuka, M.: Keyword extraction from a single document using word co-occurrence statistical information. International Journal on Artificial Intelligence Tools 13(1), 157–169 (2004)

9. Papineni, K., Roukos, S., Ward, T., Zhu, W.-J.: BLEU: a method for automatic evaluation of machine translation. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, pp. 311–318. Association for Computational Linguistics, Stroudsburg (2002)

10. Radev, D., Allison, T., Blair-Goldensohn, S., Blitzer, J., Çelebi, A., Dimitrov, S., Drabek, E., Hakim, A., Lam, W., Liu, D., Otterbacher, J., Qi, H., Saggion, H., Teufel, S., Topper, M., Winkel, A., Zhang, Z.: MEAD - a platform for multidocument multilingual text summarization. In: Proceedings of the 4th International Conference on Language Resources and Evaluation, Lisbon, Portugal (2004)

11. Nadav Rotem. Open Text Summarizer, http://libots.sourceforge.net/

12. Saggion, H., Lapalme, G.: Generating indicative-informative summaries with sumUM. Computational Linguistics 28(4), 497–526 (2002)

13. Schuemie, M.J., Weeber, M., Schijvenaars, B.A., van Mulligen, E.M., Christiaan, C., van der Eijk, Jelier, R., Mons, B., Kors, J.A.: Distribution of information in biomedical abstracts and full-text publications.. Bioinformatics 20(16), 2597–2604 (2004)

14. Silber, H.G., McCoy, K.F.: Efficient text summarization using lexical chains. In: Proceedings of the 5th International Conference on Intelligent user Interfaces, pp. 252–255. ACM Press, New York (2000)

15. Strobelt, H., Oelke, D., Rohrdantz, C., Stoffel, A., Keim, D., Deussen, O.: Document Cards: A Top Trumps Visualization for Documents. IEEE Transactions on Visualization and Computer Graphics 15(6), 1145–1152 (2009)

16. Teufel, S., Moens, M.: Summarizing scientific articles: experiments with relevance and rhetorical status. Computational Linguistics 28(4), 409–445 (2002)

Automatic Music Genre Classification Using Hybrid Genetic Algorithms

George V. Karkavitsas[1] and George A. Tsihrintzis[2]

Advantage S.E
[1] Financial System Experts, Alimos 174 55, Greece
George.Karkavitsas@fiserv.com
[2] University of Piraeus, Department of Informatics
Piraeus 185 34, Greece
geoatsi@unipi.gr

Abstract. This paper aims at developing an *Automatic Music Genre Classification* system and focuses on calculating algorithms that (ideally) can predict the music class in which a music file belongs. The proposed system is based on techniques from the fields of Signal Processing, Pattern Recognition, and Information Retrieval, as well as Heuristic Optimization Methods. One thousand music files are used for training and validating the classification system. These files are distributed equally in ten classes. From each file, eighty one (81) features are extracted and used to create 81 similarity matrices. These 81 similarity matrices constitute the training instances. During the training phase, feature selection takes place via a modified hybrid Genetic Algorithm in order to improve the class discrimination clarity and reduce the calculating cost. In this algorithm, the crossover probability is replaced by a parent pair number that produces new solutions via a taboo list. Also, an adaptive mutation, an adaptive local exhaustive search and an adaptive replace strategy are used, depending on whether the system has reached a local extreme. Local exhaustive search takes place in the most optimal up to the current solution neighboring chromosomes. The Genetic Algorithm fitness function constitutes a weighted nearest neighbors classifier. Thus, the chromosome fitness is proportional to the classifier accuracy that the chromosome creates. During the classification phase, the features selected via the Genetic Algorithm create an adjusted nearest neighbor classifier that performs the classifications. From each new music file pending classification, selected features are extracted and then compared with the corresponding features of the database music files. The music file is assigned to the class indicated by the k nearest music files.

Keywords: Signal Processing, Pattern Recognition, Music Information Retrieval, Music Genre Classification, Hybrid Genetic Algorithm, k-nearest neighbors Classifier, multi-class classification.

1 Introduction

Automatic audio signal classification constitutes a research area that has received much attention over the last years. It is considered a keystone of Music Information

Retrieval [1,5,6,7,8]. The music can be classified into categories based on genre, performer, composer, culture that represents, emotion that causes to the user [2] and - generally speaking- based on any separation can be useful to humans. Musical genres are labels created and used by humans for categorizing and describing the vast universe of music [3]. A content based classification, constitutes difficult problem for both computers and humans, since very rarely there is accuracy and clarity in musical qualities and features of each class [3]. Even though boundaries of classes can be very fuzzy, it is clear that the members of a particular genre share certain characteristics. The applications of automatic music genre classification are many and are extended in both academic and commercial level [4,2]. The present paper deals with *Automatic Music Genre Classification* and focuses on computational algorithms that can (ideally) predict in which class, a music track belongs.

Pattern Recognition and Machine Learning techniques help in dealing music genre classification problems [9]. A typical Pattern Recognition System usually consists of three steps and in fact it is a system of supervised machine learning. The three steps of Pattern Recognition system is data acquisition, feature extraction and classification. In training mode two more steps are added, the feature selection and system evaluation. In a wider sense data acquisition and feature extraction constitute the process of recognition. Classification refers to a conversion between features (F) and classes (C) [10]. Therefore pattern recognition can be described as a conversion from the area of measurement (M) in the feature area (F) and finally in the Decision area (D) : $M \rightarrow F \rightarrow D$. In the present work the conversion between features and classes is done by an adjusted k-NN classifier [11].

In most real problems, features can be noisy and many of them not relative to the particular objective function that is sought to approach. In addition it is very likely features to indicate overlapping classes. In order to simulate such systems, a feature selection takes place with probabilistic approach. The problem of feature selection can be considered as an optimization problem.

Genetic Algorithms (GA) constitute a group of probabilistic (stochastic) search and optimization techniques proposed (initially) by J. Holland at the University of Michigan in 1975 [8,13,10]. Gas follows two basic rules, the passage of information from one generation to the next (inheritance) and survival of the fittest. The GA is a seeking process of the optimal solution, through a set of solutions that are candidates for a problem. The set of all possible solutions called population. The candidate solutions which can be selected to solve a problem called chromosomes. Each chromosome consists of a construction set elements called genes. Each "parent" generation is evaluated through an evaluation function that yields a "grade". The next generation is created so as to get a higher grade, i.e. to represent a better solution [15]. In the following lines a modified hybrid GA is analyzed.

2   Musical Surface Features – Feature Extraction

Each natural entity allocates a set of features. Each feature constitutes an entity's compact representation. Thus the environment, based on this set of features, can describe, recognize or even compare the entity. Such a set of features allocate also acoustic signals. In the substance, acoustic signal feature can be any mathematical

model which represents the acoustic signal. A set of 81 features are presented to the algorithm, so the algorithm can choose those features who contributes more in the classes' segregation.

Three types of features are used: one value feature, one dimension (vector) feature and two dimension (matrix) feature. The distance between two features is measured with the City block metric.

The following features are calculated for each "analysis" window [16] of 20msec with 10msec window step (220 samples with 110 samples overlap at 11025 sampling rate). The frequency domain features calculation is based on the Short Time Fourier Transform (STFT) that can be efficiently calculated using the Fast Fourier Transform (FFT) algorithm [16].

These features are: Short-time Energy (Time domain feature), Zero Crossing Rate (Noise in signal)(Time domain feature), Spectral Centroid (Spectral Brightness)(Frequency domain feature), Spectral Rolloff (Spectral Shape) (Frequency domain feature), Spectral Flux (Spectral Change)(Frequency domain feature), Spectral Entropy (Frequency domain feature).

Six statistical moments [17] (Median value, standard derivation divided by mean value, mean value, standard derivation, max value divided by median value, max value) of these features are calculated over a "texture" window [16] of 1 second consisting of 100 "analysis" windows. Thus, thirty six features are created where each of these consists of a thirty elements vector. Two acoustic signals can belong in the same class, but this does not mean necessarily that there is simultaneous presence of similar events, among them. For this reason, it is also extracted the mean value of these vectors, generating thirty six features of one value, giving weight to the quantity of different events rather than the time instance they were took place.

RMS energy feature (signal Loudness [18]) is also extracted from the signal. The Energy RMS is calculated for each "analysis" window of 50msec with 25msec window step. The mean value of these analysis windows constitutes the -one value- Energy RMS feature. In addition, the mean value of energy RMS analysis windows that their values is lower than the energy RMS value of the corresponding texture window, constitutes the -one value- Low energy feature.

The following three features focus on aspects related to timbre and periodicities in the signal. They are matrix features and they are based on a sone / bark scale (a perceptual scale which groups frequencies to critical bands according to perceptive pitch regions) representation of the audio signal. Spectrum Histograms [19] are a simple approach to summarize the spectral shape. The SHs describe a piece by counting how many times each loudness level was exceeded in each frequency band. SHs feature calculated for each "analysis" window of 20msec, with 10msec window step, for 25 loudness levels histogram analysis and for 20 critical frequency bands. Periodicity Histograms were originally presented in the context of beat tracking [20]. The idea is to describe periodically reoccurring beats. PHs feature calculated for each "analysis" window of 46,4msec, with 23,2msec window step, for 60 modified frequencies and for 20 critical frequency bands. Fluctuation Patterns are an alternative approach to describe periodicities. The main difference between FPs [21] and PHs is that the FPs includes information on the energy distribution in the frequency spectrum which the PHs discards. They are calculated in the same manner as the PHs.

Statistical Spectrum Descriptors [22] are derived from a psycho-acoustically transformed Bark-scale spectrogram and comprise several statistical moments, which are intended to describe fluctuations on a number of critical frequency bands. The spectrogram is computed using the short time Fast Fourier Transform (STFT) (window size 23msec and 50 % overlap). The Bark scale (24 critical bands [23]) is applied to the spectrogram and The Bark scale spectrogram is then transformed into the decibel scale. Subsequently, the values are transformed into Sone values, in order to approximate the loudness sensation of the human auditory system. From this representation of a segment's spectrogram the following statistical moments are computed in order to describe fluctuations within the critical bands: mean, median, variance, skew-ness, kurtosis, min- and max-value are computed for each critical band, forming the SSD feature vector of 168 elements. Also another version of SSD feature is used. This –one value- feature for an audio file is calculated as the median of the SSD feature.

The last two features are based on chromagram. A non-overlapping analysis window of 100msec is used, and the mean over a texture window of 1 second is calculated. These calculations creates a 12 rows (notes) and 30 columns (seconds) matrix. The first -matrix- feature is constituted by the sorting version of this matrix, giving weight in the frequency of each event. The second -vector- feature is constituted by the mean of 30 seconds for each note.

3   Feature Selection

3.1   Feature Weighting

Assigning variable weights to the feature instances before applying the kNN classifier, distorts the space and modifying the importance of each feature to reflect its relevance for classification. In this way, similarity with respect to important feature becomes more critical than similarity with respect to irrelevant features. Therefore one weight value corresponds to each feature $F = \{w_1 \cdot f_1, w_2 \cdot f_2, .., w_n \cdot f_n\}$. This method is called *feature weighting* [24]. Thus, it can be said that the system is looking for the most robust feature analogy which makes more clear the class discrimination clarity. In 1991, James Kelly and Lawrence Davis [24] developed a hybrid Genetic Algorithms weighted k-nearest neighbors for analogy selection of features with very good results. In feature selection method every time a similarity matrix weighted by different features each time should be calculated and graded based on classification accuracy. In terms of time, it might be better to have similarity matrices for each feature separately instead of the initial features expression. Then, adding individual features similarity matrices with combination that indicated by the each chromosome, the classifier similarity matrix is resulted. In the present paper the City Block metric is used. Of course, other metrics suggest different ways of combining and calculating the classifier similarity matrix.

3.2   The Proposed Genetic Algorithm Approach

The flow chart of the proposed Genetic Algorithm approach is given in Figure 1.
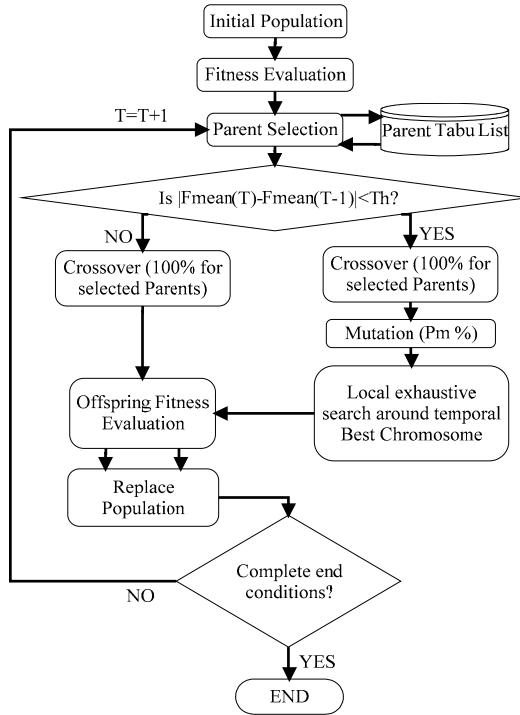
Fig. 1. Flowchart of the proposed Genetic Algorithm

3.2.1 Chromosomes

In the feature selection problem, there is a pool of features L where some of them should be selected, while others should be dismissed. Therefore a binary vector with length equal to the number of features can be used as a candidate solution's representation

$$Chromosome = [g_1, g_2, .., g_L],$$ (1)

where $g \in \{0,1\}$ and L the number of features. Given the binary encoding there are 2^L different candidate solutions.

3.2.2 Fitness Function

Critical for the success of the genetic algorithm is the choice of the evaluation (fitness) function [15]. Indeed this is the only means of communication between the genetic evolutionary process and its environment (i.e., the problem it seeks to solve). When chromosomes of the current generation are graded by the fitness function, the genetic algorithm gains feedback from the environment so as to adjust its evolution towards an improved next generation. For the problem at hand, the chromosomes fitness evaluation is done by using an adjusted k-NN nearest neighbor classifier.

In the feature selection phase, every feature is replaced by a coresponding similarity matrix of the input entities who calculated based on this particular feature. So if initially there were L features $F = \{f_i\}$, $i = 1...L$ for the selection phase, now there are L similarity matrices $SM = \{sm_i\}$, $i = 1...L$ created by these features. A weight via the L gene of chromosome is applied in each one of the L similarity matricies (the value of each gene corresponds to a specific similarity matrix and hence to a specific feature). Then, all similarity matrices are added (depending on the metric that is used) generating the classifier's similarity matrix (Eq. 2), who is normalized by his maximum value.

$$SM_{classifier} = \sum_{i=1}^{L}(sm_i \cdot g_i) \qquad (2)$$

Since it has been produced a matrix showing the correlations between all entities of the database, can be calculated easily and quickly in which class belongs each entity using the method leave one out cross validation and even for different number of nearest neighbors. For different values of k nearest neighbors, the classifier's accuracy is calculated and the k that gives the best accuracy is held. Since this work is an academic reference we choose not to integrated number k nearest neighbors as a gene on the chromosome (i.e automatic selection), but (due to its small size) to investigate it exhaustively. Thus can be said that the best chromosome is considered the chromosome that indicates such a feature selection, based in which the classifier achieves maximum accuracy. Lower fitness is better fitness (Eq. 3).

$$fitness = 1 - accuracy \qquad (3)$$

3.2.3 Parent Selection

In each generation, P_p parent pairs are selected for the offspring production via the roulette-wheel selection rule [12]. Each parent "a" can be selected again (in the same generation) to produce descendants but not with the same parent "b". This achieved via a Tabu list which is writing all the parent pairs that are selected in one generation. Thus, up to $P_p = N \cdot (N-1)$ parent pairs can be selected in each generation, where N is the chromosomes number in the population. In such a case, more suitable chromosomes are free to crossover their genetic material with more than one other chromosomes. Also, it is possible to be selected such a P_p, so that the offspring to the next generation is far more than the fixed population (Overpopulation).

3.2.4 Genetic Operations

The crossover probability is repealed, as the crossover rate is determined by the number P_p of breeding parent pairs mentioned in the above section. In this work two types of crossover were used, multi-point crossover and uniform crossover [25]. With these two aspects of crossover used, the system ensures that produced offspring in the case of multi-point crossover, containing large contiguous segments from each parent, while in the case of a uniform crossover, the produced offspring containing smaller and more alternating segments of each parent. Each crossover operation produces up to four descendants.

Mutation is a genetic operation that introduces the concept of variable in the chromosome. This variable can be global or local, and intended to prevent the entire set of solutions in the population reach a local extreme. So, this genetic operation is enabled only in the case which system reaches a local extreme before the termination condition fulfilled. The criterion to decide that the solutions of the population are very similar to each other, so it is difficult for the current generation to provide better solutions to future generations, is the mean fitness of the entire population. When mean fitness of the population remains constant (or less than a threshold value) between two consecutive generations, mutant function is activated. In essence it is adaptive mutation (Whitley and Starkweather) [26], where instead of a constant mutation rate, it varies depending the genetic homogeneity of the population. Provided that mutation activation is so rare, it is advisable to apply a relatively high mutation probability in the descendants let say 50%.

3.2.5 Local Exhaustive Search

A local exhaustive search technique is implemented if the system reached a local extreme before meeting the termination condition. In such a case the neighboring chromosomes of the most optimal up to then solution are examined as far as their fitness. Chromosomes that present equal or better fitness than the fitness of the most optimal up to then chromosome are passed on population replacement process.

3.2.6 Replacement Strategy

Chromosomes N population in each generation T remains constant. Once the descendants are produced, an adaptive replacement strategy is applied. *Case1*: During a generation where the system has not reached a local extreme. The pool of candidates for the next generation consists of the initial chromosomes and their descendants via the crossover operation. The best N chromosomes directly copied in the next generation. Thus, without mutation the system achieves a fairly rapid convergence towards a "good" extreme. *Case2*: During a generation where the system has reached a local extreme. The pool of candidates for the next generation consists of the initial chromosomes, their offspring via the crossover and mutation operation and the chromosomes that collected from the local exhaustive search. In the next generation passes a small number (elite) from the best initial chromosomes, while the remainder population is complemented by the best offspring chromosomes and those that collected from the local search.

4 Experimental Part

4.1 Training Phase

For training and validation the system is using a database that contains one thousand (1000) music files, thirty (30) seconds each. Audio files are divided equally into ten (10) classes {blues, classical, country, disco, hiphop, jazz, metal, pop, reggae, rock}, i.e. each class contains one hundred (100) music files. (Tzanetakis Music Collection [3], 10 Genres). Sampling frequency of all music samples was $Fs = 11025 Hz$. From each music track 81 features are extracted and then are converted to 81 similarity

matrices. Each similarity matrix is representing the correlation between the 1000 music files for a specific feature. These 81 similarity matrices constitute the training instances of an adjusted k-nearest neighbors classifier that the modified GA uses for rating the chromosomes. The classifier is evaluated as for his accuracy with leave one out cross validation method. Every time 999 music tracks used for training and 1 for classification. The class of music track that pending for classification is calculated using the k smallest equalized distance (nearest neighbors) from the others music tracks, as indicated by the white line in figure 2(a). The process is repeated until all tracks were used to train and classify (1000 loops for 1000 tracks). The process total error percentage is calculated as the mean of all experiments (Eq. 4).

$$E = \frac{1}{1000} \cdot \sum_{i=1}^{1000} E_i \tag{4}$$

Mean accuracy of classes constitutes the classifier's accuracy. Figure 2(d) shows the accuracy of k-NN classifier created using all 81 features, for various values of K nearest neighbors.

Both the efficiency and the success of the Genetic algorithm feature selection method depend critically on a number of parameters, like Th, T, N, Pm, Pp, Ne, Nd. For parameters choice N = 80 chromosomes, T = 30 generations, Ne = 10 number of elit chromosomes, P_p = 30 number of parent pairs, Pm = 0.5 mutation propability, Th=0.15 threshold decision whether the system has reached local extreme and Nd=3 produce descendants in each genetic operation (two of the very multi-crossover and the first uniform), Genetic algorithm converges quite quickly to the optimal solution (global minimum of the fitness function), from a set of 2^{81} candidate solutions (Figure 2(b)).

Figure 2(b) shows the best chromosome fitness (red color) in each generation (or differently the classification error (%) of the classifier produced by the best chromosome in each generation) and mean fitness (blue color) of all chromosomes in each generation. In generations 12, 17 and 20 there is a sharp increase in the mean fitness of chromosomes. This happened because the system reached a local extreme and to avoid it created several "controlled" randomized solutions probably with more useful individual genetic material, but with worse fitness.

The modified hybrid Genetic Algorithm choose that a combination of 36 features provides better classification accuracy for a 16 nearest neighbors classifier. Figure 2(c) shows accuracies of the K-NN classifier for different neigbours number K, which was created using the optimal number of features as selected by the GA (Maximum accuracy 67.6% for K = 16). The similarity matrix resulting from these 36 features is given in figure 2(a).

With base the 36 features that were selected, Figure 3 shows the estimation and mapping of the music tracks positions (as entities) in three dimensional space. Some classes seems to be moving to a different level and do not overlap with others, such as classical music, which is spatial rather far from the "Pop" music. This is also illustrated in the confusion matrix (Table 1). However there are many classes that their contents are placed very close, even after feature selection, thus decreasing the accuracy of the classifier.
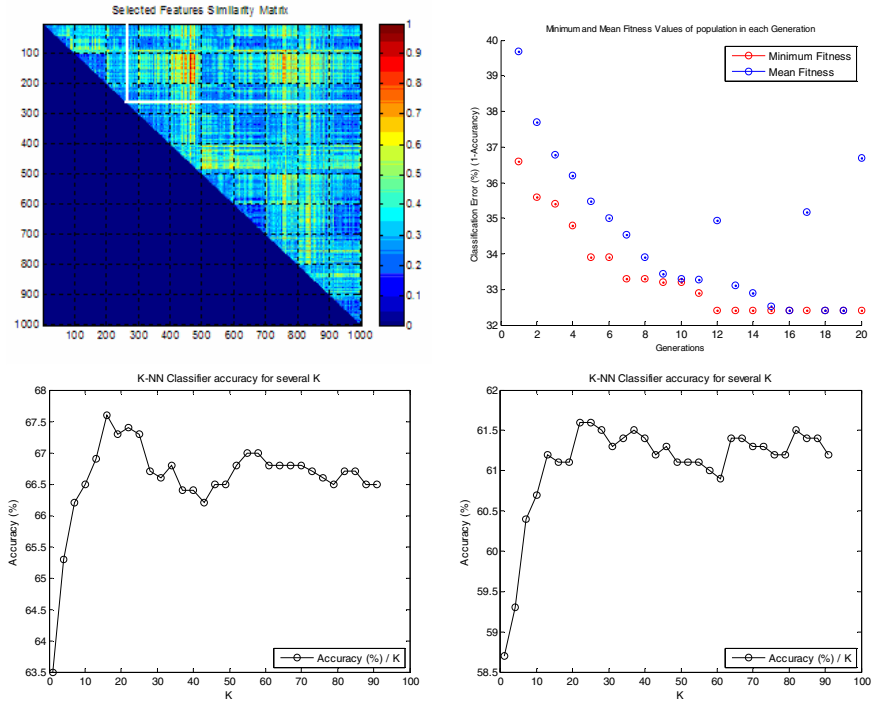
Fig. 2. (a) Resulting similarity matrix of the 36 features selected by the system to create the K-NN classifier, White color indicates the path that an entity follows to be compared with all others entities, (b) Best chromosome fitness in each generation and mean fitness of all chromosomes in each generation, (c) K-NN classifier accuracy, created using the optimal number of features selected by the GA (Maximum accuracy 67.6% for K = 16), (d) K-NN classifier accuracy, created using all 81 features (61.6% Maximum accuracy for K = 22)
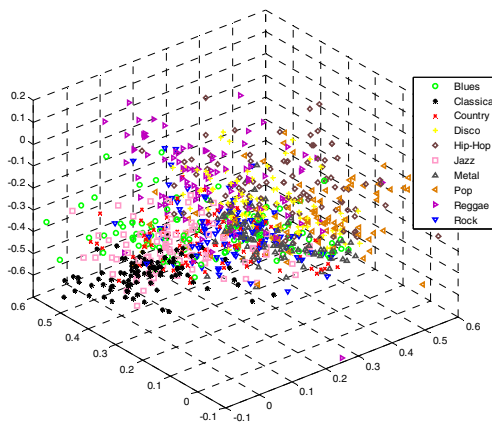
Fig. 3. Estimation and mapping of the music tracks positions (as entities) in three dimensional space.

Table 1. Classifier's Confusion Matrix for 16 nearest neighbors

		Confusion Matrix								
		PREDICTED CLASSES								
		Blues	Classical	Country	Disco	Hiphop	Jazz	Metal	Pop	Reggae	Rock
A C T U A L	Blues	68	1	7	3	0	5	3	3	5	5
	Classical	1	91	3	0	0	3	0	0	0	2
	Country	4	1	62	4	0	14	3	2	5	5
	Disco	2	1	3	68	1	0	4	9	3	9
	Hiphop	3	0	0	14	60	1	1	11	9	1
C L A S S E S	Jazz	3	11	5	1	0	77	3	0	0	0
	Metal	1	0	1	5	1	1	83	0	0	8
	Pop	5	1	5	10	6	2	1	59	4	7
	Reggae	2	0	7	7	8	0	1	3	69	3
	Rock	4	1	21	11	0	4	14	2	4	39

4.2   Confusion Matrix and Evaluation Metrics

The classifier's Confusion Matrix is given in table 1. The columns correspond to the actual genre and the rows to the predicted genre. The percentages of correct classifications lie in the diagonal of the confusion matrix. The best predicted genres are classical and metal while the worst predicted are pop and rock. This is due to the fact that the pop and rock are very broad categories and their boundaries are fuzziest than classical or metal.

Analyzing further the system's classifier, graphical representations of precision (degree of soundness, diagonal of the confusion matrix) [27] and recall (degree of completeness) [27] for all the classes that produced (figure 4).
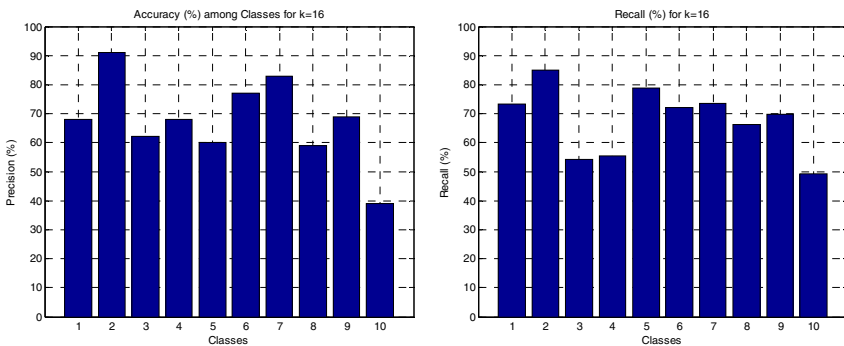
Fig. 4. (a) Accuracy, (b) Recall among classes for 16 nearest neighbors classifier.

4.3   Classification Phase

Once completed the training phase, when the system deciding which features will be used for the classifier creation, the control passes to the classification phase. From

each new to classification music file the selected features are extracted and they compared with the corresponding features of the database music files. A sixteen nearest neighbor classifier decides the new music file's class.

5   Comparison with Relative Works

In international literature two references were found which using the same music database for training the classifier [3], [28].

Table 2. Comparison Table of classification systems for the same musical database (10 classes classification).

	Tzanetakis G. (2002)	Li T. et al (2003)	Karkavitsas G. (2010)
Sampling Frequency	22050 Hz	22050 Hz	11025 Hz
Features	30 Features without DWCH	30 Features with DWCH	36 from 81 Features without DWCH
Feature Selection	No	No	Yes
Training Instances	Features	Features	Features Similarity Matrices
Evaluation	10-fold CV	10-fold CV	Leave one out CV
Classifier	Gaussian Mixture Model	Gaussian Mixture Model	k-NN
Accuracy	61%	80%	67,6% (k=16)

6   Summary – Conclusions – Future Work

Accuracy of classification systems varies and depends on several parameters. Key elements that cause variations in a music gerne classification system accuracy is the quantity and diversity of music tracks used in the training phase and quantity and diversity of music classes that they belong. Another key element is the quality of the features that are extracted from the music tracks. These elements are giving shape and details in music surface. The first two key elements are constant and fixed in most music genre classification problems, since they actually constitute the problem. Features selection not only adds quality to features set by removing noisy music data and fuzzy features that make musical classes separation less distinct, but also reduces computational cost during the classification phase. Feature selection is a quite chaotic process, which means that adding a new feature in the original feature set may actually increase the classification accuracy, but it is likely this to happen for a very different mix of features. The basic idea is that the more features are available to a feature selection system, the better quality features it will choose and thus will create greater accuracy classifier. By using similarity matrices as training instances the system saves computational cost during the training phase. The modified GA was

developed exclusively for this system and managed to cope satisfactorily in both best solution selection and fast convergence.

In the present work a kind of quantitative feature selection (selection or non-selection) was examined. A more detailed research would be a more qualitative feature selection. That is to say the features attendance in the classification with integer weights (instead of binaries) applied in each one of them. In this type of selection, the music surface would be more detailed because more features with more options may be selected but this will increase a lot the computational cost.

Computational cost constitutes a great issue in music genre classification problems. Key reason for choosing Genetic Algorithms with k-nearest neighbors classifier for fitness evaluation and similarity matrices for training instances was their easy implementation to parallel programming. Parallel programming will give to the system a quite solid and detailed music surface so the system will provide a much better solution in much less time.

References

[1] Muller, M.: Information Retrieval for Music and Motion. Springer, Heidelberg (2007); ISBN: 978-3-540-74047-6

[2] Xiao, H., Stephen Downie, J.: Exploring mood metadata: Relationships with genre, artist and usage metadata. In: Eighth International Conference on Music Information Retrieval, Vienna (2007)

[3] Tzanetakis, G., Cook, P.: Musical Genre Classification of Audio Signals. IEEE Transactions on speech and audio processing 10 (2002)

[4] Scaringella, N., Zoia, G., Mlynek, D.: Automatic genre classification of music content: a survey. Signal Processing Magazine (2006)

[5] Sotiropoulos, N., Lampropoulos, A.S., Tsihrintzis, G.A.: MUSIPER: a system for modeling music similarity perception based on objective feature subset selection. In: User Modeling and User-Adapted Interaction, vol. 18, pp. 315–348 (2008)

[6] Lampropoulou, P.S., Lampropoulos, A.S., Tsihrintzis, G.A.: Intelligent Mobile Content-based Retrieval from Digital Music Libraries. In: Intelligent Decision Technologies, vol. l. 3, pp. 123–138. IOS Press, Amsterdam (2009)

[7] Lampropoulou, P.S., Lampropoulos, A.S., Tsihrintzis, G.A.: Music Genre Classification based on Ensemble of Signals produced by Source Separation Methods. In: Intelligent Decision Technologies, vol. 4, pp. 229–237. IOS Press, Amsterdam (2010)

[8] Lampropoulos, A.S., Lampropoulou, P.S., Tsihrintzis, G.A.: A Cascade-Hybrid Music Recommender System for Mobile Services based on Musical Genre Classification and Personality Diagnosis. In: Multimedia Tools and Applications (to appear in February 2011)

[9] Duda, R., Hart, P., Stork, D.: Pattern classification. John Wiley & Sons, New York (2000)

[10] Bandyopadhyay, S.: Classification and Learning Using Genetic Algorithms. Springer, Heidelberg (2007); ISBN-13 978-3-540-49606-9

[11] Dasarathy, B.V.: Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques (1991); ISBN 0-8186-8930-7

[12] Holland, J.H.: Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence. University of Michigan Press, Ann Arbor (1975)

[13] Sivanandam, S.N., Deepa, S.N.: Introduction to Genetic Algorithms. Springer, Heidelberg (2007); ISBN 978-3-540-73189-4

[14] Tang, K.S., Man, K.F., Kwong, S., He, Q.: Genetic Algorithms and their Applications. IEEE Signal Processing Magazine, November 96, 1053–1088 (1996)

[15] Karkavitsas, G., Rangoussi, M.: Object localization in medical images using Genetic Algorithms. International Journal of Signal Processing 1(3), 204–207 (2004); ISSN:1304-4494

[16] Tzanetakis, G.: Manipulation, Analysis and Retrieval systems for audio signals. phd dissertation (2002)

[17] Lampropoulos, A.S.: Machine Learning-based Recommendation Methods for Multimedia Data. PhD Dissertation, The University of Piraeus, Piraeus, Greece (2010)

[18] Lartillot, O.: MIR toolbox 1.3", Finnish Centre of Exce!ence in Interdisciplinary Music Research. University of Jyväskylä, Finland (2010)

[19] Pampalk, E., Dixon, S., Widmer, G.: Exploring music collections by browsing different views. In: Proc of ISMIR (2003)

[20] Scheirer, E.D.: Tempo and beat analysis of acoustic musical signals. JASA 103(1) (1998)

[21] Pampalk, E., Rauber, A., Merkl, D.: Content-based organization and visualization of music archives. In: Proc of ACM Multimedia (2002)

[22] Lidy, T., Rauber, A.: Evaluation of feature extractors and psycho-acoustic transformations for music genre classification. In: Proceedings of the 6th International Conference on Music Information Retrieval, London, UK, pp. 34–41 (2005)

[23] Zwicker, E., Fastl, H.: Psychoacoustics - Facts and Models. Springer Series of Information Sciences, vol. 22. Springer, Berlin (1999)

[24] Kelly, J.D., Davis, L.: A Hybrid Genetic Algorithm for Classification (1991)

[25] Syswerda, G.: Uniform crossover in genetic algorithms. In: Schafier, J.D. (ed.) Proceedings of the 3rd International Conference on Genetic Algorithms. Morgan Kaufmann, San Mateo (1989)

[26] Whitley, D., Starkweather, T.: GENITOR-II: A distributed genetic algorithm. Journal of Experimental and Theoretical Artificial Intelligence 2, 189–214 (1990)

[27] Baeza-Yates, R., Ribeiro-Neto, B.: Modern Information Retrieval. ACM Press, Addison-Wesley (1999) ISBN 0-201-39829-X

[28] Li, T., Ogihara, M., Li, Q.: A comparative study on content based music genre classification. In: 26th annual international ACM SIGIR Conference on Research and Development in Information Retrieval, Toronto, Canada (2003)

A Remotely Accessible Exercise System for Network Security Based on an Automatic Cracking Function in a Virtual Machine Network

Yuichiro Tateiwa[1], Tomohiro Iwasaki[2], and Takami Yasuda[2]

[1] Graduate School of Engineering, Nagoya Institute of Technology,
Gokiso-cho, Showa-ku, Nagoya-shi 466-8555, Japan
`tateiwa@nitech.ac.jp`
[2] Graduate School of Information Science, Nagoya University, Furo-cho,
Chikusa-ku, Nagoya-shi 464-8601, Japan
`t-iwasaki@nagoya-u.jp, yasuda@is.nagoya-u.ac.jp`

Abstract. Recently, computer networks, including the Internet, have emerged as one of the most important networks of society. Therefore, network security problems must be taken seriously. To improve the knowledge of security, there are various forms of security education. Exercises for network security provide the advantage of hands-on experiences of attacks. However, there is an ethical problem, that is, students learn not only the method of defense against cracking but also the method of attacking. In this research, to eliminate these problems, we developed a remotely accessible exercise system with a virtual cracker function for generating automatic cracking. On realizing this system, learners will be able to experience and learn only the practical methods of defending networks against cracking.

1 Introduction

In recent years, information security problems have become more serious, and educational facilities such as universities and vocational schools provide various lectures for learning security. Such lectures consist of classroom lectures and exercises: the former is designed for learning offensive and defensive concepts systematically through books, and the latter is designed for experiencing real attacks and their defense via network equipment.

In the latter case, however, it is troublesome to set up the networks for security exercises by using actual equipment (Problem 1).

Moreover, there is a problem in the exercises for defending attacks (Problem 2). There is a shortage of staff such as teachers and teaching assistants who generates attacks against students' networks. On the other hand, teaching students to generate attacks conflicts with ethical standards [1]. Alternatively, there is also another type of exercise: students construct networks and then defend their own networks against attacks generated by crackers in the Internet. In such exercises, however, students do not get stable experiences of attack and defense because there is a possibility that the crackers may not attack their networks.

We have developed a remotely accessible exercise system for conducting network security exercises [2]. The system provides networks consisting of virtual machines with users through the Internet. Henceforth, a network consisting of virtual machines is called a "virtual machine network." The users can manage the networks from home using their PCs. Hence, the system can resolve Problem 1.

Moreover, the system contains functions for constructing virtual machine networks and attacking them automatically. The system, however, fails to attacks networks updated by the management of the users, because this function is designed for running malicious software in each network apparatus only at the time of their initialization.

For conducting network security exercises, we have proposed a remotely accessible exercise system with a virtual cracker function to resolve Problem 2. The virtual crackers identify security holes by analyzing virtual machine networks, and then attack them.

2   Related Studies

Nakama developed a curriculum for network security and reported its practical results [3]. His students constructed experimental networks, which consisted of actual machines, and connected them to the Internet. However, the report states: "Contrary to our expectation, we could suffer a few attacks from crackers in the Internet, and those attacks failed to achieve intrusion into our system." Conducting exercises that have accidental attack opportunities is a critical issue due to the time limitation of classes, although this idea is one of the solutions for addressing the issue of maintaining ethical standards.

Tele-lab IT security [4] is a virtual laboratory environment, in which students can learn to use network security tools. In this environment, students use attack/defense tools installed in the virtual machines running on servers; the students can access the exercise environment servers by using a browser and Virtual Network Computing (VNC) applets running at a remote location. SEED [5] is an exercise course that can be used to realize a virtual machine network for students. Students attack the network of their partners and defend their own network from attacks launched by their partners. Students can acquire more technical skills by using SEED than by using Tele-lab. However, these two applications do not address the issue of maintaining ethical standards (Problem 2).

3   System Design

3.1   System Overview

Fig. 1 shows the construction of our system. It consists of a server machine (exercise server), which runs the virtual machine networks of the students, and client machines, which are used to operate the virtual machine networks. The clients display GUIs for the exercise on receiving them from the server. The virtual crackers attack the networks in the server that are managed by the students by reading attack definition files (Section 3.3).
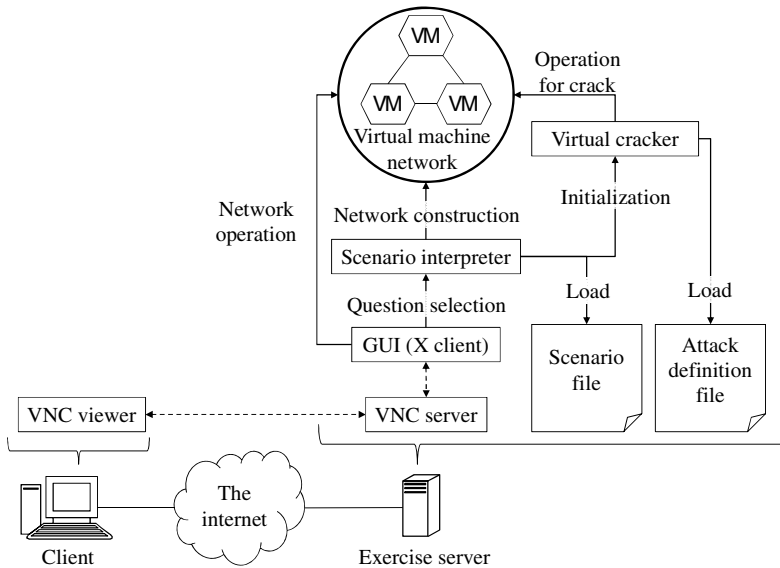
Fig. 1. System structure

Scenario interpreter constructs the networks, which are assigned to the students, based on the network definition data in the scenario files. Before starting the exercise, teachers prepare scenario and attack definition files. After starting the exercise, students login to the server through their clients and manage the networks for a fixed period.

3.2 A Remotely Accessible Exercise System for Network Security

We have developed a system for exercising network management through virtual machine networks; it consists of User-Mode Linux (UML) virtual machines [6], and it is called LiNeS (Linux Network Simulator). UML is a Linux emulator, and it is run on Linux OS. The virtual machine networks consist of virtual Linux servers, routers, clients, and firewalls and virtual switching hubs.

Furthermore, we constructed a server, which runs LiNeS and accepts remote operation through VNC [7]. Therefore, the equipment required for a student to exercise is just a PC with VNC, connected to the Internet. In this study, using this server, we have implemented the exercise server in Fig. 1.

Students design network topologies via a GUI of LiNeS with mouse operations, and they set up virtual Linux servers and routers from their console windows, and virtual Linux clients using GUIs.

LiNeS automatically constructs virtual machine networks based on network definition data. Teachers can set network topologies and initial settings of virtual network apparatuses as network definition data. Using the initial settings based on which the virtual network apparatuses run and attack, LiNeS generates virtual networks, including virtual attacks.

LiNeS, however, fails in attacking the networks that are updated by the management of the users, because the above method is designed for running the software in each network apparatus only at the time of their initialization. Accordingly, we have proposed the development of virtual crackers.

3.3 Virtual Crackers

Fig. 2 shows an attack flow which virtual cracker executes in our system. Virtual crackers act in each phase ((1)-(4)) based on the following information: A) time, B) execution results of attack tools, and C) state of virtual networks. A) means absolute and relative time. The absolute time is GMT and the relative time is the time elapsed since the exercise start time and the time when virtual crackers act. B) is the result of the nmap[8] port scan and the success or failure of the attacks by THC-Hydra[9] and Metasploit[10]. Interpreting the data, which is obtained through API2, the virtual crackers perform B). In C), the crackers analyze the information of virtual network apparatuses (for example, running process and server settings). After searching virtual network apparatuses via API1, obtaining information of the apparatuses by API2, and then analyzing the information, the crackers perform C).

Virtual crackers act based on attack definition files in Fig. 2. The files consists of production rules using XML tags in Table1. Event tags are used to define crackers' actions which are (1)-(4) in Fig. 2. Attributes are used to define their execution conditions.

LiNeS has two APIs: one, API1, is designed for editing topologies of virtual machine networks, and the other, API2, is designed for operating each virtual network apparatus. API1, for example, is used for adding new network apparatuses and changing connections between existing network apparatuses. API2 is used for booting/shutting down network apparatuses and executing Linux commands in virtual Linux servers, routers, firewalls, and clients.
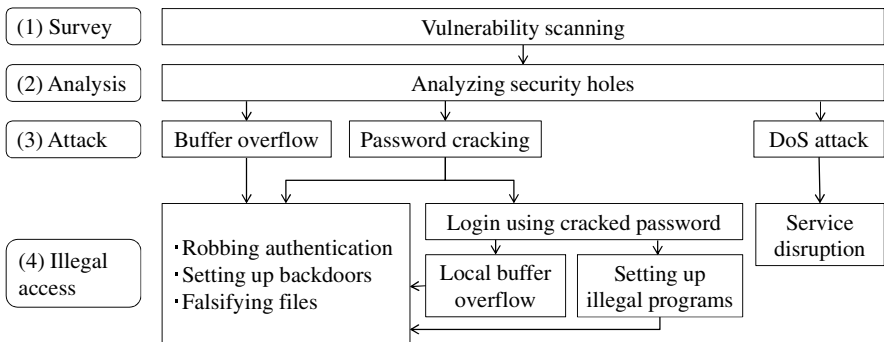
Fig. 2. Cracking flow in the exercise contents

Table 1.. A part of tags and attributes for attack definition files

Event tag	Description
netscan	Execute network scan by nmap
search	Search a keyword from the file in a virtual network machine
passwordattack	Execute password crack by THC-Hydra
metasploit	Execute attacks by Metasploit
ssh	Execute remote login by ssh. This tag is able to include command tag as a child node.
command	Execute a linux command after finishing ssh by a ssh tag
Attribute	**Description**
id	ID of event
time	Event start time
wait	Wait time before starting event
success	Next event ID when event is finished successfully
failure	Next event ID when the event is finished unsuccessfully

4   A Prototype System

A student operates LiNeS by establishing a VNC connection between his/her PC and an exercise server through the Internet. Fig. 3 shows an example of the display for an exercise. The student is using a Windows PC and operating a virtual machine network of LiNeS in a VNC viewer.

The flow of the exercise is as follows: After the student selects a question that is shown by the system, the system constructs a network, which corresponds to the question, and provides the student with the corresponding network. After time passes, a virtual cracker starts to act based on an attack definition file. In this case, the virtual cracker carries out a SSH brute force attack against a server in the student's network. As the result, the server records logs, which are shown in Fig. 4. Furthermore, the virtual cracker creates a back door by using the root privilege obtained by the SSH attack.

Subsequent actions of the cracker depend on the steps taken by the student. If the student destroys the backdoor or limits access to the server using the firewall after noticing the attack and the backdoor, the cracker would carry out DoS attacks against the server. Afterwards, the student will notice that he/she cannot browse the web pages in the server. As the result, the student starts dealing with the abnormal behavior of the server.

On the contrary, if the student did not notice the attack at all or did not deal with it appropriately, the back door would be kept available. As a result, the cracker intrudes the server through the back door and deletes web pages illegally. Then, the student notices the abnormal state of the server when he/she cannot find the web pages. Hence, the student tries to identify the abnormal logs and deal with the attack.
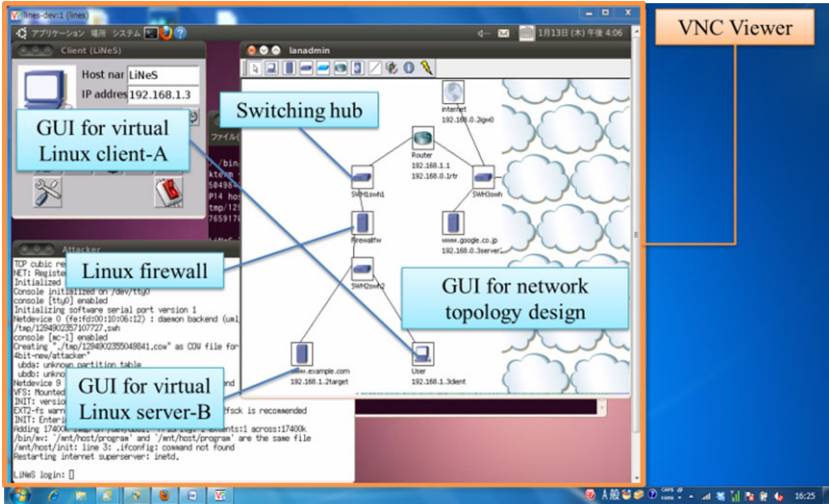
Fig. 3. Accessing an exercise server via VNC

Fig. 4. A server console which indicates SSH brute foce attack and intrusion

5   An Evaluation Experiment

5.1   Purposes of the Evaluation Experiment

There are two purposes of this evaluation experiment. One is to validate the contents of the exercise based on/with respect to the skills/knowledge of the students. Our system targets those students who have the skills/knowledge of network construction and the foundation of network security. Therefore, we analyze the actions of the users in these exercises.

The other is to clarify the necessity and the effectiveness of the exercises generated by our system. We circulated questionnaires to gather opinion of subjects who have completed the exercises using our system after learning through classroom lectures.

5.2 Summaries of the Evaluation Experiments

The subjects consisted of 12 students who have experience of network construction. The authors were the experimenters. The procedures of the experiments are as follows:

1. Each subject reads the answer books to understand the mechanisms and methods of DoS attacks, back doors, and buffer overflow. Furthermore, the experimenters explain, when necessary, to help the subject understand the material.

2. The subject is asked to manage a virtual network in the system for 20 min as a network security exercise. The experimenters record the behavior of the subject, which influences the results of the attack, while observing the subjects. The subject records the changes he/she makes to the network settings.

3. The subject answers the questionnaire.

Each subject provides the answer to a problem, which in this case is the secure management of a virtual machine network in our system. The virtual machine network has the same topology as shown in Fig. 3.

The subjects have to manage their own networks such that the networks work as per the following: Since the role of the server is to publish web pages, it is required that the server runs WWW, FTP, SSH, and TELNET services. It is required that the client machine can upload web pages to the server by FTP, remotely login to the server by password authentication of SSH and TELNET, and set the WWW server software in the server.

Furthermore, the subjects have to follow the following rules, which are called "action rules":

・All you can do to the server is view the configuration files under /etc, edit configuration files under /etc, except /etc/apache, and reboot the process of the services.

・All you can do to the firewall is interrupt communications from external hosts by iptables.

・You must not edit the network topology and other machine settings.

Needless to say, the subjects are not informed of the security holes of the server. The security holes are provided to permit remote login with root user, vulnerable passwords, and to run a process with buffer overflow vulnerability.

5.3 Evaluation of the Validity

Fig. 5 shows the action flow of a virtual cracker for the experiment. The virtual cracker decides its own actions, depending on the actions of subjects in the first and the next 10 min.

Table 2 shows the relations between the actual actions of the cracker and the expected actions of the cracker according to the actions of the subjects. No subject

prevented the virtual cracker from attacking his/her network. In addition, there was not a single subject who did not take any step for network security. Therefore, this content is valid for these subjects.
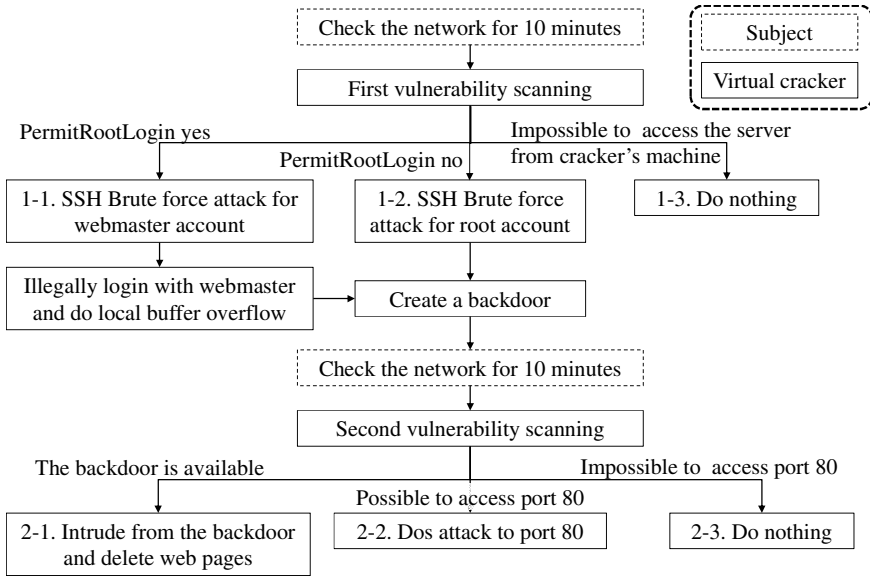
Fig. 5. Action flow of a virtual cracker

Table 2. Relation between the actual and expected actions of the cracker

The actions of the cracker	Expectation	Actuality
1-1 in Fig. 8	1	1
1-2 in Fig. 8	11	11
1-3 in Fig. 8	0	0
2-1 in Fig. 8	6	6
2-2 in Fig. 8	2	2
2-3 in Fig. 8	4	4

5.4 Evaluation of the Necessity and Effectiveness

We used the questionnaires illustrated in Table 3 with the following scale for evaluation: 5 – very much; 4 – a lot; 3 – somewhat; 2 – not much; and 1 – not at all. The questionnaire also contained space for comments.

The results of Q.1 and Q.2 indicate that the exercises generated by our system are necessary and effective in managing networks securely. We consider Comment B to be a problem of our system. For resolving this problem, we will develop a function to restrict the action of students in the system based on the settings made by the teachers.

Table 3. Usefulness of automatic attack function in promoting learning

Question	Average
Q1. Did you think the measures against cracking are necessary because of the exercise in our system?	4.9
Q2. Did you understand the ways of intrusion detection and attack defense more?	4.7
Comments	
[A] Although I understood attack methods and defense methods each except for their relations, I could not utilize them to network security. Such a practical exercise was impressive and helpful for me.	
[B] It is necessary to discuss the way to prevent students from violating the rules carelessly because there are no teachers in remote location.	

6   Conclusions

In this study, we developed a network security exercise system for exercising only the ways to detect and defend attacks by introducing virtual crackers, which automatically generate attacks in our remotely accessible network security exercise system. As the result of the evaluation experiments, the behavior and evaluation questionnaires of the subjects indicate that our purpose is accomplished. Our future works will be focused on developing user interfaces that help students to follow action rules, and to improve virtual crackers to enable them to generate more sophisticated and complex attacks.

References

[1] Harris, J.: Maintaining ethical standards for a computer security curriculum. In: Proceedings of the 1st annual conference on Information security curriculum development, pp. 46–48 (2004)

[2] Tateiwa, Y., Yasuda, T.: Multiuser network administration training in LiNeS: Connection function between virtual networks. In: Proc. of KES-IIMSS 2009. SCI, vol. 226, pp. 535–544 (2009)

[3] Masahiro, N.: A plan and methods of instruction for computer security education in teachers college students Bulletin of College of Education, vol. 60, pp. 219–228. University of the Ryukyus (2002)

[4] Hu, J., Meinel, C., Schmitt, M.: Tele-lab IT security: an architecture for interactive lessons for security education ACM SIGCSE Bulletin. SESSION: Computer security 36(1), 412–416 (2004)

[5] Du, W., Wang, R.: SEED: A suite of instructional laboratories for computer security education. Journal on Educational Resources in Computing (JERIC) 8(1), Article No. 3 (2008)

[6] The User-mode Linux Kernel Home Page,
     `http://user-mode-linux.sourceforge.net/index.html`
     (accessed March 14, 2011)
[7] RealVNC, `http://www.realvnc.com/` (accessed March 14, 2011)
[8] Nmap, `http://nmap.org/` (accessed March 14, 2011)
[9] THC-HYDRA, `http://freeworld.thc.org/thc-hydra/`
     (accessed March 14, 2011)
[10] The Metasploit Project, `http://metasploit.com/` (accessed March 14, 2011)

Lip Print Recognition for Security Systems: An Up-Coming Biometric Solution

Pawan Sharma, Shubhra Deo, S. Venkateshan, and Anurika Vaish

Department of Masters of Science in Cyber law and Information Technology,
Indian Institute of Information Technology, Allahabad
Allahabad, India
psh_099@yahoo.com, shubhra.deo@gmail.com

Abstract. Biometrics is a process for identification of a person on the basis of physical and biological attributes. It has been highly developed to a stage where identification and authenti-cation of person is done through various means of biological features. However the major concerns in biometric is about obtaining a complete accuracy for verification and identification i.e. the false acceptance rate FAR and the false rejection rate FRR. Lip printing mechanism is an upcoming method which is now being used for identification in rare cases. In this paper we have proposed a methodology for a clear reading of lip print with the help of pattern matching using brute force algorithm. Multi biometrics technique are clubbed together to find a perfect solution for identification and verification. A lip print image is matched with the stored data-base and can be then used either for authentication or identification in crime case or can also be club with other authentication technique of biometrics to make stronger and robust authentica-tion mechanism.

1 Introduction

Biometrics is a process of reflex identification of a person based on his/her physio-logical or behavioral features. This method of recognition is preferred over conven-tional methods like passwords and PIN numbers as the identification based on biometric techniques obviates the need to remember a password or carry a token. It is an approach for life measurement and is related with distinctive physiological characte-ristics for identifying individuals. Biometrics emerged as a concept of security; it is used as the computer interface and pattern matching to identify an individual. A range of biometric applications are being used for authenticating person's identity. With the use of various features including fingerprints, face, signature, and iris, a person can be identified. Lip print, body odor, gait recognition [2] are some of the newly identified unique physical characteristics to identify an individual. Multi biometrics is used now days to increase the accuracy of biometrics models. Two different biometric are com-bined to give a more accurate model of authentication and identification such as face and fingerprint of same person, palm print and hand geometry, voice and lip move-ment are combined. The coalition is done either at the score level, sensor level or at feature level. Clubbing of two features set yields more accurate results due to inde-pendency of each other.

Here in this paper we have proposed a methodology for lip print recognition that can be used in forensics labs for identification of an identity. Lip printing recognition has its own merit such as small data requirement and quick pattern matching. The uniqueness of a lip-print is typically, determined by the overall pattern of grooves. Study of lip print is known as Cheiloscopy. Cheiloscopy is derived from the Greek words cheilos, lips, skopein, to see. It is the term given to the lip print studies. Lip prints can be recovered from various objects like glasses, cigarette etc. These evidences can be vital in cases of high complexity. Unfortunately the lip print analysis has not been developed substantially. The fact that every human has unique lip print is affirmed by Yasuo Tsuchihashi [1] and Olufemi [2]. In his paper the author investigated on various subjects to study the lip print and established the result that no lip prints showed the same pattern and even lip prints for twins are not identical. They have same integrity as that in fingerprints.

A lip-print is a pattern of grooves that exist on the surface of the lips. Designing a reliable automatic lip print matching algorithm is quite challenging. However, the popularity of lip-print sensors is not yet developed and study in this area is still at a nascent stage. The critical factor with lip print matching is that they must satisfy with performance in terms of accuracy and processing speed. For the alignment of two lip-prints certain landmarks are needed. These should be automatically extracted with low misidentification rate. As landmarks we suggest is the prominent symmetry points (core-points) in the LIP PRINT. They are extracted from the complex orientation field estimated from the global structure of the LIP PRINT, i.e. the overall pattern of the grooves. Investigators often gain evidence through the use of Cheiloscopy. As in case fingerprints, experts can lift lip prints from objects found at crime scenes and compare these prints to a suspect's lip pattern. The remnants of the paper is organized as fol-lows: section (2) is about literature review section (3) focuses on methodology section (4) is about noise filtering section (5) is on pattern classification and matching, sec-tion (6) is the experimental result followed by conclusion and acknowledgment.

2   Literature Review

Study shows that there had been a smaller portion of work done in lip print technology. Personal identification with biological attribute is not an easy approach. The most common attributes used in this context are DNA and fingerprint, due to its accuracy and large database. Also because the above mentioned techniques has been established way back long, variety of scanners for the same are available in the market and is easy to install. Apart from the mentioned technique there are few more biometric techniques like iris scan, retina scan, hand geometry and facial recognition [2] but all have their own limitation. On one hand, where DNA gives more accurate measurements the fingerprint has highest usability due to large database and cost effective methodology. The external surface of lips has many elevation and depression forming a characteristic pattern called lip prints, also known as Cheiloscopy. The uniqueness of lip for each individual makes it an interesting area for pattern matching, identification and verification. Lip prints are helpful in forensic investigation that deals with

identification of a victim in the court of law. More research work in this area need to be conducted with regards to confirmation of distinctiveness, and for the collection of evidences.

With a rise in biometrics solutions for identification, verifications had pulled more attentions because they simplify the process methodology of traditional identification and verification. Considering the biological characteristics of physical attribute, human luminescence plays an important role while searching of invisible evidence at the crime scene. In fingerprints extraction we use reagents for their localization. The lip prints which are left on crime scene may contain lipstick mark or an invisible impression which act as a protective shield and doesn't leave any visible mark. Hence in case of lip print regular reagents cannot be used for the analysis. In a paper by Ana Castelló [5] it is mentioned that Nile red can be used as a potential developer for lip prints. The results demonstrate that this reagent is highly effective even on one year old lip prints. A study conducted on a methodology of lip print mentions the procedures to maintain a database of lip print for verification purposes; it states that one can take the mark of lip prints on a suitable media like paper after application of lipstick or lip rouge. Also a high resolution photograph with focused and proper lightning at a particular angle can also be used. Different angles of lip prints can be taken to enable efficient matching.

R.Fische was the first person to describe about the biological characteristics of lips. Edmond Locard, one of France's greatest criminologists, acknowledged the importance of Cheiloscopy and first proposed that lip print can be used in identification and crime investigation. Japanese research on the subject of lip print established that arrangement of lines on the red portion of person's lips is unique for every individual.

3 Methodology

Lip prints recognition is relatively not fully formed in biometrics. No public data-base is available for it. To validate our proposed methodology, we have created a database of lip images of 200 individuals. We used Nikon 300 D 12.5 MP cameras. The analog signal output of the camera is transferred to the computer using a frame grabber PIXCI. The attainment was done at room temperature in an office environ-ment. The skin colors of the subjects varied from dark to fair and the subjects were people of different age group and geographical location varied from 20-60. Images were obtained for both the genders. We haven't made any classification on the basis of gender as earlier done by Shailesh M Gondivkar and others [3][4][5].Those with a disease or deformity has not been made part of this analysis. The percentage of failure was 0.2 for whom the system was unable to generate template which measures the proportion of individuals.

The flow chart in Figure 1 represents the system's flow. The system started with image acquisition which in this case is the acquisition from the database. There would be two inputs one as a lip print of an individual which represents the stored database other as a lip print that needs to be verified. Both inputs are then processed to grooves and then the second input is matched against the first input to verify whether the lip print belongs to the same individual or not.
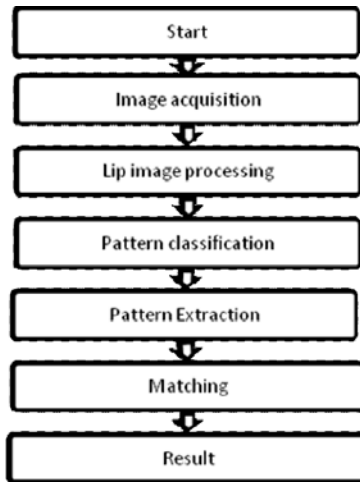
Fig. 1. Flow chart representation of methodology

4   Lip Image Processing

The extracted lip prints have a lot of unwanted details like the skin, hair etc. and ef-
fected due to noise. Hence a pre processing is done [5] for contrast enhancement of
images. It employs an adaptive filter that controls the contribution of the sharpening
path in such a way that contrast enhancement occurs in high detail areas and little or
no image sharpening occurs in smooth areas. We have used Median filter to remove
noise [6][7][8]. We propose a physical diffusion process where the concentration
balance depends upon the density gradient. In physical diffusion for the given image a
(x,y,t) the diffusivity 'g' depends on the gradient as shown in (1)

$$c \to 0 \ \ for \ \ \nabla a \ \to \ \infty \tag{1}$$

$$c \ \to 1 \ for \ | \ \nabla a \ | \to 0$$

The gray value of each pixel is calculated iteratively depending on the gray value
gradient in a 4-pixel neighborhood surrounding the pixel. The gradient is considered
using the non-linear diffusion function 'c' so that the smoothing is more over the uni-
form regions rather than the edges, and the edges remain sharp.

5   Pattern Classification

On studying the various obtained lip images we observed few patterns which were
peculiar for each lip. The primary recognition is by the width and breath of upper and
lower lips individually along with the kind of pattern which can be seen on the lips.
The metrics of lips is measured and recognition is done by the kind of lips. We have
classified lips into five major classifications (represented in the figures 3 to 7). Our

analysis is in harmony with that of JO Kim [9], El Domiaty MA [10], Vahanwala Sonal [11], Renaud M [12]. The lip print was divided into six topographical areas and each area was examined alone to establish the type of the grooves as shown in figure 2.
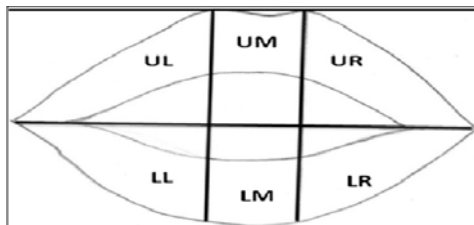
Fig. 2. The division of lips

Class I: The incomplete grooves which diminish gradually, as shown in Figure 3.The specialty of this class is that the lips are divided and distinct with the help of small patterns throughout the lips.

Fig. 3. The incomplete grooves which diminish gradually

Class II: The bifurcating grooves which form branch like structure as shown in Figure 4.These lips are more remarkable due to formation of stem like structure and divides it in small lines.
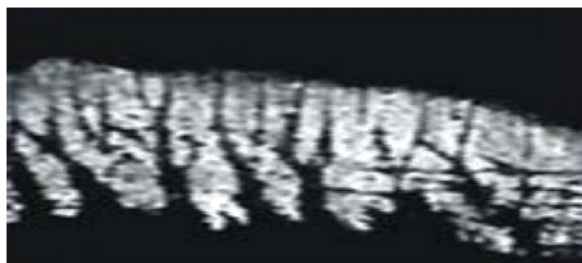
Fig. 4. The Bifurcating Grooves

Class III: The crossed grooves which have many intersections in its pattern as shown in Figure 5. These lips form an intercross web like structure giving it a diffeerent structure compared to other. It is more complicated in grooves structure and pattern.
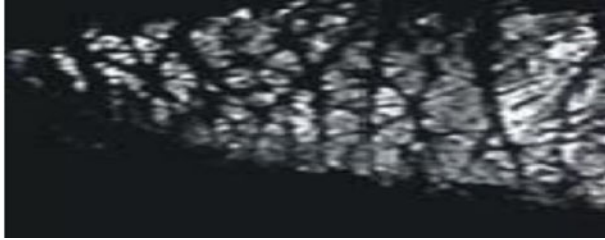
Fig. 5. Crossed Grooved

Class IV: The grooves forming a net like design as shown in Figure 6. There net like structure make it more discernible compared with others.
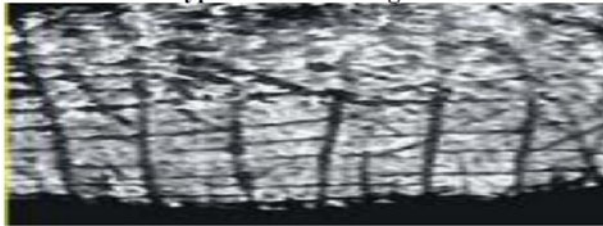
Fig. 6. The Grooves Forming A Net Like Design

Class V: These types of lip pattern doesn't have any distinct pattern and are arbitrary in nature as shown in Figure 7. Any other pattern which could not be classified in the above four falls under this category.
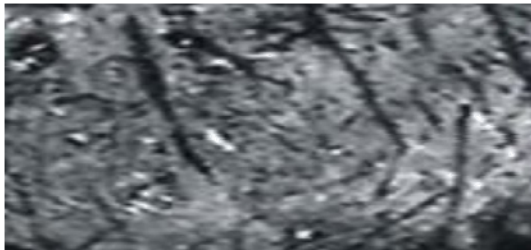
Fig. 7. Unclassified Which Do Not Form Part Of Any

After pattern classification of the lip print images the images are skeletonised so as to obtain the linear image of the lips which can be processed for pattern matching of the lips.
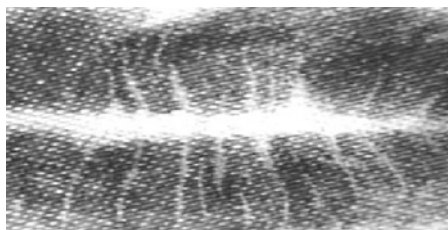
Fig. 8. Matlab veiw of region.

Figure 8 shows a zoomed snapshot of the software "MATLAB (Ver. 7.0) used for processing on the basis of pattern of lip images. This pattern is matched as step of verification of our propose modality. We involved grooves segmentation [13] which normalizes lip print image and segments grooves region. It identifies grooves regions of a lip print image and returns a mask identifying this region. It also normalizes the intensity values of the image so that the grooves regions have zeroed mean, unit standard deviation.

5.1 Pattern Matching

During the grooves matching the user lip image is acquired, processed and com-pared with the templates stored in the database. The user is accepted or rejected on the basis of the result of pattern matching. Pattern matching is the act of scrutiny for the presence of the constituents of a given pattern. We have tried to match this pattern with the help of Brute force algorithm.

5.2 Brute-Force Algorithm

The algorithm compares the pattern A with the stored lip patterns B in database for each possible shift of A relative to B until either

- A match is found or

- All placements of the pattern have tried

The algorithm Brute Force Match BFM (A,B) is presented as follows:

- Image A of size o and pattern B of size p is given as input.
- The starting index of a string of A equal to B or-1 is obtained as output.
- If no such substring exists, shift i of the pattern is tested from 0 to o-p.
- If j=p, match at 1 is found else no image match.

6 Experimental Results

Different threshold values were used for lip pattern matching to deduce the false acceptance rate and the false rejection rate. False Acceptance Rate refers to the total

number of unauthorized persons getting access to the system against the total number of people trying to use the system. False Rejection Rate refers to the total number of authorized persons not getting access to the system over the total number of people attempting to get access to the system.

6.1 False Acceptance Ratio and False Rejection Ratio

A sample from Person 2, P21 is tested against four samples from other persons with similar groves to some extent numbered P2, P3, P4, P5, and four other samples from Person 1 (P11), Person 3 (P31), Person 4 (P41) and Person 5 (P51) with a completely different grove pattern. The results obtained are summarized in Table 1. The results are obtained from MATLAB's command window.

Table 1. Matching percentage of sample P21 against other samples.

Samples Pair	Matching Percentage (%)
P21 against P2	35.7729
P21 against P3	28.3887
P21 against P4	25.1357
P21 against P5	28.4257
P21 against P11	4.0195
P21 against P31	4.1090
P21 against P41	3.9907
P21 against P51	4.1651

A scatter graph of matching against sample P21 is obtained from table 1 as shown in Figure 9. The matching percentages are very close and high for sample P21 against four samples of similar groves while the matching results against sam-ples from other
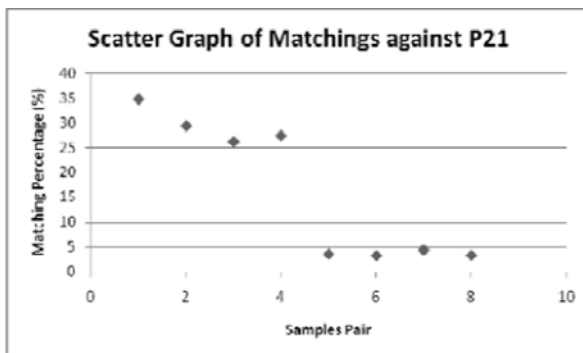
Fig. 9. Scatter graph of matching for sample P21

persons are low. This is clearly depicted in the scatter graph. It is visible that the matching percentages between two samples of similar groves are higher and more than 28%. We found that the matching percentages between two samples of different groves are smaller than 6%. Therefore, a threshold can be set to determine whether two samples are from the same lip. This threshold is used to verify the lip prints as of the same person or a different person.

Five samples from ten individuals are used and are tested against each of the sample. There are 200 verifications done in total.

6.2   False Reject Rate (FRR)

Authentic verification is a substantiation of the same entity. Matching percentage of genuine verification varies around a certain mean value. If a verification thre-shold that is too high is applied to the system, some of the genuine matching pairs are falsely rejected. Depending on the value of the threshold, the data that will be falsely rejected can be from zero to all images.

False Rejection Rate (FRR) = Number of rejected data / total data.

Table 2 shows the varying threshold values and the corresponding FRR obtained from the system's 200 verifications.

Table 2. Different threshold values and the corresponding FRR

Threshold (%)	FRR (%)
10	0
12	0
14	0
16	0
18	4
20	6
22	12
24	22
26	30
28	40
30	46
32	46
34	52

A graph is plotted for the FRR and is as shown in Figure 10. The percentage of the FRR ranges between 0 and 100 and it is increasing.
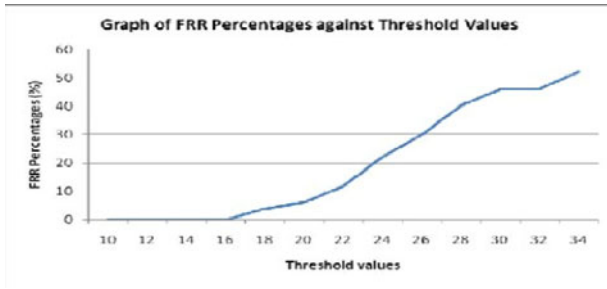
Fig. 10. FRR percentages graph against different threshold values.

6.3 False Acceptance Rate (FAR)

Depending on the choice of threshold, the arbitrary lip prints that are falsely ac-cepted by the system can be from none to all images. Impostor is an individual that is verified against a different stored database.

False Acceptance Rate (FAR) = Falsely accepted data / Number of all impostor data

Its value is one, if all impostor data are falsely accepted, and zero if none of the unknown data is accepted. Figure 11 shows the percentages of the FAR against different threshold values plotted using the data in Table 3.

Table 3. Different threshold values and the corresponding FAR

Threshold (%)	FAR (%)
10	53.33
12	51.78
14	49.56
16	42.22
18	37.11
20	27.78
22	17.33
24	10.44
26	4.89
28	0.44
30	0
32	0
34	0

The range of percentage is between 0 and 100, and is exponentially decreasing.
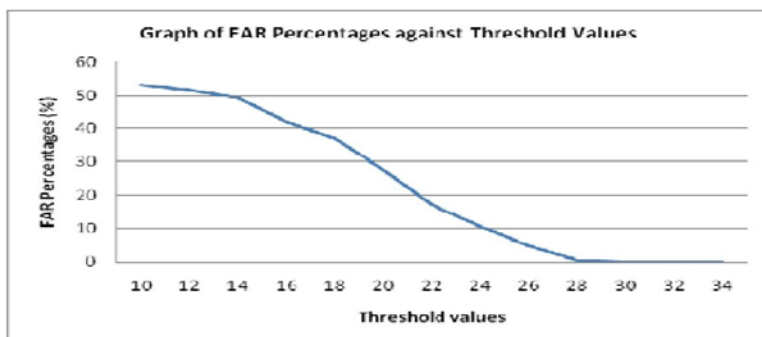
Fig. 11. FAR graph percentages against different threshold values.

6.4   Equal Error Rate (EER)

By determining the Equal Error Rate (EER) the choice of the threshold value is made easier. EER is the value where the FAR and the FRR intersect and the value is equal for both rate. The EER of a system can be used to give a threshold inde-pendent performance measure. The lower the EER is, the better is the system's performance, as the total error rate which is the sum of the FAR and the FRR at the point of the EER decreases. The EER of this system is shown in Figure 12.
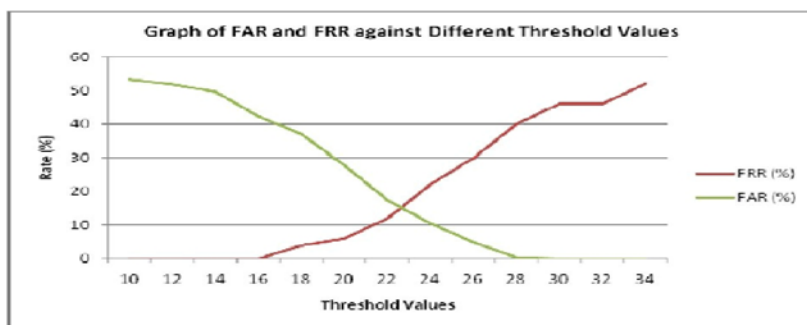
Fig. 12. Graph of FRR and FRR against a range of threshold values

The EER of the system is approximately 15% as shown in figure 12. The threshold taken for verification is 22.8.Using the determined threshold value as 23, the data obtained from 200 samples are considered. The number of times that the system re-sponds with correct and incorrect verification is counted. Total data counts are 200. There are six verifications of the same person rejected, while 15 verifications are falsely accepted. The total of correct verifications is 179. Verification rate is counted by taking the percentage of correct verifications over total data count. Therefore, the verification rate of this approach is 89.5%. Table 4 summarizes the data.

Table 4. Experimental results

Falsely rejected data	6
Falsely accepted data	15
Correct Verifications	179
Total data count	200

The rate of false acceptance and false reject can be influenced by several rea-sons. The reasons could be two different lip prints having the same principal lines, different illumination and also alignment of the lips on the capturing device. True negative rate which is the genuine rejection can also be influenced by having dif-ferent principal lines and also different lip sizes.

7 Conclusion

In this paper we have successfully implemented a modality based on unimodal biometric system for lip matching. It is an effort to understand how lip print recognition can be effectively used in forensics for crime investigation. It is the basic work flow of lip print matching which includes various stages from lip extraction till the matching of the processed lip print data. A major challenge in lip print matching is the quality of images that are being used for processing and the efficiency of the algorithm used for fast retrieval of lip print data from the database. Lip print cannot be measured on the same scale as finger prints recognition regarding the data base. However de-spite this fact some countries have started collecting database for lip prints and are working to implement it in the fields of criminal identification. Some more research is needed in this upcoming entity identification process which can tackle complexities like changes in lip patterns due to ageing and seasonal changes.

The lip print of a human is considered as personal identification and hence the analysis of lip print can help us in forensic investigation. Till date no standard methodology for lip print has been developed which has been accepted in the forensic science community.

Acknowledgement

References

1. Tsuchihashia, Y.: Studies on personal identification by means of lip prints. Forensic Science 3 (1974)
2. Adeoye, O.S.: A survey of Emerging Biometric Technologies. Interntional Journal of Computer Application 0975-8887 9 (2010)
3. Gondivkar, S.M., Indurkar, A., Degwekar, S., Bhowate, R.: Cheiloscopy for sex determination. Journal for forensic dental sceinces, 56–60 (2009)
4. Lakshmi Deepika, C., Kandaswamy, A.: An Algorithm for Improved Accuracy in Unimodal Biometric Systems through Fusion of Multiple Feature Sets. ICGST-GVIP Journal, 33–40 (2009)
5. Castello, A., Segui, M.A., Verdu, F.: Luminous lip prints as criminal evidence. In: Forensic Science International, Valencia, pp. 185–187 (2005)
6. Im, S.-K., Park, H., Kim, Y., Han, S., Kim, S., Kang, C., Chung, C.: A Biome-tric Identification System by Extracting Hand Vein Patterns. Journal of the Korean Physical Society, 268–272 (2001)
7. Shahin, M., Badawi, A., Kamel, M.: Biometric Authentication using Fast Correlation of Near Infrared hand vein patterns. International Journal of Biomedical sciences 2, 141–148 (2006)
8. Ko, T.: Multimodal Biometric Identification for Large User Population using Fingerprint, Face and Iris Recognition. In: Proceedings of the 34th Applied Imagery and Pattern Recognition Workshop (AIPR 2005), Washington, DC (2005)
9. Kim, J.: Lip print recognition for security systems by multi-resolution architecture in Future Generation, pp. 295–301. Elsevier, Seoul (2004)
10. Sonal, V.: Nayak C.D., Pagare S.S.: Study of Lip-Prints as Aid for sex De-termination. Vahanwala Sonal et al/Medico-Legal Update, Mumbai, pp. 93-98 (2005)
11. Caldasa, I.M., Magalhães, T., Afonsoa, A.: Establishing identity using cheiloscopy and palatoscopy. Forensic Science International, 1–9 (2007)
12. Sivapathasundaram, B., Ajay Prakash, P., Sivakumar, G.: Lip prints (Cheiloscopy). Indian Journal of Dental Research, 234–237 (2001)
13. Vamsi KrishnaReddy, L.: Lip prints: An Overview in Forensic Dentistry. Journal of Advanced Dental Research, 17–20 (2011)
14. Sleit, A.,, H.: saadeh, I. Al-Dhamari, A. Tareef, An enhanced sub image matching algorithm for binary imags. Recent advances in applied mathematics,Jordan, pp.565-569 (2009)
15. Cheiloscopy, K.J.: Encylopedia of Forensic Science. Elseiver, UK (2000)

Author Index