ADVANCES IN INTELLIGENT AND *SOFT COMPUTING* 60

Zdzisław S. Hippe
Juliusz L. Kulikowski (Eds.)

Human-Computer Systems Interaction

Springer

Advances in Intelligent and
Soft Computing 60

Editor-in-Chief: J. Kacprzyk

Advances in Intelligent and Soft Computing

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

Vol. 46. V.-N. Huynh, Y. Nakamori,
H. Ono, J. Lawry,
V. Kreinovich, H.T. Nguyen (Eds.)
Interval / Probabilistic Uncertainty and Non-classical Logics, 2008
ISBN 978-3-540-77663-5

Vol. 47. E. Pietka, J. Kawa (Eds.)
Information Technologies in Biomedicine, 2008
ISBN 978-3-540-68167-0

Vol. 48. D. Dubois, M. Asunción Lubiano,
H. Prade, M. Ángeles Gil,
P. Grzegorzewski,
O. Hryniewicz (Eds.)
Soft Methods for Handling Variability and Imprecision, 2008
ISBN 978-3-540-85026-7

Vol. 49. J.M. Corchado, F. de Paz,
M.P. Rocha,
F. Fernández Riverola (Eds.)
2nd International Workshop on Practical Applications of Computational Biology and Bioinformatics (IWPACBB 2008), 2009
ISBN 978-3-540-85860-7

Vol. 50. J.M. Corchado, S. Rodriguez,
J. Llinas, J.M. Molina (Eds.)
International Symposium on Distributed Computing and Artificial Intelligence 2008 (DCAI 2008), 2009
ISBN 978-3-540-85862-1

Vol. 51. J.M. Corchado, D.I. Tapia,
J. Bravo (Eds.)
3rd Symposium of Ubiquitous Computing and Ambient Intelligence 2008, 2009
ISBN 978-3-540-85866-9

Vol. 52. E. Avineri, M. Köppen,
K. Dahal,
Y. Sunitiyoso, R. Roy (Eds.)
Applications of Soft Computing, 2009
ISBN 978-3-540-88078-3

Vol. 53. E. Corchado, R. Zunino,
P. Gastaldo, Á. Herrero (Eds.)
Proceedings of the International Workshop on Computational Intelligence in Security for Information Systems CISIS 2008, 2009
ISBN 978-3-540-88180-3

Vol. 54. B.-y. Cao, C.-y. Zhang,
T.-f. Li (Eds.)
Fuzzy Information and Engineering, 2009
ISBN 978-3-540-88913-7

Vol. 55. Y. Demazeau, J. Pavón,
J.M. Corchado, J. Bajo (Eds.)
7th International Conference on Practical Applications of Agents and Multi-Agent Systems (PAAMS 2009), 2009
ISBN 978-3-642-00486-5

Vol. 56. H. Wang, Y. Shen,
T. Huang, Z. Zeng (Eds.)
The Sixth International Symposium on Neural Networks (ISNN 2009), 2009
ISBN 978-3-642-01215-0

Vol. 57. M. Kurzynski,
M. Wozniak (Eds.)
Computer Recognition Systems 3, 2009
ISBN 978-3-540-93904-7

Vol. 58. J. Mehnen, A. Tiwari,
M. Köppen, A. Saad (Eds.)
Applications of Soft Computing, 2009
ISBN 978-3-540-89618-0

Vol. 59. K.A. Cyran,
S. Kozielski, J.F. Peters,
U. Stańczyk, A. Wakulicz-Deja (Eds.)
Man-Machine Interactions, 2009
ISBN 978-3-642-00562-6

Vol. 60. Z.S. Hippe,
J.L. Kulikowski (Eds.)
Human-Computer Systems Interaction, 2009
ISBN 978-3-642-03201-1

Zdzisław S. Hippe,
Juliusz L. Kulikowski (Eds.)

Human-Computer Systems Interaction

Backgrounds and Applications

Springer

Editors

Prof.-Dr. habil. Zdzisław S. Hippe
Chair of Expert Systems and
Artificial Intelligence
WSliZ - University of Technology and Management
ul. Sucharskiego 2
35-225 Rzeszów
Poland
E-mail: zhippe@wsiz.rzeszow.pl

Prof.-Dr. habil. Juliusz L. Kulikowski
Institute of Biocybernetics and
Biomedical Engineering
Polish Academy of Sciences
ul. Trojdena 4
02-109 Warsaw
Poland
E-mail: juliusz.kulikowski@ibib.waw.pl

ISBN 978-3-642-03201-1 e-ISBN 978-3-642-03202-8

DOI 10.1007/978-3-642-03202-8

Advances in Intelligent and Soft Computing ISSN 1867-5662

Library of Congress Control Number: Applied for

Typeset & Cover Design: Scientific Publishing Services Pvt. Ltd., Chennai, India.

Printed in acid-free paper

5 4 3 2 1 0

springer.com

Foreword

Recent increases in computational processing power and expanding computational context suggest shifting paradigms in the role of computing. Computers have become mobile and embedded in our lives. They serve as media and as information channels. They have become loci of experience as well as work horses. They have become ambient and proactive as well as deskbound and reactive. They have become immersive and intrusive. We can wear them as well as sit in front of them. They can even sense.

So far our *cohabitation* with computers has been guided by the Human-Computer Systems Interaction discipline. However, there is a need for refocusing this field from a traditional discipline delivering truth into a discipline delivering value – collection of more subjectively oriented usability centered design heuristics, valued by a system stakeholders. Thus, the term Human-Computer System Interaction has been coined to grasp these *postmodern* facets of the cohabitation in question

The University of Information Technology and Management (UITM) in Rzeszów, Poland, supports research devoted the above mentioned domain within the own, extended research program. Results gathered so far has entitled us to organize (with a technical co-sponsorship of IEEE) the International Human System Interaction Conference (held on May 25-28.2008 in Kraków, Poland), becoming now a cyclic international event. 44 out of 180 papers presented during this conference have been selected for this volume entitled **Human-Computer Systems Interaction: Backgrounds and Applications.**

But it is necessary to emphasize that this book is not a collection of post conference papers. All articles include new, so far unpublished research results, and even more they have been re-edited and distinctly extended by respective authors, to fulfill expectations of the envisioned readers. For this reason the book reflects on advances in human – computer system interaction, as well as complementary research efforts in related areas. It is impossible to summarise all the papers in brief but I hope readers will find them as a truly helpful guide and a valuable source of information about the state-of-the-art status of this extremely important domain of the computer science.

To sum it up, I wish to thank all the authors for their insights and excellent contributions to this book. I would also like to express my special thanks to the UITM faculty members for their excellent job in preparing this volume. I hope it will support the position of UITM in research and in education.

Tadeusz Pomianek
UITM President

Preface

For the last decades, as the computer technology has been developing, the importance of human-computer systems interaction problems was growing. This is not only because the computer systems performance characteristics have been improved but also due to the growing number of computer users and of their expectations about general computer systems capabilities as universal tools for human work and life facilitation. The early technological problems of man-computer information exchange – which led to a progress in computer programming languages and input/output devices construction – have been step by step dominated by the more general ones of human interaction with-and-through computer systems, shortly denoted as H-CSI problems. The interest of scientists and of any sort specialists to the H-CSI problems is very high as it follows from an increasing number of scientific conferences and publications devoted to these topics. The present book contains selected papers concerning various aspects of H-CSI. They have been grouped into five Parts:

I. General H-CSI problems (7 papers),
II. Disabled persons helping and medical H-CSI applications (9 papers),
III. Psychological and linguistic H-CSI aspects (9 papers),
IV. Robots and training systems (8 papers),
V. Various H-CSI applications (11 papers).

The papers included into **Part I** illustrate a variety of basic approaches met in H-CSI problems solving. To begin with cognitive methods used in decision making in ill-structured situations, through bionic models of information processing, ontological models of knowledge acquisition and representation, formal models of dialogue between man and computer system, consistency-based approach to the prediction of states of a system, up to annotation of images based on their semantic aspects. Of course, the above-mentioned papers are no more but several samples of possible approaches to H-CSI problems solution. However, they indicate the role of models and concepts concerning human thinking mechanisms in development of the advanced H-CSI tools.

Problems of helping disabled persons by substitution of their malfunctioning sense organs by suitable artificial tools, is one of the greatest social and economical importance. Since medical techniques of natural sense organs implanting are not sufficiently effective, the role of auxiliary technical devices will be unquestionable. This area of investigations is represented in **Part II** by papers concerning the visually disabled persons mobility helping by teleassistance navigation systems, eye-driven computer mouse and eye-blink controlled human-computer interface. Computer-aided recognition of gestures also belongs to this group of presented works. Computer-assisted medical diagnostic systems are represented by papers devoted to image analysis applications in oncology, cardiology and radiology.

Part III can also be divided into two subgroups of papers. In the first one the problems of automatic emotions recognition are dominating. Such systems can also help visually disabled persons in acquiring additional information from the face observations of their interlocutors. A paper concerning the concept of a day-dreaming machine shows that no limits for the attempts to fill the gaps between human and machine thinking can be a priori marked despite the fact that they are based on different backgrounds.

The second subgroup of papers in this Part concerns linguistic aspects of H-CSI systems design. One paper is also focused on the problem of virtual reality-based visualization framework perception and capture of information.

Problems of the design of robots, including their sensor subsystems, virtual reality modeling, biologically reasoned point-of-interest image compression, as well as kinetic analysis and training of surgical robots are considered in **Part IV**. A dominating idea consists here in functional capabilities of robots' extension on one hand, and human supervisory control of utilization of robots on the other one.

Miscellaneous H-CSI applications in enterprises management, technical diagnosis, molecular modeling, virtual museums organization, etc., are presented by the papers in **Part V**. They show that the H-CSI area of investigations and practical applications touches practically all possible domains of human activity.

To summarize, the book contains a collection of 44 papers written by an international team of contributors representing academia and research institutions from sixteen countries: Austria, China (also Hong Kong), Croatia, France, India, Iran, Israel, Italy, Japan, Poland, Romania, Russian Federation, South Korea, Spain, Taiwan, and the United States. We want to thank them warmly for supplying truly interesting and innovative papers.

Our special thanks go to Prof. Janusz Kacprzyk the editor of the Series for his invaluable support and help.

We are also indebted to DSc Teresa Mroczek, Ph.D. from the University of Information Technology and Management in Rzeszów (Poland) for her assistance in preparation of index, camera-ready copy of the book and conducting most of the correspondence with the authors.

Finally, we would like to express our sincere thanks to the publishing team at Springer-Verlag, in particular to Dr. Thomas Ditzinger for his permanent, versatile and very friendly help.

Zdzisław S. Hippe
Juliusz L. Kulikowski
Editors

Contents

Part I: General Problems of H-CSI

Part II: Disabled Persons Helping and Medical H-CSI Applications

Part III: Psychological and Linguistic Aspects of H-CSI

Part IV: Robots and Training Systems

Part V: Various H-CSI Applications

Part I
General Problems of H-CSI

From Research on the Decision-Making in Ill-Structured Situation Control and the Problem of Risks

N.A. Abramova[1] and S.V. Kovriga[2]

[1] Laboratory of Cognitive Modeling and Situation Control,
Institute of Control Sciences of the Russian Academy of Sciences, Moscow, Russia
abramova@ipu.ru
[2] Laboratory of Cognitive Modeling and Situation Control,
Institute of Control Sciences of the Russian Academy of Sciences, Moscow, Russia
maxi@ipu.ru

Abstract. In this paper basic directions of cognitive approach evolution in the field of formal methods of searching and making decisions in the control of complex and ill-structured situations are briefly reviewed. The problem of risks for the results validity that arise due to the human factor in the cognitive approach is considered and the conception of cognitive risks is proposed with two kinds of risks exposed. On the example of a real-life cognitive map, modeling a complex and ill-structured situation, practically significant risks of invalid formalization related to causal influence transitivity are demonstrated. Some explanatory mechanisms and criteria for the early detection of such risks are proposed. The issues important for further evolution of the cognitive approach to decision-making in the ill-structured situation control and especially of causal mapping techniques are lighted.

1 Introduction

At the practical decision-making based on formal models and methods and applied to complex and ill-structured situation control, formalization of representations of people (experts, analysts, decision-makers) about a situation, its problems, and even about people's goals and interests inevitably turns to be the essential stage of the problem solving. Therefore, in theoretical researches on formal problem solving methods it is spoken with increasing frequency, especially in recent decade, about the cognitive approach, cognitive researches, cognitive modeling or cognitive mapping of complex objects, problems, situations, even human representations (see [1-9] for an overview).

As the analysis shows, in such researches different understanding of the cognitive approach takes place as well as different are the basic problems and considered cognitive aspects. However, application and development of formalized normative models for representation of knowledge and activity of people solving practical problems is typical for the majority of works. To represent and create knowledge

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 3–15.
springerlink.com

about ill-structured situations, various models of cognitive maps widely extend, though also other normative models of people's knowledge and activity are referred to the cognitive approach.

Today it is possible to solve a spectrum of practical problems by means of cognitive maps, depending on a kind of applied cognitive map based models, on degree of formalization and accompanying formal methods to process maps. The spectrum is stretched from conceptual modeling aimed to help the individual to better organize, structure, and understand the problem and even to improve organizational action, up to building a shared understanding of the problem (see [4] for an overview), then to most typical simulation of situations optionally including their dynamics, and finally to the solution of some strategic management problems [10, 11].

To assure validity of practical results received at the problem solving based on the cognitive approach, complex interdisciplinary researches on the fundamental problem of risks due to the human factor are assumed to be necessary, especially for complex and ill-structured objects and situations. The importance of this problem is underestimated by scientific community, despite some known researches such as researches on "logic of failure" by Dörner [12] and on psychological correctness in the formal theory of decision-making by Larichev and his school [13], supported with the newest empirical researches [14-16]. On the other hand, this problem demands really cognitive approach taking account of scientific knowledge about cognitive aspects of problem solving activities in the considered domain.

These authors' researches on the problem of risks are directed to development of knowledge about risks and mechanisms of their action, practically significant for the problem solving in the control of complex and ill-structured situations [13, 17-23]. Attempts to integrate theoretical and experimental researches with the analysis of practical situations of problem solving based on the known normative models and methods are undertaken.

The new family of risks related to the property of causal influence transitivity and its violations has been discovered quite recently. These violations have been admitted as a hypothesis by analogy to known rationally reasoned violations of transitivity of pair-wise preferences [22] and later found out in real-life causal maps. These risks are analyzed on the practical example accompanied with some explanatory mechanisms and criteria for early detection of risks.

2 About Trends in Development of the Cognitive Approach in Decision-Making

Generally the cognitive approach may be defined as solving scientific problems with methods taking account of cognitive aspects, i.e. aspects concerning the cognitive sphere of the person.

In theoretical researches on models and methods for searching and making decisions, such terms as cognitive approach, cognitive researches, cognitive modelling or cognitive mapping of complex objects, problems, and situations are used more and more often. It may be explained by growing understanding of inevitable participation of people with their cognitive resources (and also restrictions) in the decision of practical problems of control, especially in case of complex and ill-structured situations.

Recently the selective analysis of papers proposed for International Conference "Cognitive analysis and situations evolution control" (CASC) has been carried out [6], with the following similar analysis of other publications using term "cognitive" in the context of solving the applied control problems. The conference has been held at the Institute of Control Sciences of Russian Academy of Sciences for last seven years, being focused on the integration of formal and cognitive problem solving methods.

The analysis [6] has allowed see considerable distinction in understanding of what terms *cognitive approach* and *cognitive modeling* mean, in how the term *cognitive* operates, in scientific and applied problems being solved, in the formalization level, in considered cognitive aspects, in involved knowledge of the cognitive science. At all distinctions, two overlapping basic directions may be identified in accordance with understanding of the cognitive approach in the narrow and wide sense (in the context of decision-making and ill-structured situations).

2.1 Two Directions of the Cognitive Approach

The cognitive approach in the narrow sense of the term, actively developing today, means that some or other models of cognitive maps are applied as models for representation and creation of knowledge about ill-structured situations. In the most conservative branch of this direction focused on formal methods specificity of human factors and features of structurization by the person of difficult situations is not considered at all; so the word "*cognitive*" carries out purely nominative function of a label for models applied. However as a whole it is relevant to speak about two trends in development of this direction. On the one hand, the positive tendency to larger account of such human-dependent stages as formalization of primary knowledge and representations of a problem situation, targets definition, etc is observed. On the other hand, the accepted models of knowledge and activity of people solving practical problems (experts, analysts, decision-makers) are normative in relation to these people, and the justification of the models is defined by theorists at the level of a common sense and traditions. Any knowledge of how people really think and what knowledge the cognitive science has got in this respect, today are usually not involved.

The cognitive approach in a broad sense is not limited by the choice of cognitive maps as models to represent knowledge about complex objects and situations [20]. The accent initially is made on human-dependent stages. Basically, the approach covers a wide spectrum of the models applied in decision-making for ill-structured problem situations. However today in this direction the same tendencies, as in the previous one, are dominating.

Review [6] represents perspectives of more advanced development of the cognitive approach, with integration of formal methods and knowledge of psychology and the cognitive science as a whole. However, today bridging the gap appears difficult, at least, because of distinction in scientific languages. Among few exceptions it is useful to mention [14-16, 24].

2.2 Clarification of Concept of the *Cognitive Map* in Decision-Making

For further treatment of the cognitive approach in the narrow sense, it is necessary to clarify the concept of *cognitive map* in the meanings used in the decision-making context.

In this field of knowledge the concept of cognitive map is used all more widely beginning from [25], but it takes various meanings, without saying about essentially differing concept of cognitive map in psychology. Recent years have brought a number of reviews and articles with extensive reviews in which the diverse types of cognitive maps and other concept maps are compared, differing in substantial and formal interpretation, as well as in sights at their role in searching and making decision processes and control of complex objects and situations (in particular [1-6, 8, 9]).

In this work term *cognitive map* refers to the family of models representing structure of causal (or, that is the same, cause-effect) influences of a mapped situation. Formally, the obligatory base of all models of the family is a directed graph, which nodes are associated with factors (or concepts) and arches are interpreted as direct causal influences (or causal relations, connections, links) between factors [5]. Usually the obligatory base (that is the cognitive map in the narrow sense) is added with some parameters, such as an influence sign ("+" or "–") or influence intensity, and some or other interpretations both substantial, and mathematical are given to the map.

Various interpretations of nodes, arcs and weights on the arcs, as well as various functions defining influence of relations onto factors result in different modifications of cognitive maps and formal means for their analysis [5]. Owing to multitude of cognitive map modifications, one can distinguish different types of models based on cognitive maps (in short, cognitive map models). Models of this family that are often referred to as causal maps or influence diagrams cover a wide spectrum of known types of models for cognitive mapping.

The problem of risks outlined below represents advanced approach in wide sense, with taking cognitive mapping as the representative example.

3 The Problem of Risks due to the Human Factor

The problem of risks due to the human factor in the field of formal methods for searching and making decisions in the control of complex and ill-structured situations essentially is that due to inevitable and substantial humans' participation in solving practical problems (at least, for formalization of primary representations) formal methods basically cannot provide validity of received decisions [21]. Note that validity of results of a method application is understood here in wide intuitive sense as capability to rely upon these results in solving a specific practical problem. It is also possible to speak about validity of a method as its capability to yield valid results. Simply speaking, such methods (which we refer to as subjective-formal ones) are basically risky concerning validity of their results.

The pragmatic importance of the given problem of risks obviously depends on how much significant are risks obtaining invalid results in solving practical problems. By present time theoretical, experimental and even practical knowledge is accumulated, leading to under-standing or directly saying that human factors can be the significant source of risk for quality of results.

The most impressing is the research on “logic of failure”, presented in D. Dörner's known book [12]. By means of vast computer-based simulation of how people solve problems of the complex and ill-structured dynamic situation control, a number of typical errors is revealed which are inherent not only in dilettantes but also in experts at work with complex situations. His results lead to conclusion that some of the above errors should be expected for any kinds of cognitive maps whereas others should be characteristic of dynamic map modeling. Strictly speaking, in Dörner’s experiments it is possible to admit that the risks of errors stem not only from natural ways of thinking but also from the accepted model for representation of knowledge about situations in the computer. However, Dörner has found evidences in favor of his theory in the analysis of known natural-thinking decisions concerning Chernobyl.

One more class of sources of risk is revealed by Larichev [13] with his school in the traditional formal decision theory. It refers to methods and operations of receiving initial information from decision-makers (in other terms, knowledge acquisition) for subsequent processing with formal methods. Stemming from the accumulated psychological knowledge of more than 30 years, the inference has been drawn that “soft” qualitative measurements such as comparison, reference to a class, ordering are much more reliable, than appointment of subjective probabilities, quantitative estimates of criteria importance, weights, usefulness, etc. To increase reliability of such operations (in our terms, to decrease risk of human errors) the idea of psychologically correct (in other terms, cognitively valid) operations has been advanced by the school. From the above results it follows that validity of cognitive modeling should be various for different kinds of cognitive maps depending on kinds of estimates demanded from experts.

There is also a wide spectrum of psychological researches in the field of the limited rationality of the person (not relating to solving control problems) which evidence to numerous types of risks in intellectual activity of the person.

To add evidences of practical significance of risk factors in decision-making, it is relevant to mention some significant factors found out by these authors in the applied activity on safety-related software quality assurance including formal-method quality estimation and control. This practice along with further theoretical analysis have considerably expanded representations of the risk spectrum produced by theorists developing or choosing models for knowledge representation and activity of people solving practical problems, in comparison with the risks exposed by Larichev's school [17, 18]. For example, it has appeared that application of very natural, from the mathematical point of view, method of linear convolution of normalized estimates on partial quality indicators for estimation of a complex indicator creates paradoxical risk of loss of controllability on partial indicators.

Moreover, the analysis has confirmed presence of some general cognitive mechanisms of risk for diverse models of subject-matter experts’ knowledge and related formal methods what serves evidence of reasonability of the general approach to the problem of risks due to the human factor in decision-making.

The subjective aspect of the problem of risks is that even more or less widely spread peaces of knowledge relative to risks and their sources mainly are not noticed by scientific community or, at the best, are underestimated, this ignorance being quite explainable theoretically with taking into account psychological risk factors (with the models presented in [18, 21]). Thereby, there is a soil for theoretical development and practical application of subjective-formal methods, which cannot provide decision quality (adequacy, validity, reliability, safety) comprehensible for critical domains.

Among few works relevant to the problem of risks due to the human factor in the cognitive approach it is worthwhile to note [16] where complexity of decision-making is considered as a risk factor in clinical medicine and use of decision support systems and [4] where validity of cognitive maps with internal validity between the data and the conceptualization of the data, including the definitions of concepts and influences, is designated as a research problem proceeding from the general ideas of content-analysis reliability.

3.1 Cognitive Risks in Subjective-Formal Searching and Making Solutions: Two Kinds of Risk Factors

At research of human-factor risks for validity of results in subjective-formal searching and making decisions, it seems appropriate to distinguish the special class of risks which are explainable with taking into account factors (mechanisms) concerning the cognitive sphere of the person. Such risks we will refer to as cognitive risks [26].

The cognitive sphere is defined today as the sphere of psychology of the person concerned with his cognitive processes and the consciousness, including knowledge of the person about the world and himself. The aspects of perception, attention, memory, thinking, explanation, understanding, language are distinguished more or less typically, though in different scientific approaches and schools of the cognitive science the cognitive sphere is structured differently. In particular, in the Rorschach system the cognitive sphere includes three principal frames: *structurization – recognition – conceptualization* with cognitive operations to provide information processing when facing a problem situation as well as a specific problem-solving style.

All risks mentioned above are related to cognitive risks as well as a number of others, which are reported about in publications and private communications without doing them subject of the scientific analysis.

First of all, factors are interested, which, on the one hand, have regular nature (i.e. do not concern to dysfunctional cognition), and on the other hand, are hardly explainable from positions of common sense or even contradict it; so usually they are not assumed by mathematicians working out normative models for expert knowledge representation or ignored if known.

It may be expected that all named cognitive aspects, anyhow, generate risk factors for quality of results at the combined application of formal methods and cognitive resources of people solving problems in the complex and ill structured situations. However our researches have shown that influence of human risk factors is not limited only to the people solving specific practical problems. Earlier cognitive processes have been analyzed in which subjects of intellectual activity (analysts, experts, decision makers and other staff who participates in searching and making decisions at collaboration with computers) turn to be under the influence of "ambient intelligence" due to imposed forms of thinking [18, 19, 21]. This influence leads to dependence of decisions on theoretical beliefs of specialists on formal methods and computer-aided decision support systems and technologies and therefore results in risks of invalid decisions. Two kinds of risk factors explainable with the suggested models have been exposed which pertinently to consider as cognitive risks.

The risk factors psychologically influencing validity of expert methods during their application by experts belong to first-kind factors, or factors of direct action. Such factors can either objectively promote invalidity of results, or raise subjective confidence of experts of objective validity of the method application outcomes. One can tell that the latter represent themselves as factors of belief. Agents of these factors of influence are experts; just they appear in conditions which may lead, eventually, to insufficiently valid (in the objective relation) outcomes.

Second-kind risk factors or factors of an indirect action psychologically influence upon validity of expert methods during their creation and justification. Agents of influence of such factors are creators of methods, scientists and experts producing standards who, in turn, are subject to influence of scientific norms, paradigms, etc., that is the strongest factors of belief. Typical examples of first-kind risk factors are natural mechanisms of thinking with risk of errors, as in case of the errors discovered by Dörner. Typical second-kind risk factors are psychologically incorrect models of knowledge of the experts, creating risk of unreliable (unstable or inconsistent) data from experts, with the models being supported by the factor of belief (often unconscious) in their validity.

Amongst a number of cognitive risk factors having been found out by these authors in theoretical and experimental researches as well in practice, there is belief in universality of the principle of pair-wise preference transitivity in decision-making [19, 22]. The principle means that from $a \succ b$ ("a is preferable over b") and $b \succ c$ it always follows $a \succ c$, though rationally reasoned violations of this principle are known (for example, preferences based on multicriteria comparisons).

Further this paper concerns one family of cognitive risks having been discovered quite recently and for the first time presented in [26]. Risks take place at cognitive mapping based on causal maps. They are related to the property of causal influence transitivity and its violations. They have been admitted as a hypothesis by analogy to known rationally reasoned violations of transitivity of pair-wise preferences and later found out in real-life causal maps.

4 A Family of Risks Concerned with Causal Influence Transitivity

The known principle of causal influence transitivity states that from $a \to b$ ("a is a reason for b") and $b \to c$ it follows $a \to c$. Accepted as an axiom at modeling based on causal cognitive maps, the principle serves as justification to the formal inference of indirect causal influences.

The most impressive example of violation of the causal influence transitivity principle, found out in practical cognitive-map-based modeling of a complex situation, is presented and analyzed below [26].

4.1 Practical Example of Causal Influence Transitivity Violation

In Fig. 1 the fragment of a real-life cognitive map slightly simplified to demonstrate action of risks is presented. Influence in pair of factors (3,2) at the verbal level is interpreted as follows: “increase in access of manufacturers to gas export pipelines (with other things being equal) causes increase in volume of extracted gas”. This influence is positive (in mathematical sense) that means the same direction of changes of factors. Positive influence in pair (4,2) is verbalized similarly. Influence in pair (2,1) is negative: “increase in volume of extracted gas (with other things being equal) causes decrease in deficiency of gas in the country”.

All three influences, as well as the set of factors significant for an investigated situation of dynamics of the market of gas in the country, are established by the expert. (Substantially, this map corresponds to a situation when there are stocks of gas and manufacturers have resources for increase gas production in volume but their access to means for its delivery to consumers is limited.)

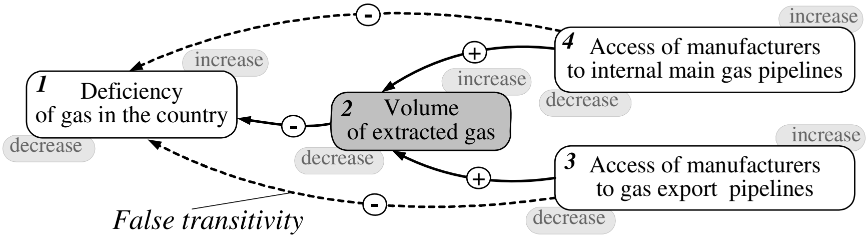

increase
increase
increase
increase
1 Deficiency of gas in the country
2 Volume of extracted gas
4 Access of manufacturers to internal main gas pipelines
3 Access of manufacturers to gas export pipelines
decrease
decrease
decrease
decrease
-
+
-
+
-
False transitivity

Fig. 1. An initial fragment of a cognitive map with false transitivity

According to formal model of causal influences and intuitive logic, from positive influence $3 \xrightarrow{+} 2$ and negative influence $2 \xrightarrow{-} 1$ follows transitive negative (in mathematical sense) influence $3 \xrightarrow{-} 1$; influence $4 \xrightarrow{-} 1$ is deduced similarly.

However later the expert has noticed that “logically deduced” influence $3 \xrightarrow{-} 1$ is absent in reality: thereby **false transitivity of influences** takes place in the map. The analysis of expert knowledge of a situation leads to following correction (fig. 2).

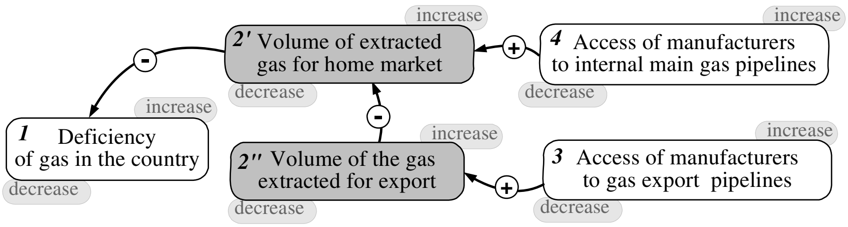

increase
increase
increase
increase
increase
2′ Volume of extracted gas for home market
4 Access of manufacturers to internal main gas pipelines
1 Deficiency of gas in the country
2″ Volume of the gas extracted for export
3 Access of manufacturers to gas export pipelines
decrease
decrease
decrease
decrease
decrease
-
+
-
+

Fig. 2. The corrected fragment of a cognitive map

It is worth while to underline that at such refinement replacements of influence $3 \xrightarrow{+} 2$ with $3 \xrightarrow{+} 2''$, $4 \xrightarrow{+} 2$ with $4 \xrightarrow{+} 2'$ and $2 \xrightarrow{-} 1$ with $2' \xrightarrow{-} 1$, in essence have not changed expert's interpretation of influences: the form of representation of knowledge has changed only. However, in this way the chain $3 \xrightarrow{+} 2 \xrightarrow{-} 1$ generating false influence $3 \xrightarrow{-} 1$ has disappeared.

Essentially other situation occurs with introduction of additional negative influence $2'' \xrightarrow{-} 2'$. This influence means that at increase of access of manufacturers to export gas pipelines (with other things being equal) it is possible to increase volume of the gas extracted for export not only by means of increase in volume of extraction, but also by simple "valve switching". Thus growth of volume of gas for export is made at the expense of decrease in volume of gas for home market. In this case knowledge is entered into a map, well known to experts, but not represented within the frame of initial system of concepts (factors).

As a result of correction new transitive influences $3 \xrightarrow{-} 2'$, $3 \xrightarrow{+} 1$ have appeared. Instead of positive (in substantial sense) situation in the map of Fig. 1, when it is possible to reduce deficiency of gas at the expense of access of manufacturers both to internal, and to external gas pipelines, more complicated and more realistic situation comes up in the map of Fig. 2. Along with positive (as a matter of fact) transitive influence $4 \xrightarrow{-} 1$, negative (as a matter of fact) influence $3 \xrightarrow{+} 1$ takes place, and their proportion at the decision of the problem of gas deficiency in the country demands comparative estimation of influences.

The invalidity of the fragment of cognitive map in Fig. 1 relative to reality, or in other words, error in cognitive mapping is obvious.

The above example from practice along with some others serve as actual evidences of that at cognitive modeling of complex situations there are possible cases of erroneous inferences by transitivity, i.e. false transitivity.

4.2 The Analysis of Revealed Cognitive Risks

In the above example the false transitivity can be explained with cumulative action of two modes of risk factors in the course of cognitive mapping: assumption of causal influence transitivity as the universal principle and disproportion of extension of concepts (Fig. 3).

First, in each of the three direct influences between factors of the initial map having been set by an expert, there is **disproportion of extension of concepts**

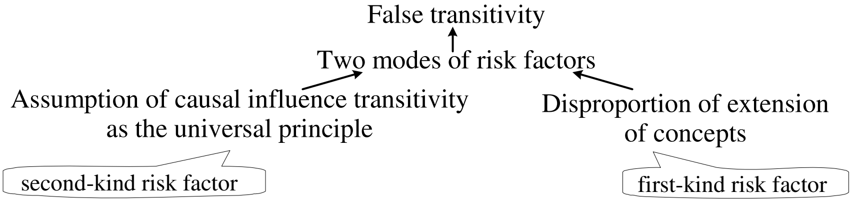

False transitivity
Two modes of risk factors
Assumption of causal influence transitivity as the universal principle
Disproportion of extension of concepts
second-kind risk factor
first-kind risk factor

Fig. 3. Two modes of risk factors in the course of cognitive mapping

with excess of extension of concept 2, which denotes the influence receiver in pairs (3,2), (4,2) and the influence source in pair (2,1).

Note that in cognitive mapping it is traditional to speak about factors (concepts) as causes and effects. However we prefer, at least in the analysis, to speak about sources and receivers of influences because at modeling of complex and ill-structured situations substantial cause-effect interpretations of individual influences in a map quite often happen more or less difficult. (There is such a situation in the above example.) Moreover, in the theoretical analysis which we spend it is more exact to distinguish "factors" and "concepts of factors" (that is concepts designating factors). It is relevant to speak about factors at the analysis of situation content, and more pertinently to speak about concepts of factors when it is a question of the logic analysis of concept extensions.

Excess of extension of concepts in some direct influences informally means that it would be possible to take concepts with smaller extension for mapping the same substantial cause-effect influences. Just this action has been made at correction.

It is hardly admissible to a priori consider such disproportions with excess of extension of concepts as errors because they are typical at the conceptualization of complex and ill-structured situations. This is evidenced both with practice of cognitive mapping and with informal reasoning of experts on such situations. Therefore we consider such disproportions only as cognitive risks. They are natural to be related to first-kind risk factors which are brought in by experts and which objectively reduce validity of cognitive modeling in complicated situations.

Assumption about the causal influence transitivity, taken by theorists as a principle for the formal modeling, due to belief in its totality, should be considered as the second-kind risk factor.

4.3 Some Criteria for the Early Detection of False Transitivity Risks

In the considered example false causal influence transitivity has been found out by backward tracing in view to explain doubtful, from the point of view of the expert analysis, results of formal modeling. Our analysis allowed to find some criteria which could help experts and analysts in early detection of risks and making decisions on possibility to correct disproportional concept extensions in case of false influence transitivity detection.

Let us more formally define these criteria. In the definition, the fact, which has been found out in practice, is taken into account: the same (as a matter of fact) causal influence may be represented in a cognitive map in different forms so that we speak about different representations of the influence. Let we have factors A, B_1 (represented with the same name concepts), which are linked by direct causal influence $B_1 \rightarrow A$, and let there exists (is found by an expert) factor B_2 such that replacement of representation of influence $B_1 \rightarrow A$ with $B_2 \rightarrow A$ does not change the influence substantially, and herewith

$$\mathcal{V}_{B_1} \supset \mathcal{V}_{B_2} \tag{1}$$

Here $\mathcal{V}_{B_i}$, i = 1,2, is a stand for the extension of the corresponding concept, and the relation between extensions is treated as usual inclusion or, that the same, verbally:

"C_1 has smaller extension than C_2". Then factor B_2 is **more proportional in its concept extension** then B_1 as the source in the direct influence on A, and factor B_1 is **extensionally excessive** in this influence.

The proposed expert criterion of extensional proportionality, K^S(A, B) is applicable to any pair of factors of a cognitive map, connected by direct influence. It allows to estimate whether factor-source of influence B is extensionally proportional to influence (or set of influences) on the receiver being modeled with link (B, A). For example, factor 2 in Fig. 1 is extensionally excessive in the influence (2,1), according to K^S(2,1). The criterion of extensional proportionality for the influence receiver K^D(A, B) is formulated and applied similarly, though in case of many influences onto one factor it is less informative at risk detection and error correction.

5 Conclusions

The problem of risks due to the human factor in the field of formal methods of searching and making decisions in the control of complex and ill-structured situations is considered as the general problem for diverse models of subject-matter experts' knowledge and related formal methods [20, 21].

Earlier the idea about productivity of the uniform approach to the problem for diverse models of experts' knowledge, solved problems and formal methods has been stated, and some theoretical and empirical evidences in favor of this idea have been found [21]. The idea has found the reinforcement and further development at current studying risks concerned to causal influence transitivity in cognitive modeling, with carrying out analogies to risks in decision-making based of pair-wise preferences. It is enough to tell that just the analogy of principles of transitivity of paired preferences and causal influences has led to a hypothesis about possible violation of the axiom of causal influence transitivity, i.e. to risk of false transitivity at formal cognitive mapping what has been confirmed in practice.

However, along with application of the uniform approach to the analysis of riskiness of those or other particular ideas, techniques, assumptions which are characteristic for diverse models and methods within the frame of the cognitive approach to the problem solving (whether it be a transitivity principle, or use of weights for various estimations or other techniques), the general approach is desirable to the analysis of cognitive risk factors as a whole.

It is essential that in search and development of such approach not only first-kind risk factors brought by people solving specific problems should be considered, but also second-kind factors brought by people who develop, theoretically justify formal models and methods for decision-making support, implement them in computer-aided technologies, support their application as intermediaries.

References

1. Eden, C.: Cognitive Mapping: A review. Eur. J. Oper. Res. 36(1), 1–13 (1988)
2. Kremer, R.: Concept mapping: informal to formal. In: Proc. Intern. Conference on Conceptual Structures. University of Maryland, MD (1994)

3. Mls, K.: From concept mapping to qualitative modeling in cognitive research. In: Proc. 1st Intern. Conference on Concept Mapping, Pamplona, Spain (2004)
4. Bouzdine-Chameeva, T.: An application of causal mapping technique ANCOM-2 in management studies. In: Proc 6th Global Conference on Business & Economics (GCBE). Harvard University, Cambridge (2006)
5. Kuznetsov, O., Kilinich, A., Markovskii, A.: Influence analysis in control of ill-structured situations based on cognitive maps. In: Abramova, N., Novikov, D., Hinsberg, K. (eds.) Human factor in control sciences. Collected papers, URSS, Moscow (2006)
6. Abramova, N.: On the development of cognitive approach to control of ill-structured objects and situations. In: Proc. 7th Intern. Conference on Cognitive Analysis and Situations Evolution Control. ICS, Moscow (2007)
7. Giordano, R., Mysiak, J., Raziyeh, F., Vurro, M.: An integration between cognitive map and causal loop diagram for knowledge structuring in river basin management. In: Proc. 1st Intern. Conference on Adaptive & Integrated Water Management, Basel, Switzerland (2007)
8. Peña, A., Sossa, H., Gutiérrez, A.: Cognitive maps: an overview and their application for student modeling. J. Comput. Sistemas 10(3), 230–250 (2007)
9. Avdeeva, Z., Kovriga, S.: Cognitive approach in simulation and control. In: Plenary papers, milestone reports & selected survey papers. The 17th IFAC World Congress, Seoul (2008)
10. Maximov, V., Kornoushenko, E.: Analytical basics of construction the graph and computer models for complicated situations. In: Proc. 10th IFAC Symposium on Information Control Problems in Manufacturing, Vienna (2001)
11. Avdeeva, Z., Kovriga, S., Makarenko, D., Maximov, V.: Goal setting and structure and goal analysis of complex systems and situations. In: Proc. 8th IFAC Symposium on Automated Systems Based on Human Skill and Knowledge, Göteborg, Sweden, pp. 889–903 (2003)
12. Dorner, D.: The logic of failure: recognizing and avoiding error in complex situations. Perseus Press, Cambridge (1997)
13. Larichev, O., Moshkovitch, E.: Qualitative methods of decision making. Physmatlit, Moscow (1996)
14. Obal, L., Judge, R., Ryan, T.: The influence of cognitive mapping, learning styles and performance expectations on learning effectiveness. In: Proc. AIS SIG-ED IAIM Conference, Montreal, Canada, pp. 1–20 (2007)
15. Gabriel, J., Nyshadham, E.: A cognitive map of people's online risk perceptions and attitudes: an empirical study. In: Proc. 41st Hawaii Intern. Conference on System Sciences, Big Island, p. 274–274 (2008)
16. Sintchenko, V., Coiera, E.: Decision complexity affects the extent and type of decision support use. In: Proc. AMIA Annual Symposium, pp. 724–728 (2006)
17. Abramova, N.: About some myths in software quality estimation. J. Reliability 1, 38–63 (2004)
18. Abramova, N.: A subject of intellectual activity under cognitive control of ambient intelligence. In: Proc. 9th IFAC Symposium on Automated Systems based on Human Skills and Knowledge, Nancy, France (2006)
19. Abramova, N., Kovriga, S.: On risks related to errors of experts and analysts. J. Control Sciences 6, 60–67 (2006)

20. Abramova, N., Novikov, D.: Evolution of representations about the human factor in control science. In: Abramova, N., Novikov, D., Hinsberg, K. (eds.) Human factor in control sciences, Collected papers, URSS, Moscow (2006)
21. Abramova, N.: On the problem of risks due to the human factor in expert methods and information technologies. J. Control Sciences 2, 11–21 (2007)
22. Abramova, N., Korenyushkin, A.: Another approach to intransitivity of preferences. In: Proc. 22nd European Conference on Operational Research, Prague, pp. 485–490 (2007)
23. Abramova, N., Avdeeva, Z., Kovriga, S.: Cognitive approach to control in ill-structured situation and the problem of risks. In: Aramburo, J., Trevino, A. (eds.) Advances in Robotics, Automation and Control. IN-TECH, Vienna (2008)
24. Vavilov, S.: Psychological space of managerial decisions. J. Sociological Investigations 5, 93–102 (2006)
25. Axelrod, R.: The Cognitive Mapping Approach to Decision Making. In: Axelrod, R. (ed.) Structure of decision. The cognitive maps of political elites. Princeton University Press, Princeton (1976)
26. Abramova, N., Kovriga, S.: Cognitive approach to decision-making in ill-structured situation control and the problem of risks. In: Proc. IEEE Conference on Human System Interaction, Cracow, Poland, pp. 485–490 (2008)

Emulating the Perceptual System of the Brain for the Purpose of Sensor Fusion

R. Velik, D. Bruckner, R. Lang, and T. Deutsch

Institute of Computer Technology, Vienna University of Technology, Vienna, Austria
{velik,bruckner,lang,deutsch}@ict.tuwien.ac.at

Abstract. This article presents a bionic model derived from research findings about the perceptual system of the human brain to build next generation intelligent sensor fusion systems. For this purpose, a new information processing principle called neuro-symbolic information processing is introduced. According to this method, sensory data are processed by so-called neuro-symbolic networks. The basic processing units of neuro-symbolic networks are neuro-symbols. Correlations between neuro-symbols of a neuro-symbolic network are learned from examples. Perception is based on sensor data and on interaction with cognitive processes like focus of attention, memory, and knowledge.

1 Introduction

The human brain is a highly complex system that is capable of performing a huge range of diverse tasks. Over millions of years, its structure and information processing principles have undergone a development and optimization process through variation and selection. One capability of the brain is to process information coming from thousands and thousands of sensory receptors and integrating this information into a unified perception of the environment.

Up to now, technical systems used for sensor fusion and machine perception can by far not compete with their biological archetype. Having available a technical system, which is capable of perceiving objects, events, and scenarios in a similar efficient manner as the brain does, would be very valuable for a wide range of applications in automation. Examples for applications are automatic surveillance systems in buildings, interactive environments, and autonomous robots.

To perceive objects, events, and scenarios in an environment, sensors of various types are necessary. The challenge that has to be faced for perceptive tasks is the merging and the interpretation of sensory data from various sources. The aim of this paper is to introduce a model for integrating and interpreting such data. As humans can perceive their environment that effectively, the perceptual system of the brain is taken as archetype for model development. Particularly, research findings from neuroscience and neuro-psychology are the guides in the development process.

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 17–27.
springerlink.com

2 The Research Field of Sensor Fusion

Among the various attempts to merge information from various sensory sources, sensor fusion is the most prominent one. Despite its prominence, several different definitions can be found in literature. All definitions share that sensor fusion deals with the combination of data – from sensors directly or derived from sensory data – to generate an enhanced internal representation of the observed process environment. According to [6], the most important achievements of sensor fusion are robustness, extended temporal and spatial coverage, increased confidence, reduced uncertainty and ambiguity, and improved resolution.

Due to the dynamics of the relatively young research field of sensor data fusion, not only different definitions exist, but the standard terminology has not yet evolved. In [1] an overview of widely used terms is given. Among them are "sensor fusion", "sensor integration", "data fusion", "information fusion", "multi-sensor data fusion", and "multi-sensor integration."

Data for sensor fusion can origin in three different sources: multiple measurements subsequently at different points in time from one single sensor, from multiple identical sensors, or from different sensors. To ease data processing, several attempts have been made to transform sensor data into symbols. An often used approach is to use a layered architecture for merging and symbolization of sensor data. Such systems are described in [2, 12, 16, 20]. Sensor fusion using data from the above mentioned three sources is generally called direct fusion. If knowledge is added to the process, it is called indirect fusion. Examples for knowledge are information about the environment and human input. A hybrid approach – different models are described in [1, 3, 5, 8, 18] – would be to fuse the outputs of direct and indirect fusion.

For sensor fusion itself, different models have been proposed. They all are strongly related to a special application. The demands originating in different domains are varying strongly. Thus, many researchers point out that a unified architecture or technique is very unlikely [7].

An approach to create sensor fusion models is to derive them from biology – which is called biological sensor fusion. One possibility is to use neural networks as starting point (see [4] and [15]). As argued in [15], it is generally accepted that sensor fusion in the perceptual system of the human brain is of far superior quality compared to sensor fusion approaches realized with existing mathematical methods. Thus, studying the capabilities of the human brain in context of sensor fusion is at hand. Not only are they maybe leading to better technical models for sensor fusion, but they also offer the possibility to gain new insights on how perception is performed in the brain.

Applications for fusion are various and range from measurement engineering and production engineering over robotics and navigation to medicine technology and military applications [13, 17].

3 Characteristics of Human Perception

The new sensor fusion model presented in this article is of the machine perception domain. It is based on scientific theories about the perceptual system of the human

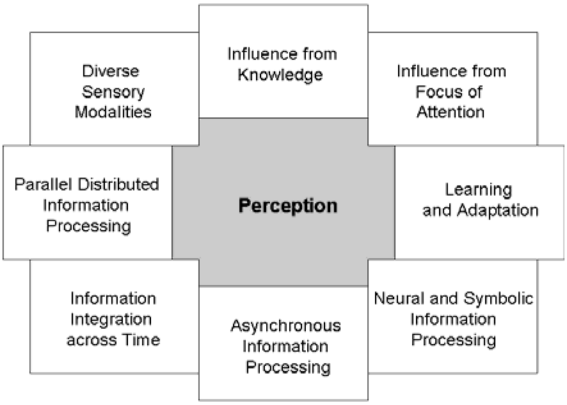

Diverse Sensory Modalities
Influence from Knowledge
Influence from Focus of Attention
Parallel Distributed Information Processing
Perception
Learning and Adaptation
Information Integration across Time
Asynchronous Information Processing
Neural and Symbolic Information Processing

Fig. 1. Characteristics of human perception

brain. Figure 1 gives an overview about the characteristics – important mechanisms and influence factors – of human perception. This list is derived from related research results of the scientific fields of neuroscience and neuro-psychology [19].

The eight identified characteristics of human perception are:

- **Diverse Sensory Modalities:** To perceive information from the external environment, our brain has access to different sensor modalities in multiple instances. The modalities include vision, touch, and audition. The use of multiple sources with different sensor modalities is the key to robust perception.
- **Parallel Distributed Information Processing:** The perceptual system is not a unitary central unit. As mentioned above, information from various sources is processed. This happens distributed and parallel.
- **Information Integration across Time:** As outlined in the previous chapter, one possibility for sensor fusion is integration across time. In human perception this approach is used to perceive objects, events, and scenarios. A single-moment snapshot of sensory information divided by all the different sources and modalities is often insufficient for unambiguous perception.
- **Asynchronous Information Processing:** The fact that the human brain uses parallel distributed information processing from multiple sources leads towards the next characteristic: asynchronous processing of the perceived information. This already starts at the sensory levels. For example, one event occurring in the environment does not necessarily trigger sensory receptors of different modalities absolutely concurrently. Also, different modalities work with different speed for information processing and transmission.
- **Neural and Symbolic Information Processing:** On the sensor level, perception is processed by interacting neurons. According to research findings, humans do not think in terms of action potentials and firing nerve cells – they think in terms of symbols. Mental processes are often considered as a process of symbol manipulation.

- **Learning and Adaptation:** At birth, only the most basic patterns are predefined by the genetic code. Hence, the not fully developed perceptual system of the human brain at birth needs to be trained during lifetime. The training includes lots of concepts and correlations concerning perception.
- **Influence from Focus of Attention:** Focus of attention is a hypothesis which states that what we see is determined by what we attend to. The environment presents far more information at every moment than what can be effectively processed. Attention helps bypassing this bottleneck by selection of relevant information and ignoring irrelevant or interfering information. To avoid processing all objects simultaneously, processing is limited to one object in a certain area of space at a time.
- **Influence from Knowledge:** Applying the concept of sensor fusion, human perception is facilitated by knowledge. It is often necessary to access prior knowledge to be able to interpret ambiguous sensory signals. Much of what we take for granted as the way the world is – as we perceive it – is in fact what we have learned about the world – as we remember it. Much of what we take for perception is in fact memory.

4 Bionic Model for Perception

This chapter introduces the bionic model for perception, based on the eight characteristics outlined in Chapter 3. Recent research findings of the organizational structure and the information processing principles in the human perceptual system are taken as archetype for the development of the model. Not all details of the working principles and functions of the human brain are already understood – the human brain is a highly complex system. In the case of insufficient research findings, for the construction of a functioning and implementable technical system, the model is supplemented by engineering considerations that fit into the overall concept.

4.1 Neuro-Symbolic Information Processing

To fulfill the first six characteristics described in the last chapter, a new concept of information processing is introduced - the *neuro-symbolic information processing*. This concept will be described in the following section.

Neuro-Symbols as Basic Processing Units

The main component of neuro-symbolic information processing is called *neuro-symbol*. The idea for development came from the consideration of interconnected neurons: It is generally accepted that information in the human brain is processed by interconnected neurons. Further it is assumed, that humans do not think in terms of action potentials and firing nerve cells, but in terms of symbols. In the theory of symbolic systems, the mind is defined as a symbol system and cognition is only symbol manipulation. Symbols can be objects, characters, figures, sounds, or colors used to represent abstract ideas and concepts. Each symbol is associated with other symbols. The symbol manipulation offers the possibility to generate complex behavior [9].

With respect to these basic assumptions, neurons are basic information processing units on a physiological basis and symbols are information processing units on a more abstract level. Within the brain, neurons have been found, which respond exclusively to certain perceptual images. It has been shown, that neurons in the secondary visual cortex respond exclusively to the perception of faces. It can be assumed, that a connection exists between neurons and symbols. For technical purpose, this connection is represented by neuro-symbols. Neuro-symbols represent perceptual images that can represent a face, a person, or a voice. A neuro-symbol contains an individual activation grade and is activated if the perceptual image that it represents is perceived in the environment. A neuro-symbol has a certain number of inputs and only one output (see figure 2).

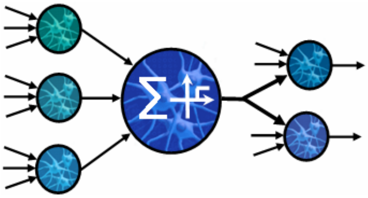


Fig. 2. Function principle of neuro-symbols

The incoming information represents the activation grade as an output of other neuro-symbols and triggered sensory receptors. The activation grades of the incoming neuro-symbols are summed up. If this sum exceeds a certain threshold, the neuro-symbol is activated and its activation grade is transmitted via the output to other neuro-symbols that are connected. To process also asynchronously arriving input data, a mechanism has been implemented that allows the processing of input data arriving within a certain time window or to consider certain successions of incoming data.

Neuro-Symbolic Networks

One single neuro-symbol is not designed to perform complex tasks. The potential of the system emerges when a certain number of neuro-symbols are interconnected. However, the way of structuring these neuro-symbols is most crucial. Once again, the structural organization of the perceptual system of the brain is taken as an archetype for this alignment. According to [14], the perceptual system of the brain is organized in cerebral layers as depicted in Figure 3.

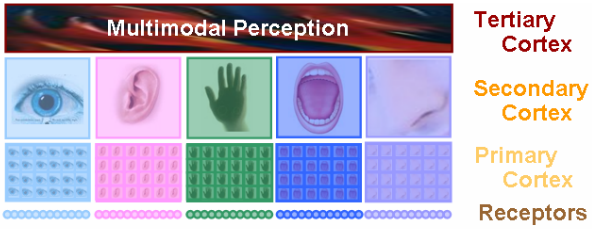


Fig. 3. Layered cerebral organization of human perceptual system

The perception during wake life starts with information coming from sensory receptors (during dreaming, the primary cortex is inactive – the secondary cortex is stimulated by higher brain functions). After the receptors, this information is processed in three levels: the primary cortex, secondary cortex, and tertiary cortex. For each sensory modality of human perception exists an own primary and secondary cortex. In the first two levels, information of different sensory modalities is processed separately and in parallel. The tertiary cortex merges information coming from all sensory modalities. The result is a unified multimodal perception. A perceptual image in the primary cortex of the visual system contains simple features like edges, lines, or movement. Information processing in the primary cortex of the auditory system can be for example sounds of a certain frequency. A perceptual image of the secondary cortex of the visual system could be a face, a person, or an object. A perceptual image in the acoustic system on this level would be a melody or a voice. One task, performed in the tertiary cortex is to merge the perceptual visual image of a face and the perceptual auditory image of a voice to the perception that a person is currently talking. The somatosensory system of the brain (commonly known as tactile system) comprises in fact a whole group of sensory systems, including the cutaneous sensations, proprioception, and kinesthesis.

As shown in figure 4, neuro-symbols are structured to so-called *neuro-symbolic networks* that are representing this structural organization of the perceptual system of the brain. Although the functional principle of each neuro-symbol follows the same principle, a neuro-symbol from a lower layer fulfills different functions than one of a higher layer. Therefore, they are labeled differently.

In the technical implementation, the raw sensory data is processed into so called feature symbols. This first layer corresponds to the primary cortex of the brain and gets the sensory data from available technical sensors that might or might not have an analogy in the human perceptual senses like video cameras, microphones, tactile sensors, chemical sensors.

Feature symbols are then combined first to sub-unimodal and finally to unimodal symbols. These two levels are corresponding to the function of the secondary

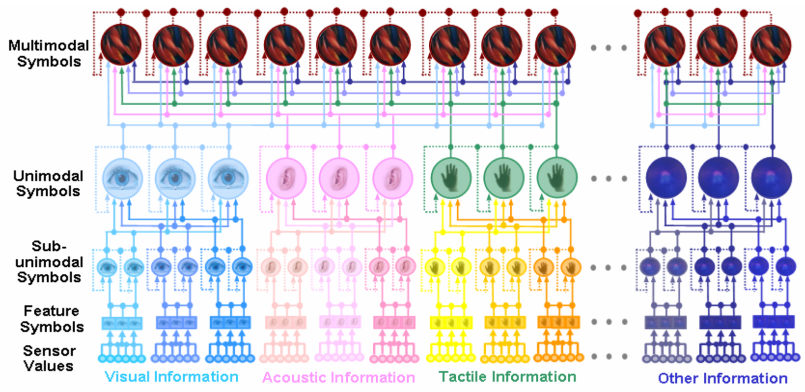


Fig. 4. Structure of a neuro-symbolic network

cortex of the brain. As with the somatosensory system of the brain, a sensory modality can consist of further sub-modalities. Similarly, a sub-unimodal level can exist between the feature level and the unimodal level. By processing information in these two steps, different sensor types, for example different video cameras, cameras mounted at different positions, or different tactile sensors like floor sensors, motion detectors, and light barriers, can be merged to one unified modality.

On the topmost level, the multimodal level, all unimodal symbols are merged to multimodal symbols. Examples and concept clarifications concerning the usage of neuro-symbols for concrete perceptive tasks can be found in [19].

Learning in Neuro-Symbolic Networks

Very similar to artificial neuronal networks, a great part of the information in the proposed model is stored in the connections between the neuro-symbols and not in the neuro-symbols themselves. It therefore has to be defined, how they are connected and what information has to be exchanged. Again, research findings from neuroscience can be consulted. As outlined in [14], higher cortical layers of the brain can only evolve if lower levels were already developed and certain correlations have to be innate – it can be said, they have to be predefined by genes. In the technical system, the connections were also defined in respect to the restrictions of this description. The lowest level of neuro-symbolic connections are therefore predefined and at system startup it is defined what feature symbols shall be extracted from the sensor data and in certain cases also how these feature symbols are built to form sub-unimodal symbols. In contrast to the lowest layers, correlations between higher layers are generally learned from examples. The system has no existing connection between the sub-unimodal layer and the unimodal level and no connections between the unimodal level and the multimodal level at initialization established. The connections are established by applying a supervised learning principle. A number of examples were shown to the system, which cover all objects, events, and scenarios the system shall perceive. With this procedure connections between sub-unimodal symbols and unimodal symbols are set in a first stage. After this procedure, the correlation between the unimodal symbols and the multimodal symbols are calculated in a next cycle. Objects, events, and scenarios are considered as perceptual images and each of them is assigned to one particular neuro-symbol. A number of examples is necessary to train a neuro-symbol to this assignment, because there can occur deviations in the representing sensory data - a generalization over these data is necessary. A detailed description about the used learning principle can also be found in [19].

4.2 Interaction between Neuro-Symbolic Network, Knowledge, and Focus of Attention

The above described processing structure as the core of the perception model covers bottom-up information processing starting from sensor values. However, perception is influenced and modified also by mechanisms called *knowledge* and *focus of attention* (Fig. 5), both corresponding to the last two points mentioned in Chapter 3. They are partly performed in brain areas with functions not being primarily dedicated to perception.

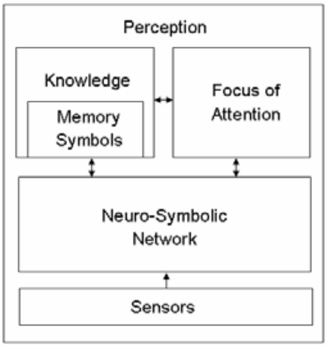


Fig. 5. Mechanisms involved into perception

Influence from Knowledge

According to [10], perception is influenced by knowledge in a top-down manner, where knowledge can be further categorized into factual knowledge, knowledge about the context within which a situation occurs, past-experience of what happened before, and expectations. The advantage of integrating knowledge into the perception process is that perception can be assisted and ambiguous sensor data can be resolved. Unfortunately, there is no global agreement between neuroscientists and neuro-psychologists on how and on what levels knowledge influences perception. In this model, knowledge influences the activation grade of affected neuro-symbols. Whereas lower levels (sensor values and feature symbols) are not influenced, the interaction can principally take place on the sub-unimodal, unimodal, or the multimodal level. Based on this cognitive information the activation grade of neuro-symbols in these layers can be increased or decreased. Imagine for example a monitored room that is empty at the beginning. Now, the system can "know" by using it's knowledgebase that certain situations cannot occur without the attendance of a person. Therefore, neuro-symbols, which are correlated with activities performed by persons, cannot be activated. The event that a person enters a room only takes a very short time and the corresponding neuro-symbol is also activated only for a brief moment. Therefore, a mechanism has to be provided that memorizes that a person entered the room. This activity would have a highly influence within the given example on future perception. In the proposed model, *memory symbols* are created to perform this task. They are interacting with knowledge and can principally receive information from neuro-symbols, coming from the sub-unimodal, unimodal, or the multimodal level. They are set when a certain event happens in the environment and reset when another, different, event happens. In the example of an entering person, a memory symbol "person is present in the room" would be set after a person enters the room and would be released when the person has left the room again.

Influence from Focus of Attention

In environments with several situations happening in parallel, it may not be possible to assign lower level neuro-symbols unambiguously to higher level neuro-symbols.

This circumstance also occurs in human perception at times of overloaded perceptions, where too many perceptual stimuli are present at one moment to integrate them all at once into a unified perception. In this case, illusionary conjunctions can occur. In the brain, a mechanism called focus of attention helps to correctly merge information coming from various sources. This mechanism has also been realized in the technical model. It constraints the spatial area, in which perceptual images of different modalities are merged. In the model, focus of attention interacts with perception on the feature symbol level. This is the level where neuro-symbols are already topographic and therefore location dependent. The activation grades of neuro-symbols are decreased in a way that they fall below the threshold value if they correspond to perceptions lying outside the focus of attention. When the neuro-symbols are no longer active the information is no longer transmitted and further processed until the focus of attention is directed back towards them. The decision towards which the focus of attention is directed is based on information coming from the neuro-symbolic network and the knowledge module. It depends on the currently perceived data and the knowledge of what is important for the particular application. The knowledgebase has to be defined with respect to the specific application field.

5 Results and Discussion

What can principally be perceived by a technical system depends on what sensor types and how many of them are used, because this influences the amount of perceptual information that is available for processing. If only a few sensors are available, there cannot be made such a fine distinction between many different objects, events, and scenarios to be detected. If more sensors are used, a better distinction can be made and – with an adequate processing mechanism – the system gets a certain degree of fault tolerance against sensor failures and false detections. However, if the amount of sensors is further increased, this does not necessarily lead to an increase of the number of objects, events, and scenarios that can be detected and differentiated, because also one and the same object, event, and scenario does not always activate exactly the same sensors. Therefore, the more sensors are used, the more the system needs the ability to generalize over data. Furthermore, the more sensory information has to be processed, the more processing power is needed. An "intelligent" mechanism is necessary to extract the currently relevant information from the flood of information being available each instant of time.

To test and evaluate the model presented in this article, it was simulated in AnyLogic. For the provision of sensor data, a simulator was used, which allows it to generate sensor values based on a virtual environment [11]. This simulator was developed to simulate sensor values in order to perceive scenarios in a virtual office environment. The reason for simulating the sensor values is on the one hand the cost reduction for testing in comparison to real physical installations. On the other hand, the simulator allows the comparison of different sensor configurations and an evaluation of the usability of the suggested model for the fusion of data from these sensor configurations.

The developed model proved to be a very efficient architecture for sensor data fusion of very distinct sensor configurations. In case that only few sensor data are available, perception can be facilitated considering stored knowledge about the environment and ambiguous situations being based on the same sensor data can in many cases be resolved that way.

By its parallel, modular, hierarchical neuro-symbolic architecture, the model is also capable of handling a very large amount of sensor data in a fast and efficient manner. What additionally increases processing speed is the fact that neuro-symbols are at the same time information processing and information storing units, which saves time for external memory access and comparison operations. An additional mechanism for increasing the amount of information that can be handled is the focus of attention. Stored knowledge also proved to be an important mechanism for perception in large sensor configurations to decrease the number of fault perceptions.

One further problem of common sensor fusion systems is that they are dedicated to a very particular application and it has to be defined in a detailed form how to fuse the sensory data and which combinations of sensor data have what meaning. Assuming a large number of different sensors, this results in a complex task to be solved by the system engineer and requires much time and very detailed knowledge about the system. In the proposed model, this problem is solved by introducing learning of neuro-symbolic correlations into the system. This allows great flexibility and strongly reduces the design effort. Furthermore, the used learning algorithm offers the ability to generalize over examples presented to the system to also correctly classify sensor data not seen before.

6 Conclusions

This paper suggests a model for emulating the perceptual system of the human brain to build next generation intelligent sensor fusion systems. Therefore, a new information processing principle was introduced called neuro-symbolic information processing. According to this method, sensory data are processed by neuro-symbolic networks to result in “awareness” of what is going on in an environment. The basic processing units of neuro-symbolic networks are neuro-symbols. Besides bottom-up information processing of sensory data, focus of attention and knowledge influence the perceptive process. The proposed model provides a powerful and flexible tool for information processing and fusion of sensor data from various sources capable of handling the demands of the upcoming future. It is a universal sensor fusion architecture, which is adaptable to many different application domains.

References

1. Beyerer, J., León, F.P., Sommer, K.: Informationfusion in der mess- und sensortechnik. Universitätsverlag, Karlsruhe (2006)
2. Burgstaller, W.: Interpretation of situations in buildings. PhD dissertation, Vienna University of Technology (2007)

3. Datteri, E., Teti, G., Laschi, C., Tamburrini, G., Dario, P., Guglielmelli, C.: Expected perception: An anticipation-based perception-action scheme in robots. In: Proc. Conference on Intelligent Robots and Systems, pp. 934–939. Las Vegas, NV (2003)
4. Davis, J.: Biological sensor fusion inspires novel system design. In: Proc. Conference on Joint Service Combat Identification Systems, San Diego, CA (1997)
5. Ellis, D.P.W.: Prediction-driven computational auditory scene analysis. Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA (1996)
6. Elmenreich, W.: Sensor fusion in time-triggered systems. Ph.D. dissertation, Vienna University of Technology (2002)
7. Elmenreich, W.: A Review on system architectures for sensor fusion applications. International Federation for Information Processing, pp. 547–559 (2007)
8. Ernst, M.O., Bülthoff, H.H.: Merging the senses into a robust percept. TRENDS in Cognitive Sciences 8(4), 162–169 (2004)
9. French, R.M.: Review of the engine of reason, the seat of the soul. Minds & Machines 6(3), 416–421 (1996)
10. Goldstein, E.B.: Wahrnehmungspsychologie. Spektrum Akademischer Verlag (2002)
11. Hareter, H., Pratl, G., Bruckner, D.: A simulation and visualization system for sensor and actuator data generation. In: Proc. 6th Conference on Fieldbus Systems and their Applications, pp. 56–63 (2005)
12. Joyce, D., Richards, L., Cangelosi, A., Coventry, K.R.: On the foundations of perceptual symbol systems: specifying embodied representations via connectionism. In: Proc. 5th Conference on Cognitive Modeling, Bamberg, Germany, pp. 147–152 (2003)
13. Luo, R.C., Kay, M.G.: Multisensor integration and fusion in intelligent systems. IEEE Transactions on Systems, Man, and Cybernetics 19(5), 901–931 (1989)
14. Luria, R.: The working brain: an introduction in neuropsychology. Basic Books, New York (1973)
15. Perlovsky, L.I., Weijers, B., Mutz, C.W.: Cognitive foundations for model-based sensor fusion. In: Proc. of SPIE Conference, vol. 5096, pp. 494–501 (2003)
16. Pratl, G.: Processing and symbolization of ambient sensor data. PhD dissertation, Vienna University of Technology (2006)
17. Ruser, H., León, F.P.: Informationfusion - eine übersicht. Technisches Messen Oldenbourg Verlag 74, 93–102 (2007)
18. Sillanpää, J., Klapuri, A., Seppäne, J., Virtanen, T.: Recognition of acoustic noise mixtures by combined bottom-up and top-down processing. In: Proc. 10th Conference on European Signal Processing EUSIPCO, Tampere, Finland, vol. 1, pp. 335–338 (2000)
19. Velik, R.: A bionic model for human-like machine perception. PhD dissertation, Vienna University of Technology (2008)
20. Bruckner, D.: Probabilistic models for building automation: recognizing scenarios with statistical methods. PhD dissertation, Vienna University of Technology (2007)

Knowledge Acquisition in Conceptual Ontological Artificial Intelligence System

M. Krótkiewicz and K. Wojtkiewicz

Institute of Mathematics and Informatics, University of Opole, Opole, Poland
{mkrotki,kwojt}@math.uni.opole.pl

Abstract. The paper deals with active knowledge acquisition based on dialogue inter AI system and its user. Presented method uses Conceptual Ontological Object Orientated System (COOS) to distinguish differences between concepts and to unequivocally process the input stream. In case of concepts, that do not exist in the system yet, adequate algorithms are being used to position them in ontological core. Separate concepts differ in attributes values or in sets of direct connections with other concepts. The communication aspects of the system deliver information that allow generating proper interpretation for user's statement.

1 Introduction

Presented knowledge acquisition unit is part of conceptual ontological object orientated artificial intelligence system (COOS). Knowledge in this system is represented on different levels of abstraction. In ontological core one can find information about essential concepts attributes and about possible and forbidden connection in between individual concept grouped in classes: association, relation, class, feature, value, data type, object. Facts and rules base is being built in help of semantic network connections between specific instances of concepts. Architecture of the system is compatible with assumption made by Sergei Nirenburg and Victor Raskin in their studies [1]. Data flow between each and every unit of COOS is implemented in the way of internal specialized language. Translation unit convert natural language statements into systems meta-language (COOL – Conceptual Ontological Object Language). After potential specification of the statement, it is used to fill the ontological knowledge stored in ontological core and to generate semantic network connections when needed.

2 Method

2.1 Conversion from Natural Language

COOS is communication with users by natural language interface. Due to this fact we may assume that it is gathering statements as series of words [2]. An example

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 29–37.
springerlink.com

will be drawn using symbols to prevent any kind of suggestion that system has any information about the words or context:

word_1 word_2 word_3 – exemplary statement

Above statement is converted into internal system language (COOL). Grammar and semantic of this language is strictly connected with structure of the ontological core, but due to limitations of this article this won't be presented here fully[3]. Translation form natural language is based on linguistic algorithms. For each inscription potential part of speech (the verb, the noun, the adjective, the adverb, the pronoun, the conjunction, the interjection, numerals, the preposition) is associated, which is then used to identify its function in the sentence(subject, predicate, object, complements, phrases, clauses). Word with associated parts of speech and sentence function is a basis to associate it with meaning – concepts stored in ontological core. It needs to be pointed out that one meaning may be represented by different words or group of words, as well as one word may be connected with different meanings in the ontological core. The next step in the process is building statement in system's meta-language, taking under consideration its grammar and limitations derived from ontological core. In case that determining any of information needed in the process of translation would be impossible or ambiguous, there will be suitable question generated by the system. This is one of the earliest phase of active knowledge acquisition.

2.2 Polysemy of Natural Language Statements

The most vital aspect of natural language processing that affect ambiguity of words is assigning particular concept from ontological core to the inscription (word). Appropriate word base is used to implement module responsible for this. As an inscription we understand series of symbols and the concept will be abstract description of existence. One inscription may be connected to many concept what leads to ambiguity. Closer and further context analysis is the most common method used to choose the most adequate association of word and its meaning. It is effective but unfortunately it is not foolproof. The easiest way to show possible problems in usage is assigning meaning to the word 'pot' in the sentence: ***Fred tried to hide his pot.***

Depending on context word 'pot' has the meaning of cooking accessory, flower holding device, overgrown stomach and finally marijuana. All of those meanings are commonly used in American English. In the example we may not judge which meaning is the most suitable. Each of them is equally possible. The analysis of further context would be helpful but probably wouldn't give any reasonable clue. Analysis of closer context is useless at all. Presented method base on possible and forbidden connections as well as let us identify the most often used connection of all. It is not simple translation but interpretation of the sentence. System would try to find all possible concepts associated with word 'pot' and assign each of them rate of probability based on its usage with corresponding word 'hide'. Flower container is rather rarely hidden by someone, that's why this meaning would get the lowest score. On the other side would be marijuana, which in most case is being connected with actions like hiding. The other two meaning would be placed in between those two. Translation module may assume it is about marijuana, but since other concepts might be used as well, it produce a question that ask which meaning is the best one to use in this example. Automatic choice would be not only risky but in wider sense pointless. System would start to promote knowledge just by statistic matter what in natural language processing is the fastest way to fail. Dialogue between human and computer is not only possible and needed but is the basis of knowledge acquisition and processing.

2.3 Ontological Core Knowledge Acquisition

If particular word has no connection with concepts in ontological core, one has to assume that the concept is totally unknown for the AI system. In this case **system** should generate questions necessary to create new instance of concept in the base.

The very first step is to determinate class of concept that the word is corresponding to. The choice is made from: class, association, relation, feature, value, data type, object and data. According to this choice the list of attributes possible to determinate is populated. Taking word 'car´ under consideration, we connect it to the concept class. After this choice, we are asked to determinate whether this concept is material (yes), enumerable (yes), collective (no), animated (no). Another kind of information is the possibility to connect this concept to others. For concepts assigned to class we may create list of object of particular class, and the list of features used to describe the class. It is also possible to assign specific relations possible to connect to the examined class. There are also attributes inherited from class C_CONCEPT. First group of attributes cover mainly descriptive values of name, symbol and comment. Second group let us determinate emotional and abstract levels of concept. Another group is used to show object orientated specific of the concept – generalization and specialization are used to build list of concepts. One can also decide whether list of specifying classes is complete or it may be changed. Other group is used to show relation of the concept in the knowledge in wider sense – relations of holonym and meronym, lists of synonyms and antonyms. The last group is used to determinate concepts that are possible to be used as left and right attributes of connection, which is defined as a class that create junctions between concepts.

Setting most of the attributes at the very first time the concept appear is not always possible, but it should be done eventually. Automatic knowledge acquisition should lead to the questions about main attributes, necessary to distinguish one concept from the other. List of possible connection between concepts are being populated automatically from information delivered to System, but may be changed by the user.

2.4 The Search of Similar Concepts

Comparison of two concepts on the base of attributes values and lists of possible connections may lead to the conclusion that, as the effect of communication, there were two concepts created, being in fact one same concept. Similarity or even equality in case of attributes is important enough to suspect those two concepts to

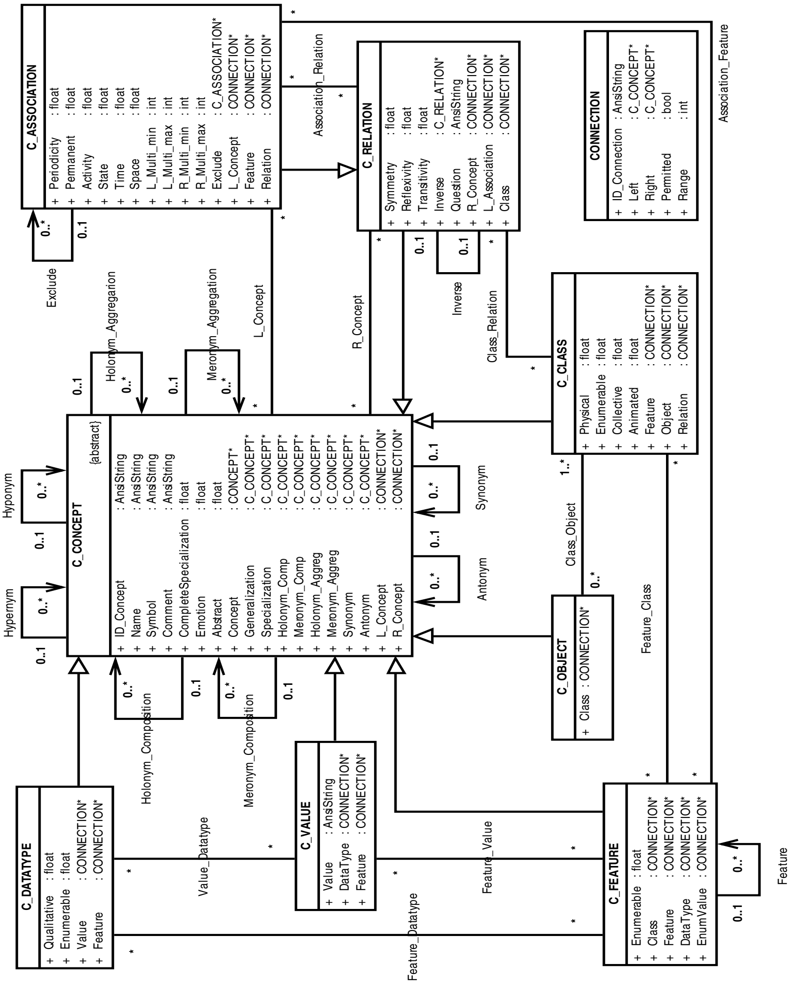


Fig. 1. Class diagram of COOS [4] showing how concepts are grouped in the lists

be synonyms, e.g. fig.2. List of possible connections, even thought it has different construction, may be the base for similar assumption. In case of such circumstance there is need for new knowledge from the human. It is implemented as question asking mechanism, that has completing information in simple attributes its priority. It would be great mistake to ask user whether those two concepts are equal. It might have been an easiest way, but there might be distortion coming from various interpretation of concepts and its symbols by different people. Due to this fact system needs to ask questions leading to completing information about specificity of the concept. After this stage renewed analysis of attributes values is needed, since examined concept might have been so greatly changed, that they no longer might be assumed synonyms. If this is not a case system needs to confront lists of possible connections. In most cases those list won't be similar, what is mainly caused by the way they are being created. Acquisition of facts that is the basis of system is directly connected with the mechanism that create connections between concepts. E.g. concept 'car' has been associated with concept 'drive'. In this case ontological core has a knowledge that it is permitted for the 'car' to be connected with association 'to drive' and that association 'to drive' may be connected with class 'car'.

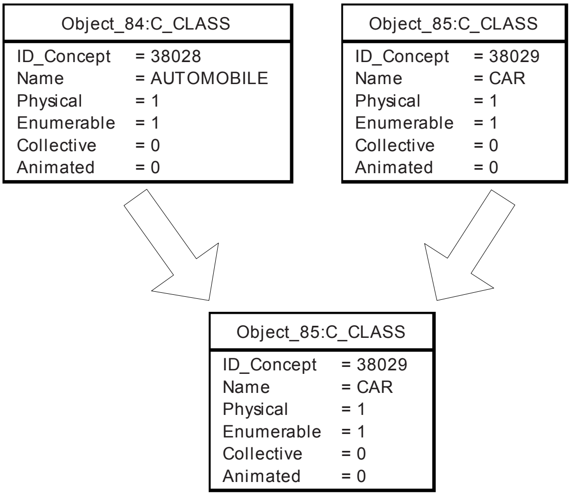


Fig. 2. Example of unification of two concepts

COOS is not using natural language inscriptions in the process of discrimination between concepts. All of the concepts are represented by an identifier, which is used to distinguish them even though they are identified by the same word in one of the natural languages. This fact is important in the process of identifying synonyms because in some cases connections might be connected to words similar in the way they are written or spoken, but they can be different concepts.

In the process of comparison of lists of possible connections for two concepts recognized to be similar, one have to take under consideration that even if the list is totally different, they still might be same concept. The process can produce one of following results:

1. lists are disjunctive,
2. lists are partly similar, but there are no contradictions, which are connections allowed on one list and prohibited on the other,
3. lists are partly similar, but there are some contradictions – at least one connection allowed on one list and prohibited on the other,
4. lists are fully similar, without any contradictions,
5. lists are fully similar, but there are contradictions.

Each of the situations require different solution. In the first one system has to determinate if it is possible to fully equalize differences of the lists by adding needed connections. If this is correct solution, system has to ask human whether two of concepts being examined are synonyms. Automatic assumption might be too risky. If equalization does not stand for correct solution, system will assume that concepts are not equal. In the second case, system has to determinate whether differences are possible to equalize. All other steps are same as in previous case.

In the third aspect, system has to find out whether contradictions shown on the lists are correct. Appropriate questions have to be asked to determinate correctness of ontological core content.

The next case can be obtained when lists are not defined fully. System need to ask questions that would find new connections for concepts that would differ them. If this process fail, system may assume after checking with user, that concepts are synonyms.

In the last possibility all of the connections are similar in the way of left and right argument, but they differ in allowance. If most of connections can be understood as contradictions, system may try to start process of determination whether two concepts are antonyms.

2.5 Separation of Concepts

Activity that is opposite to described in previous paragraph is separation of concepts. It is done when one concept has to be divided on two or more new concepts. The need for such action may arise when there are no concepts in ontological core, that would be suitable to use in specific situation, but there is another concept, which attributes and connection lists construction would be close to one needed.

The first stage of knowledge acquisition is characterized by huge lack of concepts, their attributes and possible connections between other concepts. During later stages, there are more concepts in the ontological core and more lacks are filled. Due to this facts early stages of base building will be rich in situation, when concepts will be similar in their representation in the base, but they won't be equal in ontological sense.

All actions set to distinguish synonymous concepts will be main target in early stages of knowledge building. In later stages main stress will be set to reduce number of concepts. This fact simply arise from the situation when concepts with very close or equal meaning will create separable instances in ontological core.

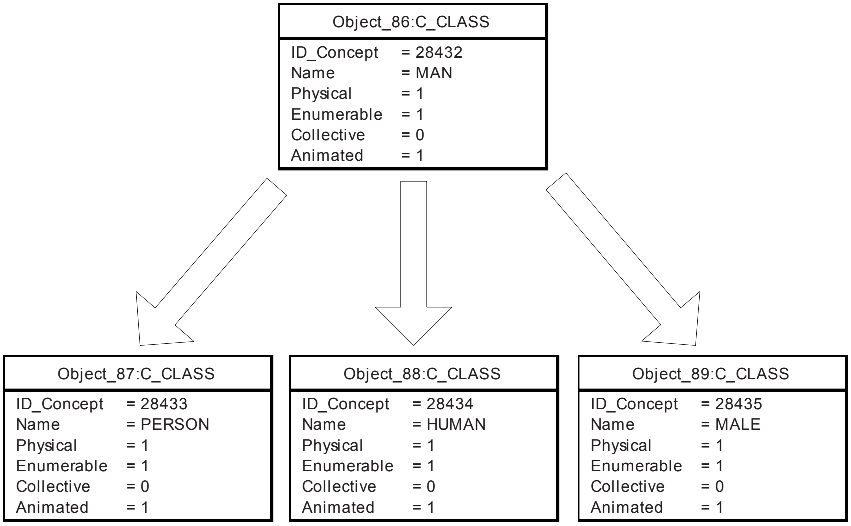


Fig. 3. Example of separation of concepts

Such reduction in early stages could lead to many mistakes and would restrain knowledge base growth.

It has to be point out that in most cases when one concept is represent as two different ones in the ontological core lead to less negative semantic consequences than when two different existences are treated as one concept. E.g. ‘a man’ and ‘human’ when treated as totally different concepts may produce great problems in advanced inference system based on ontological core processing. Figure 3 shows how concept “a man” can be separated into three other concepts: “a person”, “a human” and “a male”. Each of them has different meaning, but the difference can be found by comparing simple attributes. The main difference is stored in the lists of connections possible or prohibited to use. The possible negative outfit of treating those concept as one – “a man”, is especially visible in the aspect of generalization-specialization examining. It is destructive and should be avoided. On the other side, there are concepts like ‘to hold’ that correspond to many different meanings. Taking two of them under consideration one can hold – get a grip of something or one can hold – be able to be filled with something. If we imagine those two concepts being put into one instance in ontological core, the effect would be disastrous and much deeper. It is quite easy to imagine how those two meanings could blur any of inference processes. Those two examples shows that system has to be very careful in assumption of two concepts being synonyms. It is more easy to create relation of generalization and specification that may be checked on the later stage for correctness. This solution let system to connect two concepts and have them ready for further development.

2.6 Basic Concepts Search

Basic concepts are understood as concepts, that group common characteristics of derivative concepts. Finding such meanings is an alternative to actions set upon finding synonyms or joining concepts in one. This method is much safer, since those concepts does not have to fulfill any of the concepts we may use to describe in real life. In general sense concepts described in ontological core do not have to be in step with our method of reality description, that's why system can created new concepts that have no equivalent in natural language. Such action should be rare however, since they may lead to serious problem in communication between user and the system. On the other hand it is important to emphasize the fact that knowledge base both ontological and semantic is based on evolving concepts. Concepts are changing, they are base for creating new concepts, they are being changed and developed.

2.7 Information Scope Presented with Questions Generated by the System

In the process of decision solving in the aspect of creating, deleting and uniting of concepts, there are some information needed, mostly from attributes values. Second layer of information is stored in possible connections lists, that are grouped by classes, like shown on the fig.1. There might also be some information needed that can only be found in semantic network. It is possible since system is storing information about every and each fact that the concept is being used in. In this matter the analysis of context is possible. System store three layers of information that is needed during decision process:

- attributes values
- possible connections lists
- semantic network

The studies of context that can be understood as classical closer, and further context study, is mostly done on the level of semantic networks. Much deeper studies can be obtained through possible connections lists search done on the concept or one of its synonyms. It is also possible to search through generalized concepts to find some more general or common information, similarly one can try to find information by careful studies of specifications of the concepts, if such are present.

3 Conclusions

3.1 Active Knowledge Acquisition Role

Presented mechanism of active knowledge acquisition is supposed to acquire vital information, necessary in inference and decision making. Concepts (their attributes and connections) and structures of generalization/specialization need to be modified as soon as new information is available. Dynamics of knowledge and its structure is the most important part of learning [5], however without effective method of direct knowledge acquisition it is easy to commit mistakes and lead to ambiguity. In this case it would be degradation of knowledge already possessed. Presented method of knowledge acquisition is based on knowledge system that combines both ontological core and semantic network potential. Its main point is to be active, what means continuous work with human knowledge source leading to dynamic changes in ontological core on the level of concepts basic information, possible or prohibited connections and semantic network creation.

3.2 Further Development

Active knowledge acquisition module is part of Artificial Intelligence system. At this moment it is vital to take under consideration the best way of connecting each of separately built modules in one system. This is not trivial, since most of them have many layers of communication. Building appropriate communication system requires studies on different possible strategies and implementations. The main point is to specify internal protocol and language used by all of modules and by the system to communicate with the world. The language has to be easy to implement, capable of unambiguous information transfer, scalable and flexible. The information structure in the system requires also this language to be ready for recurrent sentence building.

References

1. Nirenburg, S., Raskin, V.: Ontological semantics. MIT Press, Cambridge (2004)
2. Dyachko, A.G.: Text processing and cognitive technologies. Moscow - Pushchino, ONTII PNTS (1997)
3. Krótkiewicz, M., Wojtkiewicz, K.: Conceptual ontological object knowledge base and language. In: Proc. 4th Conference on Computer Recognition Systems, Advances in Soft Computting, pp. 227–234. Springer, Heidelberg (2005)
4. Krótkiewicz, M., Wojtkiewicz, K.: Obiektowa baza wiedzy w ontologicznym systemie sztucznej inteligencji. In: Kozielski, S., Małysiak, B., Kasprowski, P., Mrozek, D. (eds.) Bazy danych. Modele, technologie, narzędzia. Analiza danych - wybrane zastosowania, Wydawnictwa Komunikacji i Łączności, Warszawa, pp. 221–228 (2005) (in Polish)
5. Włodarski, Z.: Psychology of learning, PWN, Warsaw (1996)

Problems of Knowledge Representation in Computer-Assisted Decision Making Systems

J.L. Kulikowski

Institute of Biocybernetics and Biomedical Engineering, Polish Academy of Sciences,
Warsaw, Poland
jkulikowski@ibib.waw.pl

Abstract. Formal knowledge representation methods in computer-assisted decision making systems are considered. A concept of infosphere and of its historical evolution is presented. Representation of knowledge in the form of factual, implicative and deontic statements as well as of evaluating meta-statements is described and their representation by formal structures is illustrated by examples. It is shown that ontological models based on formal structures can be used both to a reduction of semantic redundancy in knowledge bases and to semantically equivalent reformulation of queries ordered to the knowledge base. Application of ontological models as a source of knowledge used in decision making is illustrated by an example.

1 Introduction

Acquisition of information or more generally, its exchange with the environment, is one of the main attributes of life. During a long natural evolution process special information acquisition and processing organs in living organisms have been developed. Using its natural organs each living being can reach a type and spatially and temporally limited area of information, potentially available to it due to the contacts with the environment. This total type and area of information can be called a *natural infosphere* of the given biological (in particular – human) being. A sum of natural infospheres of a set of biological beings constituting a community, a biological species, etc. can be called its *biological infosphere*. Each individual having access to a biological infosphere plays, at the same time, the role of a source of information for other individuals. The time going, some biological communities develop some tools (systems of signs, languages) making them able to actively distribute or individually send some types of information to other members of the community. This leads to an extension of the individual biological infospheres within the community. Moreover, due to a collaboration of the community members additional sorts of information become available. As a consequence, the communities reach some total areas of available information which can be called their *social infospheres*. At the next step of development, human beings have invented technological tools and methods improving their abilities to acquire, store, transmit and/or disseminate information. As a consequence, a human being have got at his

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 39–54.
springerlink.com

disposal an extended area of available information which can be called his *individual infosphere*. The sum of *individual infospheres* of community members constitutes its *technological infosphere*. In the modern times, the last has been and is still subjected to dramatic development whose milestones were marked by the inventions of printing, optical instruments, photography, telephone, radio, electronic computers, electron microscopy, artificial satellites, nuclear magnetic resonance (NMR) device, computer networks etc., up to recently constructed Large Hadron Collider (LHD) or Gamma-ray Large Area Space Telescope (GLAST). The relationships between the biological, social and technological infospheres are illustrated in Fig. 1. They are not related by a sequence of inclusions: there is a common part of the infospheres that can be called a *bio-socio-technological infosphere* (*BST-infosphere*) and there remain some parts of bio-, socio- or technological infosphere which cannot be included into the other ones. The BST-infosphere is of particular interest: it not only stimulates human progress but also is of this progress the best visible exponent. It contains, in particular, all types of human experience and knowledge gathered for the centuries due to human co-operation and information exchange aided by any types of technological tools.

The mentioned above notion of knowledge can be in several ways defined as, e.g.:

- "A whole of stored in human mind contents being a result of accumulation of experiences and of learning processes" [1],
- "It is the application of a combination of instincts, ideas, rules, procedures, and information to guide the actions and decisions of a problem solver within a particular context" [2],
- "A capability of classification of being of interest phenomena, processes, thoughts, etc. (objects of an universe)" [3].

Some authors include into their definitions of knowledge a requirement of knowledge components credibility verification. On the other hand, knowledge becoming a part of technological or BST infosphere cannot be a simple collection of vague opinions or statements concerning a given subject; it should satisfy some requirements of clarity and it should be fitted with a structure making the knowledge resources available for retrieval, exploration and enriching. Clarity needs a language making possible precise description of facts or expression of thoughts. For this purpose three categories of languages can be used:

- natural ethnic languages,
- naturalized specialist sub-languages,
- artificial symbolic languages.

Ethnic language are the more flexible and universal ones. Due to richness and loose interpretation of their basic terms they suit well to a rough new objects or phenomena description in preliminary states of their investigation. On the other hand, they admit vagueness and various interpretations of statements and as such they do not satisfy well the requirements of exactness in knowledge presentation. Hence, natural languages are used mostly as meta-languages for creation of the

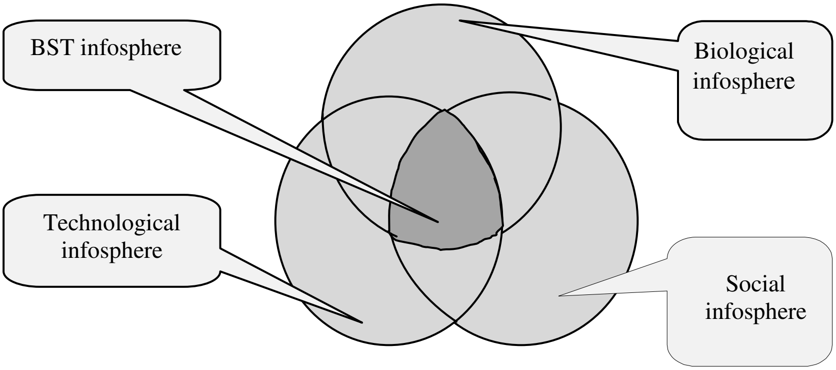


Fig. 1. Relations between Bio-Socio-Technological infosphere and its components

above mentioned two other categories of languages. Naturalized specialist sub-languages preserve grammatical rules and limited vocabularies of auxiliary words from natural languages. However, they are based on strongly defined terms closely connected with selected areas of interest. Therefore, such languages suit well to inter-communication and knowledge distribution within groups of specialists in the given areas. In principle, they also can be used to human - computer system knowledge exchange. However, difficulties connected with naturalized languages parsing make this category of languages inconvenient to computer-aided knowledge processing.

Artificial symbolic languages can be, in particular, dedicated to computer-aided knowledge resources storage and exploration purposes. A typical knowledge representing artificial symbolic language should consist of a *knowledge units description core* and a *basic terms vocabulary*. The last may consist of symbolic equivalents of a naturalized specialist sub-language notions corresponding to a described universe, while the knowledge units description core should be adjusted to the expected types of tasks or queries addressed to the computer knowledge processing system.

The aim of this paper is to show the way of knowledge expressed in natural language through adequately chosen logical models transforming into formal data structures convenient to computer-aided decision making. The following problems in the next sections of the paper are presented: basic types of knowledge representing statements and their relational models have been described in Sec. 2, some problems of redundancy reduction in knowledge bases are analyzed in Sec. 3, the role of formal ontological models in knowledge bases exploration have been described in Sec. 4. Conclusions have been collected in Sec. 5.

2 Knowledge Representing Statements

From a formal point of view knowledge resources stored in a computer systems consist of collections of knowledge units which from a logical and semantic point of view can be specified as assertive statements of the following types:

- factual statements,
- implicative statements,
- deontic statements,

as well as:

- evaluating meta-statements.

Basic characteristics of the above-mentioned notions are presented below. The statements are expressed in the terms of a basic terms vocabulary containing, in general, symbolic expressions of nouns, adjectives, numerals, verbs, adverbials etc., as well as logical operators (in particular: *negation* (¬)**,** *conjunction* (∧), *disjunction* ∨, etc.), linguistic operators (e.g.: *separators* |, *quotation marks* " " *brackets* ( ), etc.), mathematical symbols, etc. chosen according to the expected application needs. In particular, it will be shown below that phrases or statements taken into quotation marks in certain cases can be considered as nouns.

2.1 Factual Statements

Assertive statements of factual type may concern real or abstract objects, persons, phenomena and/or processes about which lot of questions can be posed, like: "What is this?", "Who is he?", "What are the properties of this?", "What is it caused by?", "What is it similar to?", "What will be caused by it?", "Where is it described?", etc. From a linguistic point of view simple factual statement answering one of the above-mentioned questions can be presented in the form of a concatenation (co-occurrence, ∗) of triplets of phrases:

$$f_f = S * P * Q \qquad (1)$$

where:
S – denotes subjective phrase consisting of a *subject,* i.e. a *noun* pointing out an (abstract or real) object, a person, an event, a process, etc., and, possibly, some *attributes* describing quantitative or qualitative *features* of the subject expression by *adjectives* or *numerals*,
P – predicative phrase consisting of a *predicate* in the form of a *verb* and (eventually) some *adverbials* of time, place, manner, concession, number, degree, etc. or in the form of a mathematical or logical operator, whereas
Q – is objective phrase whose construction is similar to this of a subjective phrase (however, it may occur optionally).

Examples. The following examples illustrate simple factual statements:

[helium] ∗ [is] ∗ [noble gas] ≡ "Helium is a noble gas";
[smoking] ∗ [increases] ∗ [risk of lung cancer] ≡ "Smoking increases the risk of lung cancer" ●

Using logical and/or linguistic operators the form of subjective, predicative and objective phrases can be extended on composite phrases. Formula (1) holds also if *S, P, Q* are respectively replaced by *S', P', Q'* denoting composite phrases. Let us denote, correspondingly, by $F_{S'}$, $F_{P'}$ and $F_{Q'}$ the sets of all admissible simple or composite subjective, predicative and objective phrases constructed on the basis of terms of a given *basic terms vocabulary*; taking into account that such terms are usually chosen for a description of an assumed universe U it can also be told that the phrases have been described over the universe U under consideration. Therefore, a Cartesian product $F_{S'} \times F_{P'} \times F_{Q'}$ represents all possible triples of simple and composite phrases whose concatenations according to formula (1) generate factual statements. However, only some of them satisfy also a *semantic meaningfulness* constraint.

Examples. Among the concatenations of phrases:

[high caloric diet] * [causes] * [(diabetes)∨(circulatory diseases)];
[sinus] * [is] * [(periodic function)∧(everywhere differentiable function)];
[high caloric diet] * [causes] * [everywhere differentiable function]

The first two are *semantically meaningful* while the third one is *semantically meaningless* in the sense that it cannot be related to any situation arising in the universe U; however, some semantically meaningful sentences may be acceptable in a metaphoric context. ●

Let us take into consideration a subset:

$$\Phi_f \subseteq F_{S'} \times F_{P'} \times F_{Q'} \tag{2}$$

of all triples of phrases semantically meaningful in the universe U under consideration. It should be noticed that, in a formal sense, Φ_f constitutes a *relation* described on the given Cartesian product. Resources of factual statements in a knowledge base can be thus considered as families Σ_f of relations described by formula (2). It will be shown below that other types of assertive statements in similar form can be presented. This fact plays an important role in the knowledge representation methods.

2.2 Implicative Statements

Implicative statements are widely used in decision-aiding expert systems. They have the following general form:

$$f_i = \textbf{\textit{If}}\ p\ \textbf{\textit{then}}\ q \tag{3}$$

where p and q are *semantically correct factual statements*, p being called a *premise* and q – a *conclusion*. Like before, the formal structure of p and q is described by formula (1).

Examples. The following statements illustrate the form described by formula (3):

If (the number of measurements is too small) ***then*** (the measurement error is high);
If (the glucose level in patients' blood is high) ***then*** (the patient is endangered by diabetes);
If (sinus is a continuous function) ***then*** (Rome is a capital city of Italy) ●

Like factual statements, some formally and logically correct implicative statements, despite the fact that their premise and conclusion are meaningful, in certain context, may be semantically meaningless.

If Φ_p and Φ_q denote the sets of all semantically meaningful factual statements used, respectively, as premises and conclusions then a subset

$$\Phi_i \subseteq \Phi_p \times \Phi_q \qquad (4)$$

of all pairs of premises and conclusions that may form semantically meaningful implicative statements takes also the form of a formal relation. Resources of implicative statements in a knowledge base can be thus considered as families Σ_i of relations described by formula (4).

2.3 Deontic Statements

Deontic statements express recommendations, instructions of actions, rules of behavior, etc. (gr. *deon = a duty, what should be*). In general, they have the following form:

If *r* ***and it is desired*** *s* ***then do*** *t*

where:
r – is an assertive factual statement describing an **initial state** of an universe,
s – a linguistic phrase semantically equivalent to the description of a desired **final state** of the universe,
t – an imperative mood statement or a phrase semantically equivalent to the description of an **action** or their ordered sequence (**algorithm**, **program,** etc.).

Examples. The above-given definition can be illustrated by the following statements:

If [red light goes on] ***and it is desired*** [protection of the device from destroying] ***then do*** [switch off the device];
If [you want to login] ***and it is desired*** [acceptance] ***then do*** [enter your password] ●

Let us denote by: Φ_{is} – a set of all assertive factual statements describing initial states of a universe U, Φ_{fs} – a set of all assertive factual statements describing final states of U, Ψ – a set of all imperative mood statements or phrases semantically equivalent to the description of action(s). Then the subset:

$$\Phi_d \subseteq \Phi_{is} \times \Phi_{fs} \times \Psi \qquad (5)$$

is a formal relation describing the triples semantically equivalent to all possible deontic statements over U. Resources of deontic statements in a knowledge base can be considered as a family Σ_d of relations of the type described by (5).

In order to assign logical values to the deontic statements let us remark that they are logically equivalent to the following implicative statements:

If [*r*] ***and*** [*t* has been done] ***then*** [*s* will be reached].

2.4 Evaluating Meta-statements

Evaluating meta-statements can take one of the following forms:

$$h_v = f * L \tag{6a}$$

or

$$h_r = f * R * g \tag{6b}$$

where:
f, g – are any factual, implicative or deontic statements,
$L = [V, v]$ – is a pair of elements characterizing logical value assigned to the statement; *V* being a linearly ordered set of logical (probabilistic, etc.) values, $v \in V$ being its element assigned to the given statement,
R – relation of logical semi-ordering of the given pair of statements.

Examples. The evaluating meta-statements can be illustrated as follows:

[sinξ and cosξ are periodic functions] [*V, true*] ≡
≡ "***The statement*** [sinξ and cosξ are periodic functions] **is *true***";
[flight No 16 today will be delayed more than 30 min] [*V*, 0.6] ≡
≡ "***The probability of the event*** [flight No 16 today will be delayed more than 30 min] ***is 0.6***";
[David is a good student] $\preceq$ [Joan is a good student] ≡
≡ "***The statement*** [David is a good student] ***is less justified than*** [Joan is a good student]" ●

Let us remark that in a knowledge base several different logical evaluation systems *L* can be used; a family of such systems will be denoted by *Λ*.

If ***Φ*** denotes a set of all factual, implicative or deontic statements over an universe *U* then the set of all admissible evaluating statements of the (6a) type takes the form of a super-relation:

$$\Phi'_e \subseteq \Phi \times \Lambda \tag{7a}$$

In similar way, the set of all admissible evaluating statements of the (6b) type can be presented by a super-relation:

$$\Phi''_e \subseteq \Phi \times W \times \Phi \tag{7b}$$

In particular, in a version of relative logic the following logical relationships between the statements are admitted: $W = \{\preceq, \succeq, \approx, ?\}$, where $\preceq$ – *is not more logically justified than,* $\succeq$ – *is not less logically justified than,* ≈ – *are equally logically justified, ? – are logically incomparable* [4].

Taking into account that ***Φ*** can be defined as a relation, Φ'_e and Φ''_e represent a sort of *super-relations* (relations between relations). A family Σ_e of super-relations Φ'_e and Φ''_e described on the relations belonging to Σ_f, Σ_i and Σ_d is the fourth component of knowledge-base resources. The last can be thus formally defined as a sum:

$$\Sigma = \Sigma_f \cup \Sigma_i \cup \Sigma_d \cup \Sigma_e \tag{8}$$

The elements of Σ will be considered as *structured knowledge entities* (*SKE)* stored in the knowledge bases.

3 Structural and Semantic Redundancy in Knowledgebases

Knowledge presentation in the form of the families of *SKE*s is a step toward making it more convenient for computer processing than using a natural language for this purpose. Its superiority is connected both with standardization of form and conciseness of the statements. On different knowledge presentation levels redundancy is caused by different factors, as it is shown in Fig. 2.

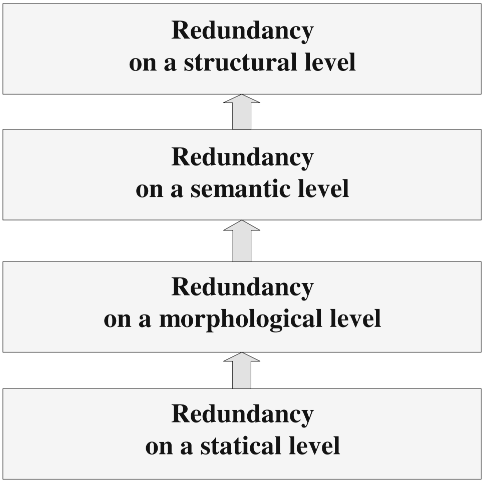

Redundancy on a structural level
Redundancy on a semantic level
Redundancy on a morphological level
Redundancy on a statical level

Fig. 2. A hierarchy of redundancy levels in knowledgebases

The type of redundancy on the basic and morphological levels is well known due to the basic C. Shannon works [5]; effective methods of its reduction in the form of data compression standards for communication systems and multi-media databases have been elaborated [6, 7]. Our attention below will be focused on the redundancy on semantic and structural knowledge presentation levels.

The following factors may cause semantic redundancy in knowledge bases:

a) using extended terms instead of their abbreviations or more concise synonyms (e.g. [Anno Domini nineteen hundred forty five] instead of [AD 1945];
b) expression of thoughts or description of facts by several semantically equivalent statements (e.g. [X is a normally distributed random value; its probability

density function is described by a Gauss function], [The measurement error should be small, less than 0.05%, not exceeding 0.03%"];

c) detailing of objects, actions etc. instead of using their collective names (e.g. [Integer, rational, algebraic and transcendental real numbers] instead of [All real numbers].

Structural redundancy occurs mainly in large (in particular, distributed) knowledgebase. It consists in repeating the same or semantically equivalent knowledge entities in various data structures distributed and stored in the system. Structural redundancy may be caused by imperfection of the knowledge base managing system. In particular, supplying the system from different sources without semantic data verification may lead not only to semantic and/or structural knowledge redundancy but also to a knowledge-base logical inconsistency. On the other hand, a strongly controlled structural and semantic redundancy level may help in achieving higher knowledge retrieval effectiveness. This will be illustrated below.

Example. Let us assume that a query:

"What are the chemical compounds of argon and oxygen?

is ordered to a knowledge base. However, no statements like:

[X][is][(chemical compound) of][argon∧oxygen],

in the knowledge base do exist. Instead, there can be found the following statements:

[Ar][is](chemical element)],
[O][is][(chemical element)],
[(Ar∨He∨Kr∨Ne∨Xe)][is][(noble gas)],
[(noble chemical element)][¬(forms)][(chemical compound) of]∗[(noble chemical element)∧(chemical element)],
[(noble chemical element)][is][(noble gas)∨(noble metal)],

etc. They are not sufficient to a direct deduction of a logical answer to the above formulated query, however, in two cases it could be possible:

- if the query is reformulated as:

  "What are the chemical compounds of Ar and O?

- if additional, semantically redundant statements like:

  [argon][is][(chemical element)],
  [oxygen][is][(chemical element)],
  [(argon∨helium∨krypton∨neon∨xenon)][is][(noble gas)],

have been included into the knowledge base. In both cases a construction of a logical path leading from the initial query to a logical answer:

"Argon forms no chemical compounds with other chemical elements"

is possible. However, an automatic reformulation of the query also needs additional statements:

[Ar][means][argon],
[O][means][oxygen]

establishing synonymic (i.e. semantically redundant) relationships between the terms to be included into the knowledge base. ●

The users of a queries answering system expect that their queries will be understood and correctly replied despite a large variety of semantically equivalent forms of their formulation. The computer system should thus be able to **automatically** reformulate the queries in order to find in the knowledge base the adequate *SKE*s and to use them to a logical inference of correct replies. This leads thus to the question, which of the above-mentioned two ways of problem solution is better? It seems that using standard terms in connection with the statements establishing synonymic relationships is, in general, a more effective way than multiplication and storage of semantically equivalent statements.

4 Knowledge Representation by Ontological Models

The problems of the nature and general theory of existence drew attention of the philosophers since the ancient times. In XVII century a term *ontology* considered as a part of metaphysics has been assigned to them. Some basic ontological problems, like those of a priority of spiritual versus material principle of a reality are still kept alive. However, ontology has also narrower aspects. If considered as "a common understanding of some domain" [8], "explicit specification of a conceptualization" [9], or "an abstract view of the world we are modeling, describing the concepts and their relationships" [10] ontology can be involved into advanced computer-aided decision making systems. A general concept of universal ontology as a tool for reality description is proposed in [11].

For practical applications, an ontology $\boldsymbol{O}$ can be represented by a logically consistent set of *ontological models* $\boldsymbol{OM}_j$ defined as ordered quadruples [12]:

$$OM_j = [C_j, R_j, A_j, Top_j], j = 1,2,...,J, \tag{9}$$

where:
$\boldsymbol{C}_j$ – is a non-empty set of **concepts;**
$\boldsymbol{R}_j$ – a family of **relations** between the elements of C_j containing, in particular, a **taxonomy** Ξ_j establishing hierarchical relationships between the concepts and their instances;
$\boldsymbol{A}_j$ – is a subset of **axioms** concerning the relations in R_j;
$\boldsymbol{Top}_j \in \boldsymbol{C}_j$ is the **highest element** in the taxonomy Ξ_j.

Logical consistency of ontological models within a given ontology is reached due to a family $\boldsymbol{F}$ of cross-relations $\boldsymbol{R}_{p,q,...,r}$ defined on the sets of concepts belonging to different ontological models. Therefore, an ontology can be presented in the form:

$$\boldsymbol{O} = [\boldsymbol{OM}_1, \boldsymbol{OM}_2, \ldots, \boldsymbol{OM}_k, \boldsymbol{F}] \tag{10}$$

Besides the taxonomies some other partial ordering (reciprocal, symmetrical and transitive) relations in ontological models are used. Among them *mereological relations,* establishing hierarchies between some objects and their components should be mentioned. The following examples in graphical form illustrate some relations used in ontological models. In Fig. 3 a typical taxonomic relation between the *Organic pressure* and subordinated to it concepts is shown.

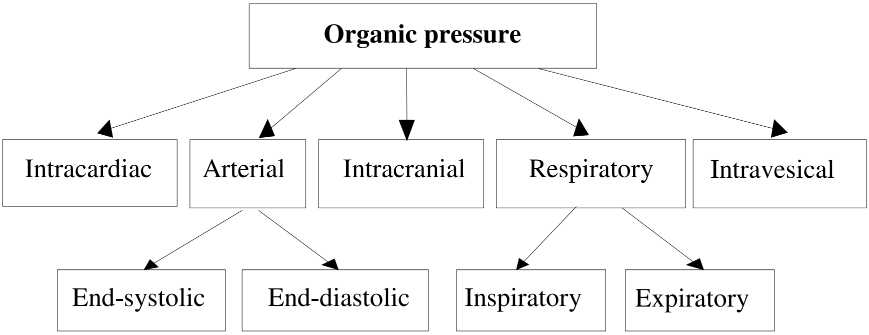


Fig. 3. A taxonomic tree describing an *Organic pressure* concept

A mereological relation between cardiac components is illustrated in Fig. 4.

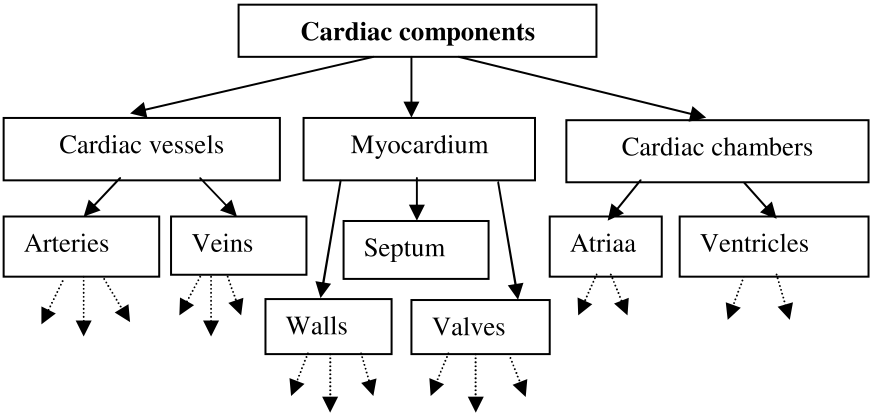


Fig. 4. A mereological tree of cardiac components

Ontological models provide large possibilities of reinterpretation of queries ordered to a knowledge base and formulation of extended replies based not only on the statements directly fitting to the queries but also on larger knowledge resources stored in the form of *SKE*s. A simple example of using a taxonomic tree to formulate an extended reply to a query is given below.

Example. Let us assume that a query:

"What are the arterial pressure values in norm?"

has been ordered to a knowledge base. An answer should be found in a factual database. However, no *SKE* of the form:

"[Arterial pressure value in norm] [is] [120-140 hPa]"

in the factual knowledge base exists. Instead, there can be found there:

"[Arterial end-diastolic pressure value in norm] [is] [60-90 hPa]",
"[Arterial end-systolic pressure value in norm] [is] [120-140 hPa]".

A correct answer to the query can be found by taking into account the part of taxonomic tree shown in Fig. 3 establishing relationships between the concept *Arterial pressure* and its instances *Arterial end-systolic pressure* and *Arterial end-diastolic pressure*. The right answer thus will be:

[The arterial pressure values in norm][are]:
[end-diastolic pressure value in norm] [is] [60-90 hPa]",
[end-systolic pressure value in norm][is] [120-140 hPa]" ●

In the above-presented case some (in general, different) properties have been assigned to the objects represented by the instances of a given higher-level concept. The result thus has been obtained by application of a rule:

> ***A query concerning attributes of an object represented by a concept in a taxonomic tree is equivalent to a set of similar queries concerning the objects represented by the subordinated concepts.***

Another situation arises if a property is assigned to an object represented by a superior concept and a query concerns one of its instances.

Example. There are given in a knowledge base:

- a taxonomic tree presented in Fig. 3,
- an implicative statement:
  If (the respiratory pressure for a long time is abnormally low)
  then (it may affect arising anoxaemia).

Let the following query be ordered to the system:

"What can be affected by an abnormally low expiratory pressure?

Using the *modus ponens* syllogism:

$$\frac{\begin{array}{c}\textit{The respiratory pressure is abnormally low;}\\ \textit{Expiratory pressure is a respiratory pressure;}\end{array}}{\textit{Expiratory pressure is abnormally low}}$$

the implicative statement can be reformulated as follows:

***If** (the expiratory pressure for a long time is abnormally low)*
***then** (it may affect arising anoxaemia)*

and finally, the following reply to the query will be given:

"Anoxaemia can be affected by an abnormally low expiratory pressure" ●

Similar possibility of reformulation of queries is connected with mereological relations if a whole object's state or property can be extended on all its components.

Example. Let us take into consideration a deontic statement:

***If** [cardiac chambers are to be diagnosed] **and it is desired** [real-time examination of action] **then do** [use RTG, USG or NMR visualization modality].*

Assume that there is given a query:

"What are the methods of real-time examination of cardiac ventricles' action?"

On the basis of mereological relation presented in Fig. 4 the above-given deontic statement can be reformulated as follows:

***If** [cardiac atrii or cardiac ventricles are to be diagnosed] **and it is desired** [real-time examination of action] **then do** [use RTG, USG or NMR visualization modality].*

This leads to the answer:

"If cardiac ventricles are to be diagnosed and a real-time examination of action is desired then use a RTG, USG or NMR visualization modality" ●

Among other types of ontological models there can be mentioned: head – subordinate, ancestor – descendant, producer – product, cause – result,etc.

Unlike the taxonomic or mereological relations represented by trees, in the first three of the above-given cases the corresponding models have the form of contour-less directed graphs. E.g., each human descendant has two ancestors – a situation that cannot be represented by a tree. In the last case, a formal model of the *cause-result* relation can be constructed on the basis of *Petri nets* (a specific form of labelled directed graphs [13]).

The idea of using cross-relations to guarantee logical consistency of a set of ontological models is illustrated below.

Example. Let us assume that in industrial enterprise licences of operators working on some workstations should be regularly updated. The types and time-periods of validity of licences are for different workstations different and a computer system for their verification is to be designed.

The verifying procedures are based on a knowledge concerning the workstations, the state of health of the workers and the rules of licensing to work on the workstations. The necessary knowledge can be represented by formal models constituting a part of the domain ontology of the enterprise. The structure of this part of ontology should be designed.

Basic concepts concerning the enterprise are collected in a *Taxonomic tree* whose selected part is shown in Fig. 5; it plays the role of a source of keywords used to create the reasoning processes. For licenses updating the path *Enterprise – Functions - Security protection - Work/health protection - Protection procedures – Checking working licence validity* in the taxonomic tree is of particular interest.

The node *Checking working licence validity* should contain a link to a deontic statement which in natural language has the form:

“If admittance of *n* to work as *p* on *x* is asked then check the validity of all his necessary health certificates in order to prove if they are presently valid”.

Here n, $n \in N$, denotes a personal identifier (e.g. name of a worker); p, $p \in P$, – a post of a worker; x, $x \in X$, – a workstation. Some branches of the taxonomic tree are terminated by symbols of relations $\boldsymbol{R_1}$, $\boldsymbol{R_2}$, $\boldsymbol{R_3}$ which will be explained below. For computer processing purposes the above-given deontic statement should be presented in a more concise form:

If *[n is to be employed as p on x]* ***then check*** [$\boldsymbol{R_1}(n,p,x)\cap\boldsymbol{R_2}(p,x,T)\cap\boldsymbol{R_3}(d,T)$] ***in order to confirm that*** *[licence of n to work on x as p is valid].*

Here d, $d \in D$, denotes a current date; T is standing for a set of all possible medical certificates that can be required from the workers employed in the given enterprise.

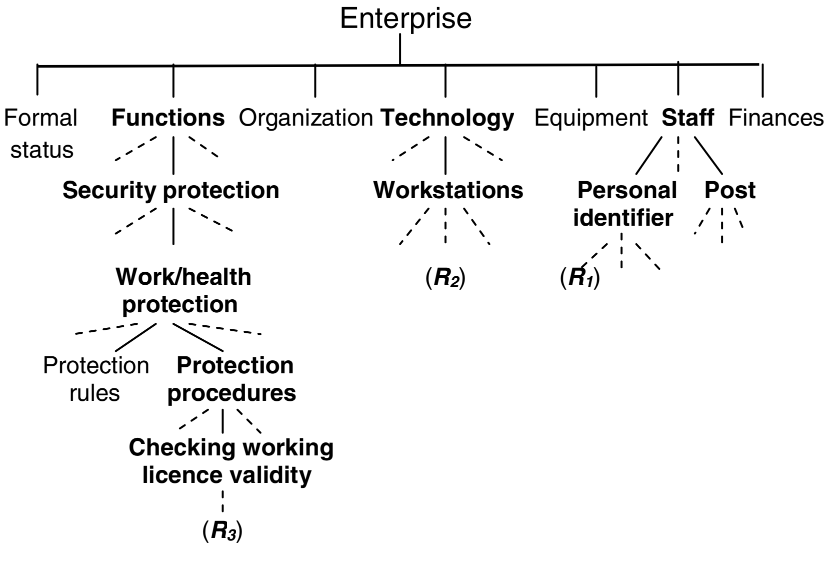


Fig. 5. A part of a taxonomic tree of the ontology of *Enterprise*

The second term of the statement has the form of a conjunction of hyper-relations [14]. $\boldsymbol{R_1(n,p,x)}$ is a formal representation of a factual statement saying that "*n is employed as p on x*". $\boldsymbol{R_2(p,x,T)}$ is a hyper-relation representing a factual statement saying that "*anybody working as p on x should submit medical certificates* $t_1, t_2,\ldots,t_k$", where $t_1, t_2,\ldots,t_k \in T$. $\boldsymbol{R_3(d,T)}$ is a hyper-relation corresponding to a factual statement that "*medical certificates* $t_1, t_2,\ldots,t_k$ *are valid on d*". Using hyper-relations instead of relations is here justified by the fact that the number k of the arguments $t_1, t_2,\ldots,t_k$ depends on the values of the remaining arguments (while the number of arguments in relations is strongly fixed). $\boldsymbol{R_1}$, $\boldsymbol{R_2}$ and $\boldsymbol{R_3}$ are thus formal models included into the ontology of the enterprise.

The conjunction of hyper-relations forces the instances of $\boldsymbol{R_1(n,p,x)}$ and $\boldsymbol{R_2(p,x,t_1, t_2,\ldots,t_k)}$ to keep common values p, x and those of $\boldsymbol{R_2(p,x,t_1, t_2,\ldots,t_k)}$ and $\boldsymbol{R_3(d, t_1, t_2,\ldots,t_k)}$ to keep common values $t_1, t_2,\ldots,t_k$. Therefore, it is a formal tool guaranteeing a compactness of the sequence of ontological models used to construct a decision rule based on the given knowledge resource. ●

5 Conclusions

It follows from the above-given considerations that:

- Knowledge bases consist of factual, implicative, deontic and evaluating statements playing the role of knowledge entities;
- The above-mentioned knowledge entities can be formally presented as instances of relational formal structures;
- Formal representation of knowledge entities makes possible a reduction of semantic redundancy in knowledge bases;
- Formal structures describing real objects, processes etc. can be considered as their ontological models used to support decision making concerning the given domain of reality;
- Ontological models make possible semantically equivalent reformulation of queries ordered to the knowledge bases;
- In construction of ontological models the concepts of extended algebra of relations and of hyper-relations can be effectively used.

It also should be remarked that formal ontology combined with novel concepts in the theory of relations and with fuzzy reasoning methods seems to be an important step to fill the gap between natural human thinking and computer-aided reasoning methods.

Acknowledgement

This work has been partially supported from the project N N518 4211 33 financed by the Ministry of Science and Higher Education of Poland.

References

1. Great Universal Encyclopedia, PWN, Warsaw, vol 12, p 251 (1969) (in Polish)
2. Xodo, D.: Multi-participant decision making and balanced scorecard collaborative. In: Rivero, L.C., et al. (eds.) Encyclopedia of database technologies and applications. Idea Group Reference, Hershey (2006)
3. Pawlak, Z.: Knowledge and rough sets. In: Traczyk, W. (ed.) Artificial intelligence problems. Wiedza i Życie, Warsaw (1995) (in Polish)
4. Kulikowski, J.L.: Decision making in a modified version of topological logic. In: Proc. Seminar on Nonconventional Problems of Optimization. Part 1, Prace IBS PAN, 134, Warsaw (1986)
5. Goldman, S.: Information theory. Constable and Co., London (1953)
6. Skarbek, W. (ed.): Multimedia. Algorithms and compression standards. AOW LPJ, Warsaw (1998) (in Polish)
7. Kunt, M., Ikonomopoulos, A., Kocher, M.: Second generation image coding techniques. In: Proc. IEEE Intern. Conference, Yilan, Taiwan, vol. 73(4), pp. 549–575 (1985)
8. Fernandez-Lopez, M., Gomez-Perez, A.: Overview and analysis of methodologies for building ontologies. The Knowledge Eng. Rev. 17(2), 129–156 (2002)
9. Gruber, T.R.: A translation approach to portable ontologies. Knowledge Acquisition 5(2), 199–220 (1993)
10. Zilli, A., Damiani, E., et al. (eds.): Semantic knowledge management.: an ontology-based framework. Information Science Reference, Hershey (2009)
11. Abdoullaev, A.: Reality, universal ontology, and knowledge systems: toward the intelligent world. IGI Publishing, Hershey (2008)
12. Kulikowski, J.L.: Structural image analysis based on ontological models. In: Kurzyński, M., et al. (eds.) Computer recognition systems 2, pp. 68–75. Springer, Heidelberg (2007)
13. Winkowski, J.: Towards an algebraic description of discrete processes and systems. Institute of Computer Science, Polish Academy of Sciences, Warsaw (1980)
14. Kulikowski, J.L.: Description of irregular composite objects by hyper-relations. In: Wojciechowski, K., et al. (eds.) Computer vision and graphics, pp. 141–146. Springer, Heidelberg (2004)

A Dialogue-Based Interaction System for Human-Computer Interfaces

G. Bel-Enguix[1], A.H. Dediu[1,2], and M.D. Jiménez-López[1]

[1] Rovira i Virgili University, Tarragona, Spain
{gemma.bel,adrian.dediu,mariadolores.jimenez}@urv.cat
[2] University of Technology, Bucharest, Romania

Abstract. Modeling man-machine interaction based on human conversation can provide flexible and effective user interfaces. Conversational interfaces would provide the opportunity for the user to interact with the computer just as they would do to a real person. In this paper, we introduce a formal model of dialogue that may contribute to the building of better man-machine interfaces.

1 Introduction

Many researchers believe that natural language interfaces can provide the most useful and efficient way for people to interact with computers. According to (Zue 1997), "for information to be truly accessible to all anytime, anywhere, one must seriously address the problem of user interfaces. A promising solution to this problem is to impart human-like capabilities onto machines, so that they can speak and hear, just like the users with whom they need to interact." Human-computer interfaces require models of dialogue structure that capture the variability and unpredictability within dialogue. The study of human-human conversation and the application of its features to man-machine interaction can provide valuable insights, as has been recognized by many authors [2, 11, 14, 15, 21]. To be truly conversation-like, a human-computer dialogue has to have much of the freedom and flexibility of human-human conversations.

Different degrees of conversational-like interfaces can be distinguished according to [21]:

1. in one extreme, the computer can take complete control of the interaction by requiring that the user answer a set of prescribed questions;
2. at the other extreme, the user can take total control of the interaction while the system remains passive;
3. in the middle, we find the mixed-initiative goal-oriented dialogue, in which both the user and the computer participate actively to solve a problem interactively using a conversational paradigm.

The dialogue system we introduce in this paper can be considered a mixed-initiative model. In order to build this mixed-initiative model we have examined

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 55–65.
springerlink.com

human-human interactions. Even though many researchers agree that a complete simulation of human conversation is very difficult (maybe impossible) to be reached, it seems clear that knowledge of human language use can help in the design of efficient human-computer dialogues. It can be argued that users could feel more comfortable with an interface that has some of the features of a human agent. Therefore, our model is based on human-human interactions. The result is a highly formalized dialogue-based framework that may be useful for human-machine interfaces.

Section 2 introduces the formal definition of a dialogue-based interaction system. Section 3 offers a simple example with its implementation for illustrating how the model works. Section 4 presents some final remarks.

Throughout the paper, we assume that the reader is familiar with the basics of formal language theory, for more information see [16].

2 A Dialogue-Based Interaction System

In this section, we introduce a formal model of dialogue that may contribute to the building of better human-computer dialogues through a simulation of human language use. Note that we do not intend to fully simulate human language use, but only to take advantage from research on human language in order to improve conversational interfaces. Our formal model of dialogue tries to capture human-conversation main features and is based on Eco-Grammar Systems (EGS). Eco-grammar systems theory is a subfield of grammar systems theory, a consolidated and active branch in the field of formal languages [7]. Eco-grammar systems have been introduced in [8] and provide a syntactical framework for eco-systems; this is, for communities of evolving agents and their interrelated environment.

Taking as a starting point the notion of an eco-grammar system, we introduce the notion of Conversational Grammar Systems (CGS). CGS present several advantages useful for dialogue systems, we emphasize here only several aspects like the generation process is:

1. highly *modularized* by a distributed system of contributing agents;
2. *contextualized* permitting to linguistic agents to re-define their capabilities according to context conditions given by mappings;
3. *emergent*, from current competence of the collection of active agents emerges a more complex behaviour.

Definition 1. A Conversational Grammar System (CGS) of degree n, $n \geq 2$, is an (n+1)-tuple $\Sigma = (E, A_1,...,A_n)$, where:

- $E = (V_E, P_E)$,
  – V_E is an alphabet;
  – P_E is a finite set of rewriting rules over V_E.

- $A_i = (V_i, P_i, R_i, \varphi_i, \psi_i, \pi_i, \rho_i)$, $1 \leq i \leq n$,
  – V_i is an alphabet;
  – P_i is a finite set of rewriting rules over V_i;

- R_i is a finite set of rewriting rules over V_E;
- φ_i: $V^*_E \to 2^{P_i}$;
- ψ_i: V^*_E x $V^+_i \to 2^{R_i}$;
- π_i is the start condition;
- ρ_i is the stop condition;
- π_i and ρ_i are predicates on V^*_E.

The items of the above definition have been interpreted as follows:

1. E represents the environment described at any moment of time by a string w_E, over alphabet V_E, called the *state of the environment*. The state of the environment is changed both by its own evolution rules P_E and by the actions of the agents of the system, A_i, $1 \leq i \leq n$.
2. A_i, $1 \leq i \leq n$, represents an agent. It is identified at any moment by a string of symbols w_i, over alphabet V_i, which represents its current state. Such state can change by applying evolution rules from P_i, which are selected according to mapping φ_i and depend on the state of the environment. A_i can modify the state of the environment by applying some of its action rules from R_i, which are selected by mapping ψ_i and depend both on the state of the environment and on the state of the agent itself. Start/Stop conditions of A_i are determined by π_i and ρ_i, respectively. A_i starts/stops its actions if context matches π_i and ρ_i.

CGSs intend to describe dialogue as a sequence of context-change-actions allowed by the current environment and performed by two or more agents.

Definition 2. By an action of an active agent A_i in state $\sigma = (w_E; w_1, w_2, \ldots, w_n)$ we mean a direct derivation step performed on the environmental state w_E by the current action rule set $\psi_i(w_E, w_i)$ of A_i.

Definition 3. A state of a conversational grammar system $\sum = (E, A_1, \ldots, A_n)$, $n \geq 2$, is an n+1-tuple: $\sigma = (w_E; w_1, \ldots, w_n)$, where $w_E \in V_E{}^*$ is the state of the environment, and $w_i \in V_i{}^*$, $1 \leq i \leq n$, is the state of agent A_i.

This rule is applied *by an active agent* and it is a rule selected by $\psi_i(w_E, w_i)$. We define an *active agent* in relation to the allowable actions it has at a given moment. That is, an agent can participate in conversation –being, thus, active— only if its set of allowable actions at that moment is nonempty:

Definition 4. An agent A_i is said to be active in state $\sigma = (w_E; w_1, w_2, \ldots, w_n)$ if the set of its current action rules, that is, $\psi_i(w_E, w_i)$, is a nonempty set.

Since conversation in CGS is understood in terms of *context changes*, we have to define how the environment passes from one state to another as a result of agents' actions.

Definition 5. Let $\sigma = (w_E; w_1, \ldots, w_n)$ and $\sigma' = (w'_E; w'_1, \ldots, w'_n)$ be two states of a conversational grammar systems $\sum = (E, A_1, \ldots, A_n)$. We say that σ' arises from σ by a simultaneous action of active agents $A_{i1}, \ldots, A_{ir}$, where $\{i_1, \ldots, i_r\} \subseteq \{1, \ldots, n\}$,

$i_j \neq i_k$, for $j \neq k$, $1 \leq j, k \leq r$, onto the state of the environment w_E, denoted by $\sigma \Rightarrow^a{}_\Sigma$ σ' iff:

- $w_E = x_1x_2...x_r$ and $w'_E = y_1y_2...y_r$, where x_j directly derives y_j by using current rule set ψ_I (w_E, w_{ij}) of agent A_{ij}, $1 \leq j \leq r$;

there is a derivation:

- $w_E = w_0 \Rightarrow {}^{a*}{}_{Ai1}$ $w1 \Rightarrow {}^{a*}{}_{Ai2}$ $w2 \Rightarrow {}^{a*}{}_{Ai3} \ldots \Rightarrow {}^{a*}{}_{Air}$ $w_r = w'_E$

such that, for $1 \leq j \leq r$, $\pi_{ij}(w_{j-1})$ = true and ρ_{ij} (w_j) = true. And for $f \in \{t, \leq k, \geq k\}$ the derivation is:

- $w_E = w_0 \Rightarrow {}^{af}{}_{Ai1}$ $w1 \Rightarrow {}^{af}{}_{Ai2}$ $w2 \Rightarrow {}^{af}{}_{Ai3} \ldots \Rightarrow {}^{af}{}_{Air}$ $wr = w'_E$

such that, for $1 \leq j \leq r$, $\pi_{ij}(w_{j-1})$ = true, and $w'_I = w_i$, $1 \leq I \leq n$.

Definition 6. Let $\Sigma = (E, A_1,...,A_n)$ be a conversational grammar system. And let $w_E = x_1x_2...x_r$ and $w'_E = y_1y_2...y_r$ be two states of the environment. Let consider that w'_E directly derives from w_E by action of active agent A_i, $1 \leq i \leq n$, as shown in Definition 5. We write that:

- $w_E \Rightarrow^{\leq k}{}_{Ai}$ w'_E iff $w_E \Rightarrow^{\leq k'}{}_{Ai}$ w'_E, for some $k' \leq k$;
- $w_E \Rightarrow^{\geq k}{}_{Ai}$ w'_E iff $w_E \Rightarrow^{\leq k'}{}_{Ai}$ w'_E, for some $k' \geq k$;
- $w_E \Rightarrow^{*}{}_{Ai}$ w'_E iff $w_E \Rightarrow^{k}{}_{Ai}$ w'_E, for some k;
- $w_E \Rightarrow^{t}{}_{Ai}$ w'_E iff $w_E \Rightarrow^{*}{}_{Ai}$ w'_E, and there is no $z \neq y$ with $y \Rightarrow^{*}{}_{Ai}$ z.

In words, $\leq$ *k*-derivation mode represents a time limitation where A_i can perform at most *k* successive actions on the environmental string. $\geq$ *k*-derivation mode refers to the situation in which A_i has to perform at least *k* actions whenever it participates in the derivation process. With *-mode, we refer to such situations in which agent A_i performs as many actions as it wants to. And finally, *t*-derivation mode represents such cases in which A_i has to act on the environmental string as long as it can.

However, in the course of a dialogue, agents' states are also modified and the environmental string is subject to changes due to reasons different from agents' actions. So, in order to complete our formalization of dialogue development, we add the following definition:

Definition 7. Let $\sigma = (w_E;w_1,\ldots,w_n)$ and $\sigma' = (w'_E;w'_1,\ldots,w'_n)$ be two states of a conversational grammar system $\Sigma = (E,A_1,\ldots,A_n)$. We say that σ' arises from σ by an evolution step, denoted by $\sigma \Rightarrow^e{}_\Sigma$ σ', iff the following conditions hold:

- w'_E can be directly derived from w_E by applying rewriting rule set P_E;
- w'_i can be directly derived from w_i applying rewriting rule set $\varphi_i(w_E)$, $1 \leq i \leq n$.

Definition 8. Let $\Sigma = (E, A_1,...,A_n)$ be a conversational grammar system as in Definition 1. Derivation in Σ terminates in:

- Style (ex) iff for $A_1,...,A_n$, $\exists A_i$: $w_i \in T_i$, $1 \leq i \leq n$;
- Style (all) iff for $A_1,...,A_n$, $\forall A_i$: $w_i \in T_i$, $1 \leq i \leq n$;
- Style (one) iff for $A_1,...,A_n$, A_i: $w_i \in T_i$, $1 \leq i \leq n$.

According to the above definition, derivation process ends in style *(ex)* if there is *some* agent A_i that has reached a terminal string. It ends in style *(all)* if *every* agent in the system has a terminal string as state. And it finishes in style *(one)* if there is *one* distinguished agent whose state contains a terminal string.

In CGS, the development of dialogue implies that both the *state of the environment* and *state of agents* change. Such changes take place thanks to two different types of processes: *action steps* and *evolution steps*. At the end, what we have is a *sequence of states* reachable from the initial state by performing, alternatively, action and evolution derivation steps:

Definition 9. Let $\sum = (E,A_1,\ldots,A_n)$ be a conversational grammar system and let σ_0 be a state of $\sum$. By a state sequence (a derivation) starting from an initial state σ_0 of $\sum$ we mean a sequence of states $\{\sigma_i\}^{\infty}_{i=0}$, where:

- $\sigma i \Rightarrow^{a}_{\sum} \sigma_{i+1}$, for $i = 2j$, $j \geq 0$; and
- $\sigma i \Rightarrow^{e}_{\sum} \sigma_{i+1}$, for $i = 2j + 1$, $j \geq 0$.

Definition 10. For a given conversational grammar system $\sum$ and an initial state σ_0 of $\sum$, we denote the set of state sequences of $\sum$ starting from σ_0 by Seq $(\sum,\sigma_0)$.

The set of environmental state sequences is:

- $\text{Seq}_E (\sum,\sigma_0) = \{\{w_{Ei}\}^{\infty}_{i=1} \mid \{\sigma_i\}^{\infty}_{i=0} \in \text{Seq} (\sum,\sigma_0), \sigma_i = (w_{Ei}; w_{1i},\ldots, w_{ni})\}$.

The set of state sequences of the j-th agent is defined by:

- $\text{Seq}_j(\sum,\sigma_0) = \{\{w_{ji}\}^{\infty}_{i=1} \mid \{\sigma_i\}^{\infty}_{i=0} \in \text{Seq} (\sum,\sigma_0), \sigma_i = (w_{Ei}; w_{1i},\ldots, w_{ji},\ldots, w_{ni})\}$.

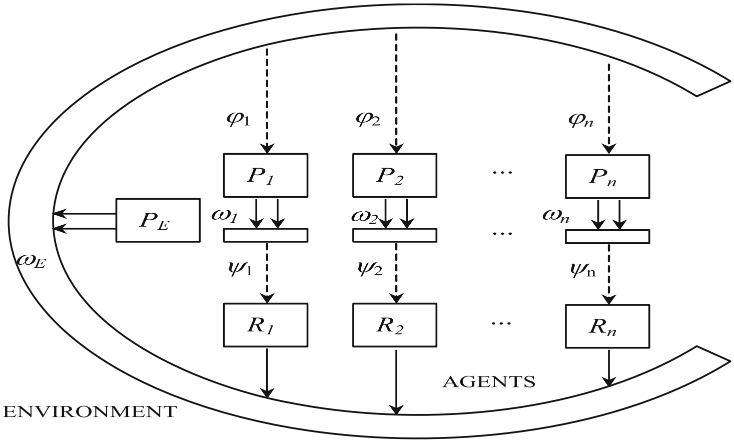


Fig. 1. Conversational grammar systems

Definition 11. For a given conversational grammar system $\sum$ and an initial state σ_0 of $\sum$, the language of the environment is:

- $L_E(\Sigma,\sigma_0) = \{w_E \in V^*_E \mid \{\sigma_i\}^{\infty}_{i=0} \in \text{Seq}\ (\Sigma,\sigma_0), \sigma_i = (w_{Ei}; w_1,\ldots, w_n)\}$.

and the language of j-th agent is:

- $L_j(\Sigma,\sigma_0) = \{w_j \in V^*_A \mid \{\sigma_i\}^{\infty}_{i=0} \in \text{Seq}\ (\Sigma,\sigma_0), \sigma_i = (w_{Ei}; w_1,\ldots w_j,\ldots, w_n)\}$, for $j = 1,2,\ldots, n$.

The formal apparatus constitutes what we have called conversational grammar system. Figure 1 offers a graphic view of the model.

Features such as *cooperation, coordination, emergence, dynamism* and flexibility and its capacity to capture the main elements and mechanisms in a dialogue, let us to consider CGS as a conversational-like model that emulates human linguistic behaviour in some specific situations such as natural language interfaces. In a CGS, the essential structure in conversation can be well and easily formalized. The most important elements and mechanisms at work in conversation are modelled by means of formal tools. The conversational setting has been defined as one in which there are two elements: *context* and *agents*. Both are described at any moment of time by a string of symbols over a specific alphabet. The functioning of the system --development of conversation-- is understood in terms of *state changes*. Both context and agents change their state in the course of conversation. A change in agents' strings is due to the updating of agents' state according to what is happening in the conversation. On the other hand, modification on environmental string is due both to agents' actions and to its own evolution. In order to allow agents to change the state of the environment, we have endowed them with sets of action rules. Action to be performed at any moment of time is selected according to the state of the environment and the state of the agent itself. In this way, we guarantee that agents act appropriately, in accordance with what is required by the state of conversation at a given moment. So, conversation proceeds by alternating between *action* and *evolution* steps. What we have at the end is a *sequence of states*, that is, a series of states that have been reached, from the initial state, during conversational interchange. Two formal tools, namely derivation modes and stop/start conditions, capture – in a very simple way – the turn-taking found in conversation which is considered as one of its constitutive features. Mapping ψ_i has been taken as the formal language counterpart of adjacency pairs and similar notions: it constrains the set of allowable actions according to the current state of the context. Different ways of closing a conversation have been defined, where closing is understood as the reaching of a terminal string. Coherence, dynamism and emergence in conversation are also preserved in our model. By limiting agents in CGS to apply only such rules included in ψ_i we keep coherence. The lack of any external control and mappings φ_i and ψ_i make of CGS a dynamic, flexible and emergent framework that accounts for the unplanned and opportunistic nature of conversation.

3 An Example

The following fragment of a dialogue illustrates how the dialogue-based interaction system we have introduced above works. Dialogues of this type can be handled by our system in a simple way.

Let us consider the following dialogue that can take place between a customer (the user) who wants to buy a table and the shop assistant (the system) that guides the customer in his choice. The interaction is collaborative, with neither the system nor the user being in control of the whole interaction. Each of them contributes to the interaction when they can.

User: Good morning.
System: Good morning. Can I help you?
User: Yes, I would like to buy a table.
System: Which type of table, for which room?
User: I need a living room table.
System: We have different shapes, which do you prefer?
User: I would like a rectangular table.
System: What about the material?
User: I would like a glass table.
System: We have different glass colours, any preference?
User: I would like a transparent glass table.
System: I guess this is the table you are looking for.
User: Yes, it is wonderful. Thank you very much.
System: You are welcome.

Starting from this example, we define the following CGS with two agents. The alphabet of the environment is:

- **I** is the initial state
- **h** stands for "Hello"
- **Q** stands for "Can I help you?"
- **t** stands for "I need a table"
- **R** stands for "What kind of room?"
- **l** stands for "A living room"
- **S** stands for "What shape?"
- **r** stands for "Rectangular"
- **M** stands for "What material?"
- **g** stands for "Glass"
- **C** stands for "Colour?",
- **w** stands for "White",
- **K** stands for "OK, something else?"
- **e** stands for "End of my request, bye"
- **B** stands for "Bye"

The environment is in the initial state *I*. For this dialogue we do not need special functions φ, they just copy the environment to the agents' status.

```
Hashtable R1 = new Hastable ();
if (environment.Text.Equals("I")) {R1.Add("I","h");}
else if (environment.Text.Equals.EndsWith("Q"))
{R1.Add("Q","Qt");}
else if (environment.Text.Equals.EndsWith("R"))
{R1.Add("R","R1");}
else if (environment.Text.Equals.EndsWith("S"))
{R1.Add("S","Sr");}
else if (environment.Text.Equals.EndsWith("M"))
{R1.Add("M","Mg");}
else if (environment.Text.Equals.EndsWith("C"))
{R1.Add("C","Cw");}
else if (environment.Text.Equals.EndsWith("K"))
{R1.Add("K","Ke");}
else if (environment.Text.Equals.EndsWith("b"))
{R1.Add("b","bB");}
```

Fig. 2. Implementation in C# of the function ψ_1

The function ψ_1 adds a single rule R_1 to its set of rules, as follows:

- $R_1 = \{I \rightarrow h\}$ for $\omega_E = xI$, $x \in V^*_E$,
- $R_1 = \{Q \rightarrow Qt\}$ for $\omega_E = xQ$, $x \in V^*_E$,
- $R_1 = \{R \rightarrow Rl\}$ for $\omega_E = xR$, $x \in V^*_E$,
- $R_1 = \{S \rightarrow Sr\}$ for $\omega_E = xS$, $x \in V^*_E$,
- $R_1 = \{M \rightarrow Mg\}$ for $\omega_E = xM$, $x \in V^*_E$,
- $R_1 = \{C \rightarrow Cw\}$ for $\omega_E = xC$, $x \in V^*_E$,
- $R_1 = \{K \rightarrow Ke\}$ for $\omega_E = xK$, $x \in V^*_E$,
- $R_1 = \{b \rightarrow bB\}$ for $\omega_E = xb$, $x \in V^*_E$.

The function ψ_2 is defined in a similar way:

- $R_2 = \{h \rightarrow hhQ\}$ for $\omega_E = xh$, $x \in V^*_E$,
- $R_2 = \{t \rightarrow tR\}$ for $\omega_E = xt$, $x \in V^*_E$,
- $R_2 = \{l \rightarrow l\ S\}$ for $\omega_E = xl$, $x \in V^*_E$,
- $R_2 = \{r \rightarrow rM\}$ for $\omega_E = xr$, $x \in V^*_E$,
- $R_2 = \{g \rightarrow gC\}$ for $\omega_E = xg$, $x \in V^*_E$,
- $R_2 = \{w \rightarrow wK\}$ for $\omega_E = xw$, $x \in V^*_E$,
- $R_2 = \{e \rightarrow eb\}$ for $\omega_E = xe$, $x \in V^*_E$.

This implementation gives only one possible dialogue that is:

$$I \rightarrow hhQtRlSrMgCwKebB.$$

If we want to get different dialogues on the same topic, we can imagine that after asking about the table, the dialogue continues with the specifications of details, colour, shape, material, room in an arbitrary order like:

$$I \rightarrow hhQtRlCwMgSrKebB \text{ or } I \rightarrow hhQtSrRlMgCwKebB.$$

To do this, we keep the definition of the function ψ_1 and we should modify a little bit the function ψ_2 in the following way:

- $R_2 = \{a \rightarrow aB\}$ for $\omega_E = xa$, $x \in V^*_E$, $a \in \{t,l,r,g,m,w\}$ and B randomly selected from the set $\{R,S,M,C\}$.

```
Random r = new Random();
if (environment.Text.Equals("h")) { R2.Add("h", "hhQ"); }
else if (environment.Text.EndsWith("t") ||
    environment.Text.EndsWith("l") ||
    environment.Text.EndsWith("r") ||
    environment.Text.EndsWith("g") ||
    environment.Text.EndsWith("m") ||
    environment.Text.EndsWith("w"))
    {
       String mc = "";
       if (environment.Text.LastIndexOf("R") < 0) { mc += "R"; }
       if (environment.Text.LastIndexOf("S") < 0) { mc += "S"; }
       if (environment.Text.LastIndexOf("M") < 0) { mc += "M"; }
       if (environment.Text.LastIndexOf("C") < 0) { mc += "C"; }

       String lc = environment.Text[environment.Text.Length - 1].ToString();
       lc = lc.ToLower();
       if (mc.Length == 0)
       {  R2.Add(lc, lc + "K");}
       else
       {
          int i = r.Next(mc.Length);
          R2.Add(lc, lc + mc[i].ToString());
       }
    }
else if (environment.Text.EndsWith("e")) { R2.Add("e", "eb"); }//end, bye
```

Fig. 3. Implementation in C# the modifications produced in the environment by the function ψ_2

We can see as an example the implementation code in C# for functions ψ_1 in Figure 2 and ψ_2 with random discussions in Figure 3. Applying productions is implemented by the function *applyPrductions,* whose implementation is given in Figure 4.

```
private String applyPrductions(String s, Hashtable P1)
{
     String t = "";
     for (int i = 0; i < s. Length ; i++)
     {String st=s[i].ToString() ;
              ...
              if (P1.ContainsKey(st))
              {       t += Pl[st];}
              else
              {       t += st;}
     }
     return t;
}
```

Fig. 4. Implementation in C# the function *applyProductions*

The result is quite surprising. It seems that for normal conversations we do not need the status of the agents and the functions φ or the rewritten of the environment with P_E productions. We need only the functions ψ and the environment.

We observe that in our system, the language of the environment has monotonically increasing length (being therefore context sensitive).

Actually we could transform our system (the simplified version without permutations of the dialogue sentences) into a cooperating distributed grammar system working in a terminal mode derivation, described in (Csuhaj-Varjú et al. 1997). To do that, we should have introduced distinct non-terminals for each agent. There are many theoretical results about CD grammars, one saying that there is a hierarchy of languages hierarchy given by the number of grammars. Another important result is the following theorem.

Theorem 1. [Cooperative Distributed (CD) Grammars] *Two context-free grammars cooperating under the terminal mode of derivation can always be replaced by a single context-free grammar. Three context-free grammars cooperating under the terminal mode are as powerful as ET0L systems.*

It might be the case that we need the status of the agents to realize some kind of synchronization and to prevent two agents to speak in the same time. We consider these aspects as promising directions for future research.

4 Final Remarks

According to [21] human language technology plays a central role in providing an interface that will drastically change the human-machine communication paradigm from *programming* to *conversation*, enabling users to efficiently access, process, manipulate and absorb a vast amount of information. Effective conversational interfaces must incorporate extensive and complex dialogue modelling. In this paper, we have introduced a formal model of dialogue that may contribute to the building of more effective and efficient human-computer interaction tools through the simulation of human-human conversations.

The dialogue system we have introduced can be considered a mixed-initiative interaction model in which there is a dynamic exchange of control of the dialogue flow. CGS are able to model dialogue with a high degree of flexibility, what means that they are able to accept new concepts and modify rules, protocols and settings during the computation. The main characteristic of the model is the use of simple grammars in order to generate a dialogue structure. It should not to be seen as a psychologically realistic cognitive model, but as a model that might successfully emulate human linguistic behaviour in some specific situations such as natural language interfaces. CGS can be considered a generic dialogue system that can be adapted to different tasks, according to the idea of (Allen et al. 2001) that while practical dialogues in different domains may appear quite different at first glance, they all share essentially the same underlying structures.

Effective implementation on CGS can lead to design and simulate the emergence of consciousness and linguistic capabilities in both humans and computers. This is why future efforts in this research will focus on the development of efficient strategies of communication, turn-taking protocols and closing algorithms as well as the inclusion of semantic items in the dialogue games.

The final goal, to achieve successful communicative exchanges between humans and computers, is getting closer every day.

References

1. Allen, J.F., Byron, D.K., Dzikovska, M., Ferguson, G., Galescu, L., Stent, A.: Towards conversational human-computer interaction. AI Magazine 22(4), 27–37 (2001)
2. Bunt, H.: Non-problems and social obligations in human-computer conversation. In: Proc. 3rd Intern. Workshop on Human-Computer Conversation, Bellagio, Italy (2000)
3. Bunt, H.: Designing an open, multidimensional dialogue act taxonomy. In: Proc. 9th Intern. Workshop on Semantics and Pragmatics of Dialogue (Dialor 2005), Nancy, France (2005)
4. Te'eni, D., Carey, J.M., Zhang, P.: Human computer interaction: developing effective organization information systems. Wiley & Sons Inc., New York (2006)
5. Clark, H.H.: Using languages. Cambridge University Press, Cambridge (1996)
6. Cohen, P.: Dialogue modeling. In: Cole, R., Mariani, J., Uszkoreit, H., Varile, G., Zaenen, A., Zampolli, A., Zue, V. (eds.) Survey of the state of the art in human language technology (1998), http://cslu.cse.ogi.edu/HLTsurvey/HLTsurvey.html (accessed May 6, 2009)
7. Csuhaj-Varjú, E., Dassow, J., Kelemen, J., Păun, G.: Grammar systems: a grammatical approach to distribution and cooperation. Gordon and Breach, London (1994)
8. Csuhaj-Varjú, E., Kelemen, J., Kelemenová, A., Păun, G.: Eco-grammar systems: a grammatical framework for studying lifelike interactions. Artificial Life 3, 1–28 (1997)
9. Dix, A.J., Finlay, J.E., Abowd, G.D., Beale, R.: Human-computer interaction. Prentice Hall, Upper Saddle River (2004)
10. Kirlik, A.: Adaptive perspectives on human-technology interaction: methods and models for cognitive engineering. Oxford University Press, Oxford (2006)
11. Larsson, S.: Dialogue systems: simulations or interfaces? In: Proc. 9th Intern. Workshop on Semantics and Pragmatics of Dialogue, Dialor 2005 (2005), http://www.ling.gu.se/~sl/activities.html (accessed April 14, 2009)
12. Lazar, J.: Universal usability: designing computer interfaces for diverse user populations. Wiley & Sons Inc., New York (2007)
13. Moulin, B., Rousseau, D., Lapalme, G.: A multi-agent approach for modeling conversations. In: Proc. 14th Intern. Conference on Natural Language Processing (AI 1994), Paris, vol. 3, pp. 35–50 (1994)
14. Ogden, W.C., Bernick, P.: Using natural language interfaces. In: Helander, M.G., Landauer, T.H., Prabhu, P.V. (eds.) Handbook of human-computer interaction, pp. 137–161. Elsevier, Amsterdam (1997)
15. Reichman, R.: Getting computers to talk like you and me. Discourse context, focus, and semantics (An ATN model). MIT Press, Cambridge (1985)
16. Rozenberg, G., Salomaa, A.: Handbook of formal languages. Springer, Heidelberg (1997)
17. Searle, J.: Speech acts: an essay in the philosophy of language. Cambridge University Press, Cambridge (1969)
18. Sears, A.: Handbook for human computer interaction. CRC Press, Boca Raton (2007)
19. Sharp, H., Rogers, Y., Preece, J.: Interaction design: beyond human-computer interaction. Wiley & Sons Inc., New York (2007)
20. Smith-Atakan, S.: Human computer interaction. Thomson Learning Co., Andover (2006)
21. Zue, E.: Conversational interfaces: advances and challenges. In: Proc. 5th European Conference on Speech Communication and Technology (Eurospeech 1997), vol. 1, pp. 9–18 (1997)

Consistency-Based vs. Similarity-Based Prediction Using Extensions of Information Systems – An Experimental Study

K. Pancerz[1,2]

[1] Institute of Biomedical Informatics,
University of Information Technology and Management, Rzeszów, Poland
kpancerz@wsiz.rzeszow.pl
[2] Chair of Computer Science and Knowledge Engineering,
Zamość University of Management and Administration, Zamość, Poland

Abstract. The paper is devoted to the application of the extensions of information systems to solve prediction problems. An information system (in the Pawlak's sense [4]) can describe the states of processes observed in a given system of concurrent processes. If we extend a given information system by adding some new states which have not been observed yet, then we are interested in the degrees of consistency (called also consistency factors) of added states with the knowledge of state coexistence included in the original information system. State coexistence is expressed by a proper kind of rules extracted from the information system. Such information can be helpful in predicting the possibility of appearing given states in the future in the examined system. The consistency factor computed can be between 0 and 1, 0 for the full inconsistency and 1 for the full consistency. Consistency-based prediction is compared with prediction based on a simple similarity measure. The experiments show that the states from extensions of original information systems, having greater values of consistency factors, appear significantly more often in the future. Consistency-based prediction seems to be a promising alternative for similarity-based prediction.

1 Introduction

The notion of partially consistent extensions of information systems has been introduced in [11]. In fact, it is a generalization of the notion of consistent extensions of information systems considered, among others, in [8, 10]. A partially consistent extension of a given information system consists of all objects which appeared in the original system and those with known attribute values which can be added to the system. A new object added to the original information system can be consistent with the knowledge extracted from the system and expressed by rules, but not necessarily. We allow the situation that some objects can be consistent with this knowledge only partially. The essential thing is to determine a consistency factor of

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 67–81.
springerlink.com

a new object with the knowledge included in the original system. The approach to computing a consistency factor has been proposed in [11]. An information system can be used to describe a system consisting of some separated processes (cf. [5]). Such a system is called a system of concurrent processes. Then, a partially consistent extension of an information system can be used for predicting future states which can appear in the system of concurrent processes. Intuitively, we suppose that states with the greater consistency factor should appear more often in the future. Our assumption has been verified on the experimental data coming from economy and meteorology. The results show that the formulated hypothesis is close to the truth. Experiments show also that consistency-based prediction seems to be a promising alternative for similarity-based prediction.

The rest of the paper is organized as follows. A brief review of the basic concepts underlying the information systems and their extensions is given in Section 2. The experimental results are shown in Section 3. Finally, Section 4 consists of some conclusions and further plans.

2 Theoretical Background

First, we recall basic concepts concerning the information systems (in Pawlak's sense) [4] and their extensions (cf. [10, 11]) used in the paper.

2.1 Extensions of Information Systems

An information system is a pair $S=(U,A)$, where U is a nonempty, finite set of objects, called universe, A is a nonempty, finite set of attributes, i.e., a: $U\rightarrow V_a$ for a belonging to A, where V_a is called a value set of a.

The idea of a concurrent system representation by means of information systems has been proposed in [5]. Then, elements of the set U can be interpreted as global states of a given concurrent system CS, whereas attributes (elements of the set A) as processes in CS. For each process a from A, the set V_a of local states is associated with a. If the information system is represented by a data table, then the columns of a data table are labeled with names of processes. Each row of a data table (labeled with an element from the set U) includes a record of local states of processes of CS. Each record can be treated as a global state of CS. Therefore, in the rest of the paper, for an information system S describing a concurrent system, we can interchangeably use the following terminology: attributes of S are called processes in S, objects of S are called global states in S, and values of attributes of S are called local states of processes in S. It is worth noting that a notion of concurrent systems can be understood widely. In a general case, a concurrent system is a system consisting of some processes whose local states can coexist together and they are partly independent. For example, we can treat as concurrent systems the ones consisting of economic processes, financial processes, biological processes, genetic processes, meteorological processes, and so forth.

We can represent the knowledge included in a given information system by means of rules in the form of *IF* ... *THEN* ... In the proposed approach, we consider rules in the form:

r: $IF[(a_{i_1}, v_{i_1}) AND (a_{i_2}, v_{i_2}) AND ... AND (a_{i_k}, v_{i_k})] THEN [(a_d, v_d)]$

where $a_d \in A, v_d \in V_{a_d}, a_{i_j} \in A - \{a_d\}$, and $v_{i_j} \in V_{a_{i_j}}$ for $j = 1,...,k$. Moreover, *AND* is understood as a classical propositional connective called conjunction. Each pair (*a*,*v*) in a rule is called a descriptor. Each descriptor determines a considered local state of a given process, i.e., the situation while the process *a* has the state *v*. The left part of a rule, i.e., $(a_{i_1}, v_{i_1}) AND (a_{i_2}, v_{i_2}) AND ... AND (a_{i_k}, v_{i_k})$ is referred to as the predecessor of a rule whereas the right part of a rule, i.e., (a_d, v_d) is referred to as a successor of a rule.

The knowledge expressed by rules is the useful information on coexistence of local states of processes in the examined system. If we have a rule *r*, then we have the following knowledge: if the process a_{i_1} has the state v_{i_1} , the process a_{i_2} has the state v_{i_2} , ..., the process a_{i_k} has the state v_{i_k} , then the process a_d may have the state v_d .

Rules considered by us satisfy three requirements, namely each rule should be true, minimal and realizable in a given information system.

A rule *r* is called true in the information system *S*=(*U*,*A*) if and only if for each object $u \in U$ if $a_{i_1}(u) = v_{i_1}$ and $a_{i_2}(u) = v_{i_2}$ and ... and $a_{i_k}(u) = v_{i_k}$ then also $a_d(u) = v_d$. If we have $a_{i_1}(u) = v_{i_1}$ and $a_{i_2}(u) = v_{i_2}$ and ... and $a_{i_k}(u) = v_{i_k}$ and $a_d(u) \neq v_d$ for an object $u \in U$, then we say that a rule *r* is not satisfied by the object *u*. This fact will be denoted by $u \not\models r$.

A rule *r* is called minimal in *S*=(*U*,*A*) if and only if removing any descriptor from the predecessor of a rule makes this rule not true in *S*. The set of all minimal rules true in *S* will be denoted by *Rul*(*S*).

A rule *r* is called realizable in *S*=(*U*,*A*) if and only if there exists at least one object $u \in U$ such that $a_{i_1}(u) = v_{i_1}$ and $a_{i_2}(u) = v_{i_2}$ and ... and $a_{i_k}(u) = v_{i_k}$ and $a_d(u) = v_d$. Any object satisfying this requirement is referred to as the object supporting a rule *r*. The set of all objects from *U* supporting a rule *r* from *Rul*(*S*) will be denoted by *U*$_r$. The strength factor of a rule *r* in the information system *S* is defined as:

$$str_S(r) = \frac{card(U_r)}{card(U)}.$$

In some applications, it is not advisable to take into consideration all minimal rules for a given information system. For example, rules with a low strength factor can be too detailed and they can be treated as a certain kind of an information noise. Therefore, we can consider a set of rules such that each rule in this set has a strength factor greater or equal to a fixed threshold value τ, where $0 \leq \tau \leq 1$.

Let S be an information system, and $Rul(S)$ be the set of all minimal rules true in S. By $StrRul^{(\tau)}(S)$ we denote the set of all τ-strong minimal rules true in S such that $StrRul^{(\tau)}(S) = \{r \in Rul(S) : str_S(r) \geq \tau\}$, where $0 \leq \tau \leq 1$.

Other important notions apply to extensions of information systems.

Definition 1 (Extension). *Let S=(U,A) be an information system. An information system S^*=(U^*,A^*) is an extension of S if and only if the following conditions are satisfied:*

- $U \subseteq U^*$,
- $card(A) = card(A^*)$,
- *for each* $a \in A$, *there exists* $a^* \in A^*$ *such that a function* $a^* : U^* \to V_a$ *is an extension of a function* $a : U \to V_a$ *to* U^*.

Definition 2 (Cartesian extension). *Let S=(U,A) be an information system,* $\{V_a\}_{a \in A}$ *the family of value sets of attributes from A, U' a universe including U. An information system* $S^\#$=($U^\#$,$A^\#$) *such that:*

- $U^\# = \{u \in U' : \underset{a^* \in A^\#}{\forall} a^*(u) \in V_a\}$,
- *for each* $a \in A$, *there exists* $a^* \in A^\#$ *such that a function* $a^* : U^\# \to V_a$ *is an extension of a function* $a : U \to V_a$ *to* $U^\#$.

  is called a Cartesian extension of S.

It is worth mentioning that the Cartesian extension of a given information system is a maximal (with respect to the number of objects) extension of this system.

Remark 1. *Let S=(U,A) be an information system,* S^*=(U^*,A^*) *its extension, and* $S^\#$=($U^\#$,$A^\#$) *its Cartesian extension. For the rest of the paper, the sets A,* A^**, and* $A^\#$ *will be marked with the same letter A. This inaccuracy seems to be not important in further considerations.*

2.2 Consistency Measure

For each object from the extension of a given information system S we can define the so-called consistency factor with the knowledge included (hidden) in S. Proper definitions are given below.

Let S=(U,A) be an information system and $Rul'(S) \subseteq Rul(S)$ a set of rules true in S.

By $K_{Rul'(S)}$ we denote the knowledge about *S* included in *Rul'*(*S*). By $\overline{Rul'_u}(S)$ we denote a set of all rules from *Rul'*(*S*) which are not satisfied by an object *u* from an extension $S^*=(U^*,A)$ of *S*, i.e., $\overline{Rul'_u}(S)=\{r \in Rul'(S): u \not\models r\}$.

The strength of a set $\overline{Rul'_u}(S)$ of rules in *S* is computed as follows:

$$str\left(\overline{Rul'_u}(S)\right)=\frac{card(U_{\overline{Rul'_u}(S)})}{card(U)}$$

where:

$$U_{\overline{Rul'_u}(S)} = \bigcup_{r\in\overline{Rul'_u}(S)} U_r$$

For each object *u* from an extension S^* of *S*, we assign a consistency factor of *u* with the knowledge $K_{Rul'(S)}$ defined below.

Definition 3 (Consistency factor). *Let S=(U,A) be an information system, S*=(U^*,A) its extension, $Rul'(S)\subseteq Rul(S)$, and $u \in U^*$. The consistency factor of u with the knowledge $K_{Rul'(S)}$ is defined as*

$$\xi_{K_{Rul'(S)}}(u)=1-str\left(\overline{Rul'_u}(S)\right).$$

We have $0 \leq \xi_{K_{Rul'(S)}} \leq 1$ for each $u \in U^*$. It is obvious that if $u \in U$, then $\xi_{K_{Rul'(S)}}=1$ because $\overline{Rul'_u}(S)=\varnothing$.

Algorithms for computing consistency factors for new objects added to original information systems have been presented in [13]. It is worth noting that there exists algorithm which does not need computing any rules in information systems. This is an important property from the computational complexity point of view, especially, if we have high dimensional data (for example, in genetics).

Now, we recall an efficient algorithm (given in [13]) (see Algorithm 1) for computing a consistency factor of any object u^* from the extension of a given information system *S* with the knowledge included in *S* and expressed by all minimal rules true and realizable in *S* (the set *Rul*(*S*)). The algorithm presented here allows us to determine a set of objects from an original information system *S* supporting minimal rules from *Rul*(*S*), but not satisfied by the object u^*. A consistency factor is computed as a complement to 1 of the strength of the set of rules not satisfied. This algorithm takes advantage of the theorem proposed in [2]. Here, we recall this theorem.

Theorem 1. *Let S=(U,A) be an information system, S^*=(U^*,A) its extension, Rul(S) a set of all minimal rules true and realizable in S and $u^* \in U^*$. For each $u \in U$ let $M_u=\{a \in A: a(u^*)=a(u)\}$, and for each $a \in A-M_u$:*

$$P_u^a = \{a(u') : u' \in U \text{ and } \underset{a' \in M_u}{\forall} a'(u') = a'(u)\}.$$

The object u^* satisfies all rules from *Rul*(*S*) if and only if for any $u \in U$ we have $card(P_u^a) \geq 2$ for each $a \in A - M_u$.

Proof. For the proof see [2, 13].

We immediately obtain the following corollary.

Corollary 1. The object u^* satisfies all rules from *Rul*(*S*) which are supported by an object $u \in U$ if and only if $card(P_u^a) \geq 2$ for each $a \in A - M_u$.

Algorithm 1. Algorithm for efficient computing a consistency factor of an object belonging to the extension of an information system.

Input: An information system *S*=(*U*,*A*), an object u^* belonging to the extension of *S*.

Output: A consistency factor $\xi_S(u^*)$ of the object u^* with the knowledge included in *S*.

```
begin
$\overline{U} \leftarrow \varnothing$;
$S_{orig} = (U_{orig}, A) \leftarrow S = (U, A)$;
for each $u \in U$ do
        for each $a \in A$ do
                if $a(u) \neq a(u^*)$ then
                        $a(u) \leftarrow *$;
Remove each object $u \in U$ such that $\underset{a \in A}{\forall} a(u) = *$;
for each $u \in U$ do
        $M_u = \{a \in A : a(u) \neq *\}$;
        for each $d \in A - M_u$ do
                $P_u^d = \{d(u') : u' \in U \text{ and } \underset{a' \in M_u}{\forall} a'(u') = a'(u)\}$, where $d(u')$ is
                determined on the basis of $S_{orig}$;
                if $card(P_u^d) = 1$ then
                        $\overline{U} \leftarrow \overline{U} \cup \{u\}$;
                        break;
$\xi_S(u^*) \leftarrow 1 - \frac{card(\overline{U})}{card(U)}$;
end.
```

2.3 Similarity Measure

For each new object added to a given information system *S* we can determine its similarity to objects from *S*. In the presented approach we consider information systems with discrete values of attributes. Therefore, we can define a simple similarity measure called a similarity factor. A similarity factor can be computed according to Algorithm 2.

Definition 4 (Similarity measure). *Let S=(U,A) be an information system, S*=(U^*,A) its extension, and* $u^* \in U^* - U$ *. The similarity factor of* u^* *to objects from U is defined as*

$$\sigma_U(u^*) = \frac{\sum\limits_{u \in U} \frac{card(\{a \in A : a(u^*) = a(u)\})}{card(A)}}{card(U)}.$$

Algorithm 2. Algorithm for computing a similarity factor of a new object belonging to the extension of an information system.

Input: An information system *S*=(*U*,*A*), a new object u^* belonging to the extension of *S*.

Output: A similarity factor $\sigma_U(u^*)$ of the object u^* to objects from *S*.

```
begin
$sim\_sum \leftarrow 0$;
for each $u \in U$ do
    $sim \leftarrow 0$;
    for each $a \in A$ do
        if $a(u) = a(u^*)$ then
            $sim \leftarrow sim + 1$;
    $sim\_sum \leftarrow \frac{sim}{card(A)}$;
$\sigma_U(u^*) = \frac{sim\_sum}{card(U)}$;
end.
```

3 Experiments

To perform experiments we use a software tool called ROSECON [12]. Our experiments have been performed on data coming from two domains: macroeconomics and meteorology. One can say that both of experiments confirm, on some conditions, our presumption that the state with the greater consistency factor with the original knowledge should appear more often in the future. Experiments have been carried out according to the following steps:

- We split a given information system into two parts. The first part was a training system S_{train}. The second part was a test system S_{test}.
- For the training system S_{train} we determined its Cartesian extension $S^{\#}_{train}$.
- For each new object from the Cartesian extension $S^{\#}_{train}$ we computed its consistency factor with the knowledge included in S_{train} and expressed by a chosen set $Rul'(S_{train})$ of rules.
- We checked our prediction on the test system S_{test}. For each new object from the Cartesian extension $S^{\#}_{train}$ of S_{train} we counted its appearances in the test system S_{test}.
- For each new object from the Cartesian extension $S^{\#}_{train}$ we computed its similarity factor to objects from S_{train}.
- We checked our prediction on the test system S_{test}. For each new object from the Cartesian extension $S^{\#}_{train}$ of S_{train} we counted its appearances in the test system S_{test}.

3.1 Macroeconomics Data

The first data set includes information on changes of monthly macroeconomics indexes like: unemployment (marked *unempl*), inflation (marked with *infl*), usd exchange rate (marked with *usd*), euro exchange rate (marked with *euro*), export (marked with *exp*) and import (marked with *imp*). An information system is built in the following way. Attributes correspond to indexes whereas objects correspond to consecutive months. For each attribute, its value set consists of three values: -1, 0, and 1. The meaning of these values is the following: -1 denotes decreasing a given index in relation to the previous month, 0 denotes remaining a given index on the same level in relation to the previous month, 1 denotes increasing a given index in relation to the previous month. An original data table consists of 88 objects (global states). A fragment of the information system is shown in Table 1.

Table 1. Information system (fragment) for macroeconomics indexes

U/A	*unempl*	*infl*	*usd*	*euro*	*exp*	*imp*
m_1	0	1	0	0	1	1
m_2	-1	-1	-1	-1	0	1
…	…	…	…	…	…	…
m_{88}	-1	-1	1	0	-1	-1

Experiments were repeated for different sets $StrRul^{(\tau)}(S)$ of rules, where $\tau = 0$, 0.05, 0.1, 0.15, 0.2, i.e., in each case, the knowledge included in the original information system was represented by means of rules having a strength factor greater or equal to a given threshold value. In each case, a training system included objects (global states) from m_1 to m_{40} whereas a test system included objects (global states) from m_{41} to m_{88}. Results of prediction tests (for each set $StrRul^{(\tau)}(S)$) are collected on graphs presented in Figures 1 - 5. On these graphs,

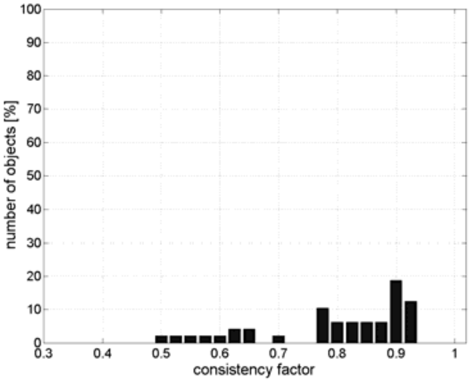


Fig. 1. Results of consistency-based prediction test for all rules

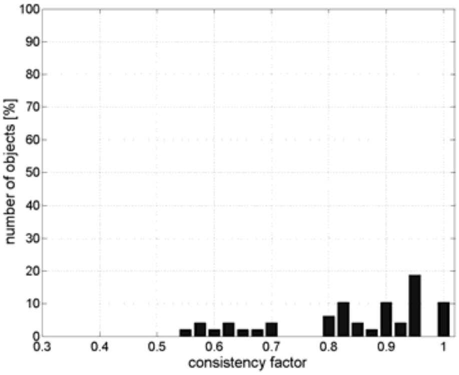


Fig. 2. Results of consistency-based prediction test for rules with minimum strength 0.05

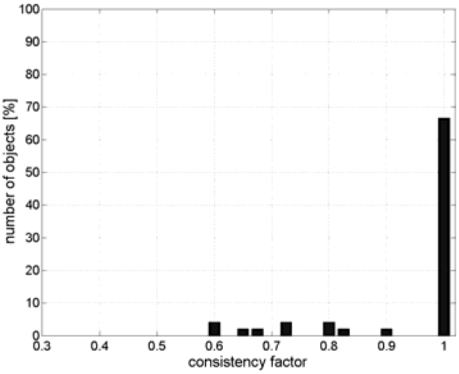


Fig. 3. Results of consistency-based prediction test for rules with minimum strength 0.10

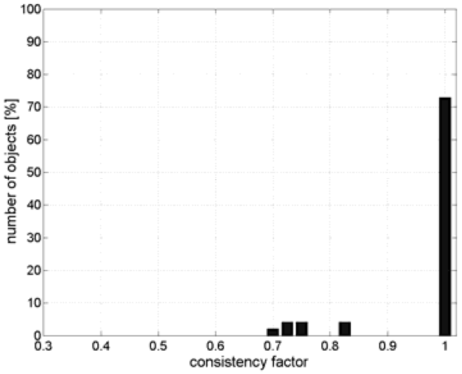


Fig. 4. Results of consistency-based prediction test for rules with minimum strength 0.15

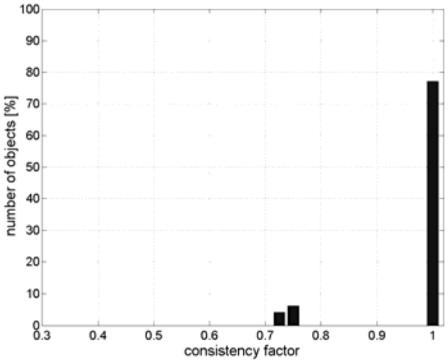


Fig. 5. Results of consistency-based prediction test for rules with minimum strength 0.20

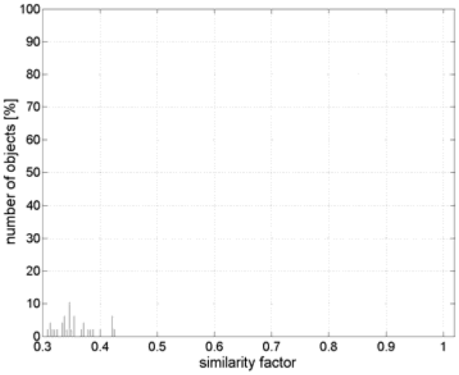


Fig. 6. Results of similarity-based prediction test

percentage appearances of objects in the test systems, having suitable values of consistency factors (computed for training systems), are presented. Analogously, an experiment using a similarity measure was conducted. A result of a prediction test is shown in Figure 6. On this graph, percentage appearances of objects in the test systems, having suitable values of similarity factors (computed for training systems), are presented.

3.2 Weather Data

The second data set includes information on the weather processes like temperature (marked with *temp*), dew point (marked with *dew*), humidity (marked with *hum*), pressure (marked with *press*), and wind speed (marked with *wind*). An information system is built in the following way. Attributes correspond to the weather processes whereas objects correspond to consecutive days. Sets of attribute values (local states of processes), after the discretization by means of the Equal Frequency Binning method from the ROSETTA software tool [6], are the following: V_{temp}={38,50,58} [F], V_{dew}={31,40,47} [F], V_{hum}={40,61,70,77} [%], V_{wind}={0,4,6,8,10} [mph], V_{press}={2906,2982,2995,3004,3016} [100×in]. An original data table consists of 122 objects (global states). A fragment of the information system is shown in Table 2.

Table 2. Information system (fragment) for weather processes

U/*A*	*temp*	*dew*	*hum*	*press*	*wind*
d_1	50	40	70	3004	6
d_2	50	47	61	3004	6
...	...	…	…	…	…
d_{122}	50	47	86	2995	4

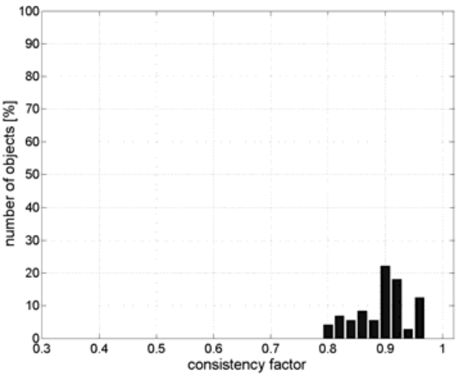


Fig. 7. Results of consistency-based prediction test w.r.t. all rules

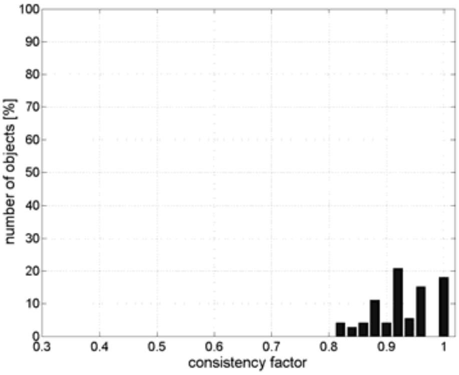


Fig. 8. Results of consistency-based prediction test for rules with minimum strength 0.025

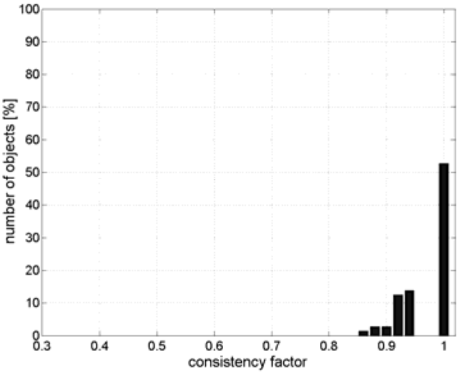


Fig. 9. Results of consistency-based prediction test rules with minimum strength 0.05

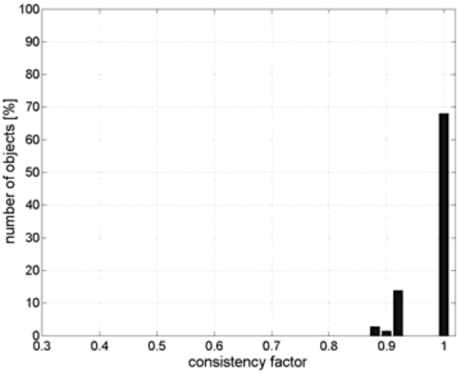


Fig. 10. Results of consistency-based prediction test for rules with minimum strength 0.075

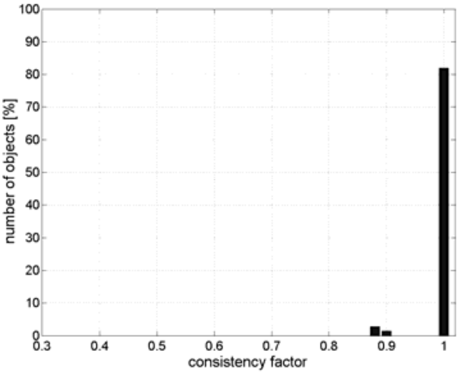


Fig. 11. Results of consistency-based prediction test for rules with minimum strength 0.1

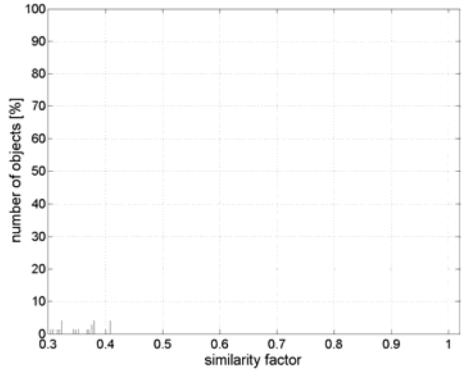


Fig. 12. Results of similarity-based prediction test

Experiments were repeated for different sets $StrRul^{(\tau)}(S)$ of rules, where $\tau = 0$, 0.025, 0.05, 0.075, 0.1, i.e., in each case the knowledge included in the original information system was represented by means of rules having a strength factor greater or equal to a given threshold value. In each case, a training system included objects (global states) from d_1 to d_{50} whereas a test system included objects (global states) from d_{51} to d_{122}. Results of prediction tests (for each set $StrRul^{(\tau)}(S)$) are collected on graphs presented in Figures 7-11. On these graphs, percentage appearances of objects in the test systems, having suitable values of consistency factors (computed for training systems), are presented. Analogously, an experiment using a similarity measure was conducted. A result of a prediction test is shown in Figure 12. On this graph, percentage appearances of objects in the test systems, having suitable values of similarity factors (computed for training systems), are presented.

3.3 Summary

Performing experiments on the real data by means of our approach we can draw the following conclusions. The quality of prediction depends on the chosen set of rules which express the knowledge included in information systems. If we take a set of all minimal rules true and realizable in a given information system, then it is difficult to observe the relationship between percentage appearances of objects in the test systems and values of consistency factors of these objects computed for training systems. Probably, rules with a low strength factor are too detailed and they can be treated as a certain kind of an information noise. On the one hand, if we omit rules with a strength factor below a given threshold value, then we can observe the relationship between percentage appearances of objects in the test systems and values of consistency factors of these objects computed for the training systems. However, on the other hand, the threshold value for a strength factor cannot be too large because then the set of rules expressing the knowledge included in the information system does not consists of many rules and the majority of new states added to the system is consistent with such knowledge to the degree 1.0. In this case, we obtain the majority of the new states can appear in the future with the same possibility and it is too weak information for prediction. If we use a simple similarity measure for prediction, we cannot see any regularity, i.e., it is difficult to observe that the relationship between percentage appearances of objects in the test systems and values of similarity factors of these objects computed for training systems. Obviously, it is needed to check other similarity measures, but consistency-based prediction seems to be a promising alternative for similarity-based prediction. It can be important in methodologies based on similarity, for example, the Case-Based Reasoning (CBR) methodology [1] or the Episode Based Reasoning (EBR) methodology [9]. Some attempt to replacing similarity by consistency has been presented in [3].

4 Conclusions

In this paper we presented an experimental study of consistency-based prediction by means of extensions of information systems. The experiments showed that partially consistent extensions can be used for predicting possibilities of future appearing given states in the examined systems. Moreover, consistency-based prediction seems to be a promising alternative for similarity-based prediction. The important tasks for further research are:

- to consider a more general case, i.e., when processes of examined systems can have new local states which have not been observed yet,
- to propose other ways of the knowledge representation, and what follows, other ways of computing consistency factors,
- to check prediction quality for other similarity measures.

References

1. Aamodt, A., Plaza, E.: Case-based reasoning: foundational issues, methodological variations, and system approaches. Artificial Intelligence Communications 7(1), 39–59 (1994)
2. Moshkov, M.J., Skowron, A., Suraj, Z.: On testing membership to maximal consistent extensions of information systems. In: Greco, S., Hata, Y., Hirano, S., Inuiguchi, M., Miyamoto, S., Nguyen, H.S., Słowiński, R. (eds.) RSCTC 2006. LNCS (LNAI), vol. 4259, pp. 85–90. Springer, Heidelberg (2006)
3. Pancerz, K.: An attempt to use extensions of dynamic information systems in episode-based reasoning. In: Nguyen, H.S., Huynh, V.N. (eds.) Proc. Workshop SCKT 2008, Hanoi, Vietnam, pp. 90–99 (2008)
4. Pawlak, Z.: Rough sets - theoretical aspects of reasoning about data. Kluwer Academic Publishers, Dordrecht (1991)
5. Pawlak, Z.: Concurrent versus sequential the rough sets perspective. Bulletin of the EATCS 48, 178–190 (1992)
6. The ROSETTA Home Page, `http://rosetta.lcb.uu.se/` (accessded May 5, 2009)
7. Rough Set Database System, `http://rsds.univ.rzeszow.pl/` (accessed April 24, 2009)
8. Rząsa, W., Suraj, Z.: A new method for determining of extensions and restrictions of information systems. In: Alpigini, J.J., Peters, J.F., Skowron, A., Zhong, N. (eds.) RSCTC 2002. LNCS (LNAI), vol. 2475, pp. 197–204. Springer, Heidelberg (2002)
9. Sánchez-Marré, M., Cortés, U., Martínez, M., Comas, J., Rodríguez-Roda, I.: An approach for temporal case-based reasoning: Episode-based reasoning. In: Muñoz-Ávila, H., Ricci, F. (eds.) ICCBR 2005. LNCS (LNAI), vol. 3620, pp. 465–476. Springer, Heidelberg (2005)
10. Suraj, Z.: Some remarks on extensions and restrictions of information systems. In: Ziarko, W.P., Yao, Y. (eds.) RSCTC 2000. LNCS (LNAI), vol. 2005, pp. 204–211. Springer, Heidelberg (2001)
11. Suraj, Z., Pancerz, K., Owsiany, G.: On consistent and partially consistent extensions of information systems. In: Ślęzak, D., Wang, G., Szczuka, M.S., Düntsch, I., Yao, Y. (eds.) RSFDGrC 2005. LNCS (LNAI), vol. 3641, pp. 224–233. Springer, Heidelberg (2005)
12. Suraj, Z., Pancerz, K.: The ROSECON system - a computer tool for modelling and analysing of processes. In: Mohammadian, M. (ed.) Proc. Internat. Conference CIMCA 2005, pp. 829–834. IEEE Computer Society, Los Alamitos (2006)
13. Suraj, Z., Pancerz, K.: Towards efficient computing consistent and partially consistent extensions of information systems. Fundamenta Informaticae 79, 553–566 (2007)

Image Annotation Based on Semantic Rules

A.L. Ion

Software Engineering Department, Faculty of Automation, Computers and Electronics,
Craiova, Romania
anca.ion@software.ucv.ro

Abstract. For developing image navigation systems, we need tools to realize the semantic relationship between user and database. In this paper we develop algorithms that automatically generate semantic rules that identify image categories and introduce the cognitive dimension in the retrieval process. The semantic rules are represented in Prolog and can be shared and modified depending on the updates in the respective domain.

1 Introduction

Content-based image retrieval (CBIR) is described as any technology that helps to organize digital picture archives by their visual content. By this definition, anything ranging from an image similarity function to a robust image annotation engine falls under the purview of CBIR [6, 8].

While the effort in solving the fundamental open problem of robust image understanding continued, we also see people from different fields, as computer vision, machine learning, information retrieval, human-computer interaction, database systems, Web and data mining, information theory, statistics, and psychology contributing and becoming part of the CBIR community [5].

One problem with all current approaches is the reliance on visual similarity for judging the semantic similarity, which may be problematic due to the *semantic gap* [1] between low-level content and higher-level concepts [6].

The popular online photo-sharing Flickr [3], which hosts hundreds of millions of pictures with diverse content, the video sharing and distribution forum YouTube have brought in a new revolution for multimedia usage. In [6] it is supposed that image retrieval will enjoy a success story in the coming years. Although great progress has been made in image retrieval and browsing systems since the early, none of the existing methods captures enough of the semantic-related information to be used as a navigation tool in a general content image database. Our inability to capture "image semantics" comes from the incompatibility of the information we are able to compute directly from image data, and our subjective interpretation of the same data [4].

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 83–94.
springerlink.com

The proposed study is based on:

a) the understanding of semantic categories taking into account the human visual perception,
b) the study of human judgment regarding the similitude between images for extracting the significant and discriminate attributes of semantic categories,
c) the design and implementation of algorithms for extracting the image characteristics, similitude metrics, semantic vocabulary, that offers important semantic elements in the retrieval and categorization, semantic rules generation, image classification.

Our study is done on nature categories, where subjects as *water*, *sky/clouds*, *mountains*, *snow*, *rain*, *sun rise* are considered to be important cues. Also the colour composition and colour features played an important role in comparing nature images. Within these categories, spatial organization, spatial frequency, or dimension do influence similarity judgments. Exceptions from these ideas are some categories as flowers, fruits, exotic animals, which contain strong hues (dark, medium red, yellow, blue, green, pink, etc) that are not in the description of nature. This study is an extension of the previous work [7] and started from the limitations regarding the researches in multimedia semantic modelling. In this study, we propose new approaches for image annotations, like: methods for generation of rules which identify image categories, a method for mapping low level features to semantic indicators using the Prolog declarative language, the creation of a representation image vocabulary.

2 The Image Segmentation

The selection of the visual feature set and the image segmentation algorithm is the definitive stage for the semantic annotation process of images. By doing a large set of experiments, we deduce the importance of semantic concepts in establishing the similitude between images. Even if the semantic concepts are not directly related to the visual features (colour, texture, shape, position, dimension, etc.), these attributes capture the information about the semantic meaning.

Before to be segmented, the images are transformed from RGB to HSV colour space and quantized to 166 colours. The extraction of single colour regions is realized by applying the modified colour set back projection algorithm [10].

The implementation of this algorithm is described in pseudo-code:

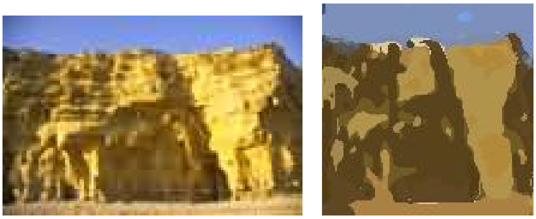
```
Algorithm. Segmentation of an image in regions having a single colour
  Input: image I, colour set C
  Output: the set of single colour regions, R and the set of spatial coherencies of each
          region M.
  Method:
          InitStack(S)
          Visited = ∅
```

```
foreach node P in I do
    if P unvisited and colour(P) is in C then
        PUSH(P)
        R = ∅
        SpatialCoherency = 0
        PointsofRegion = 0
        while not Empty(S) do
            CrtPoint = POP()
          if  CrtPoint is unvisited then
              Add CrtPoint at R
              Visited = Visited ∪ {CrtPoint}
               Connectivity8 = 0
               foreach neighbour N of CrtPoint do
                     if colour(N) = colour(CrtPoint) then
                      Increment Connectivity8
                      PUSH(N)
                     end.
              end.
              SpatialCoherency = SpatialCoherency + Connectivity8
              Increment  PointsofRegion
           end.
        end.
        SpatialCoherency = SpatialCoherency/ PointsofRegion
        Add region at R, SpatialCoherency at M
     end.
   end.
```

Each region is described by the following visual characteristics:

- The colour characteristics are represented in the HSV colour space quantized at 166 colours. A region is represented by a colour index, which is in fact an integer number between 0-165;
- The spatial coherency represents the region descriptor, which measures the spatial compactness of the pixels of same colour;
- A seven-dimension vector (maximum probability, energy, entropy, contrast, cluster shade, cluster prominence, correlation) represents the texture characteristic;
- The region dimension descriptor represents the number of pixels from region;
- The spatial information is represented by the centroid coordinates of the region and by minimum bounded rectangle;
- A two-dimensional vector (eccentricity and compactness) represents the shape feature.

The following figure illustrates the process of segmentation applied on an image belonging to *Cliff* semantic category. As we observe, four significant regions of different hues are detected from the cliff image.



Original Image Colour Regions

Fig. 1. The results of the colour segmentation algorithm on mountain image

3 From Visual Features to Semantic Descriptors

Using the experiments, we construct a vocabulary based on the concepts of semantic indicators, whereas the syntax captures the basic models of human perception about patterns and semantic categories. The representation language is simple, because the syntax and vocabulary are elementary. The language elements are limited to the name of semantic indicators as in Fig. 2. Being visual elements, the semantic indicators and their values are: colour (colour-light-red), spatial coherency (spatial coherency – small, spatial coherency-medium, spatial coherency - big), texture (energy-small, energy-medium, energy-big, etc.), dimension (dimension-small, dimension-medium, dimension-big, etc.), position (vertical-upper, vertical-center, vertical-bottom, horizontal-upper, etc.), shape (eccentricity-small, compactness-small, etc.).

The syntax is represented by the model, which describes the images in the terms of semantic indicators values. The values of each semantic indicator are mapped to a value domain, which corresponds to the mathematical descriptor.

A figure is represented in Prolog by means of the terms of form *figure-(ListofRegions)*, where *ListofRegions* is a list of image regions.

The term region(ListofDescriptors) is used for region representation, where the argument is a list of terms used to specify the semantic indicators. The term used to specify the semantic indicators is of form:

descriptor(DescriptorName, DescriptorValue).

The mapping between the values of low-level (mathematical) descriptors and the values of semantic indicators is based on experiments effectuated on images from different categories and the following facts are used:

mappingDescriptor(Name,SemanticValue,ListValues).

The argument Name is the semantic indicator name, SemanticValue is the value of the semantic indicator, ListValue represents a list of mathematical values and closed intervals, described by the following terms: interval(InferiorLimit, SuperiorLimit).

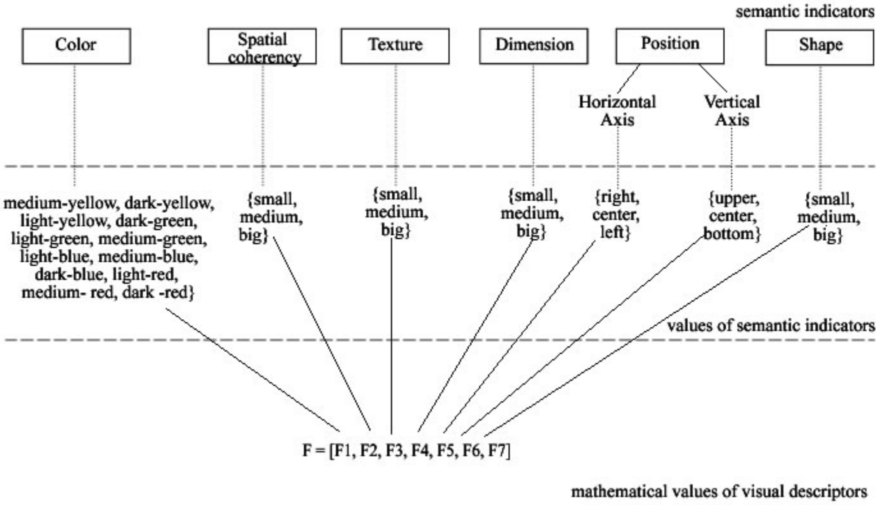

semantic indicators
Color
Spatial coherency
Texture
Dimension
Position
Shape
Horizontal Axis
Vertical Axis
medium-yellow, dark-yellow, light-yellow, dark-green, light-green, medium-green, light-blue, medium-blue, dark-blue, light-red, medium- red, dark -red}
{small, medium, big}
{small, medium, big}
{small, medium, big}
{right, center, left}
{upper, center, bottom}
{small, medium, big}
values of semantic indicators
F = [F1, F2, F3, F4, F5, F6, F7]
mathematical values of visual descriptors

Fig. 2. Vocabulary definition: mapping low-level descriptors values to semantic indicators values

For example, the facts are used to map the semantic indicators red-medium and red-dark to values or intervals of values: the semantic indicator red-medium has the values 144 and 134, and red-dark has the values 100, 123 and all the values from the interval [105,111].

```
mappingDescriptor(colour, red_medium, [144, 134]).
mappingDescriptor(colour, red_dark, [100, interval(105, 111), 123]).
```

The mapping mechanism has the following Prolog representation:

```
mapDescriptor(descriptor(Name,MathematicalValues),descriptor(Name,
SemanticValue)):-
        mappingDescriptor(Name,SemanticValue,ListValues),
        containValue(ListValues, MathematicalValue).

containValue([Value|_], Value).
containValue([interval(InferiorLimit,SuperiorLimit)|_], Value):-
        InferiorLimit=<Value,Value=< SuperiorLimit.
containValue([_|ListValues], Value):- containValue(ListValues, Value).
```

4 Rule-Based Image Annotation

The annotation method includes two phases:

- the learning/training phase in which the rules are generated for each image category, and

- the testing/annotation phase in which new images are annotated using the semantic rules.

We suppose that we have an image database *DB*. In the learning/training phase, the set *U = {S1,..., Sn}* is a subset of *DB* and contains n image-examples, labelled by semantic concepts, for which we train the system and generate semantic rules. In the learning phase, the scope is to automatically generate semantic rules *R* based on categorized image-examples, which identify the semantic concepts of images. A rule determines the set of semantic indicators, which identify the best a semantic concept. In the testing/annotation phase, for each image of the testing subset, namely the images from *DB*, but that are not in *U*, we use the generated semantic rules to label them with one or more concepts.

Since the images and rules are represented in Prolog, its interpreter is used for rules inference to recognize the semantic high-level concepts.

The process of the automated generation of semantic rules and image annotation is the following:

1. The learning phase: rules generation
   A semantic rule is of the form:
   "semantic indicators => semantic concepts"
   The stages of the learning process are:
   - relevant images for a semantic concept are used for learning it;
   - each image is automatically processed and segmented and the primitive visual features are computed;
   - for each image, the primitive visual features are mapped to semantic indicators;
   - the rule generation algorithms are applied to produce rules, which will identify each semantic category from database.
2. The image testing/annotation phase has as aim the automatic annotation of images.
   - each new image is processed and segmented in regions,
   - for each new image, the low-level characteristics are mapped to semantic indicators,
   - the classification algorithm is applied for identifying the image category/semantic concept.

4.1 Generation of Semantic Association Rules

In the described system, the learning of semantic rules is continuously made, because when a categorized image is added in the learning database, the system continues the process of rules generation.

The developed algorithm that generates semantic rules is based on A-priori algorithm [9] for discovering semantic association rules between primitive characteristics extracted from images and categories/semantic concepts, which images belong to. The aim of discovering image association rules is to find semantic relationships between image objects. For using association rules that discover the semantic information from images, the modelling of images in the terms of itemsets and transactions is necessary:

- the image set with the same category represents the transactions,
- the itemsets are the colours of image regions,
- the frequent itemsets represent the itemsets with support bigger or equal than the minimum support. A subset of frequent itemsets is also frequent,
- the itemsets of cardinality between 1 and k are iteratively found (k-length itemsets),
- the frequent itemsets are used for rule generation.

For each frequent colour, all values of the other semantic indicators existed in the images are joined.

The semantic association rules have the body composed by conjunctions of semantic indicators, while the head is the category/semantic concept. A semantic rule describes the frequent characteristics for each category, based on the Apriori rule generation algorithm.

The rules are represented in Prolog as facts of the form:

rule(Category, Score, ListofRegionPatterns).

The patterns from ListofRegionPatterns are terms of the form:

regionPattern(ListofPatternDescriptors).

The patterns from the descriptors list specify the set of possible values for a certain descriptor name. The form of this term is:

descriptorPattern(descriptorName,ValueList).

The values list has the same form as the argument used for mapping the semantic descriptors.

One of the semantic rules used to identify the cliff category is illustrated bellow. This rule has the score (confidence) equal to 100%.

```
rule(cliff,100,
[regionPattern([
        descriptorPattern(colour,[dark-brown]),descriptorPattern(horizontal-
        position,[center]),descriptorPattern(vertical-position,[center]),
        descriptorPattern(dimension,[big]),descriptorPattern(eccentricity-
        shape,[small]),descriptorPattern(texture-
        probability,[medium]),descriptorPattern(texture-inversedifference,[medium]),
        descriptorPattern (texture-entropy,[big]),descriptorPattern (texture-energy,[big]),
        descriptorPattern (texture-contrast,[big]),descriptorPattern (texture-correlation,
        [big])])],
[regionPattern ([
        descriptorPattern(colour,[medium-brown]),descriptorPattern(horizontal-
        position,[center]), descriptorPattern (vertical-position,[bottom]),
        descriptorPattern(dimension,[small]),descriptorPattern(eccentricity-
        shape,[small]), descriptorPattern(texture-
        probability,[big]),descriptorPattern(texture-inversedifference,[big]), descriptor-
        Pattern (texture-entropy,[medium]),
        descriptorPattern (texture-energy,[big]),descriptorPattern (texture-
        contrast,[medium]), descriptorPattern (texture-correlation,[big])])],
    [regionPattern ([
        descriptorPattern(colour,[medium-blue]), descriptorPattern (horizontal-
        position,[center]), descriptorPattern (vertical-position,[upper]),
        descriptorPattern(dimension,[medium]),descriptorPattern(eccentricity-
        shape,[big]), descriptorPattern(texture-probability, [medium]), descriptorPat-
        tern(texture-inversedifference, [big]),
        descriptorPattern (texture-entropy,[big]), descriptorPattern (texture-energy,[big]),
        descriptorPattern (texture-contrast,[small]), descriptorPattern (texture-correlation,
        [big] )])]).
```

4.2 Image Categorization

The set of rules, which was selected after elimination process, represent the classifier. The classifier is used to predict which category the image from test database belongs to. Being given a new image, the classification process searches in the rules set for finding its most appropriate category, as in the bellow figure.

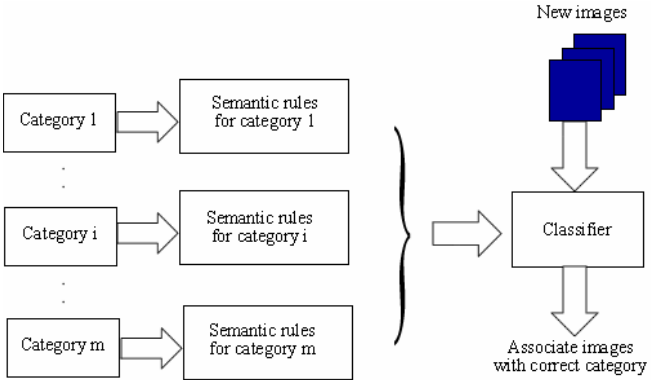

New images
Category 1
Semantic rules for category 1
Category i
Semantic rules for category i
Category m
Semantic rules for category m
Classifier
Associate images with correct category

Fig. 3. The image classification/annotation process

Before classification, the image is automatically processed: the mathematical and semantic descriptors are generated; the semantic descriptors are saved as Prolog facts, and the semantic rules are applied to the facts set, using the Prolog inference engine.

In this study, a new method called the „perfect match classification method” for semantic annotation /classification of images, using semantic rules is proposed and developed. A semantic rule matches an image if all the characteristics, which appear in the body of the rule, also appear in the image characteristics. The mechanism of rules application is called by means of the predicate *isCategory (Figure, Category, Score)*, which has a single clause:

isCategory(figure(ListofRegions),Category,Score):-rule(Category, Score, ListofRegionPatterns), verify(ListofRegions, ListofRegionPatterns).

The predicate verify tests if, for each rule pattern, there is a region, which matches the pattern:

```
verify(_, []).

verify(ListofRegions, [RegionPattern |  ListofRegionPatterns]):-
        containsRegion(ListofRegions,  RegionPattern), verify(ListofRegions,
        ListofRegionPatterns ).

  containsRegion([region(ListofDescriptors)|_,RegionPattern(ListofPatterns
  Descriptors)):- matches(ListofDescriptors, ListofPatternsDescriptors).

  containsRegion([_|Rest], RegionPattern):- containsRegion (Rest, RegionPat-
  tern).
```

The predicate containsRegion searches in the list of regions, which describes a figure, a region, which matches a region pattern.

A region matches a region pattern if for each descriptor pattern of the region pattern there is a region descriptor whose value belongs to the values set of the pattern descriptor.

```
matches(_, []).

matches(ListofDescriptors,[PatternDescriptor|RestPatterns]):-
        matchesPattern (ListofDescriptors,PatternDescriptor),
        matches(ListofDescriptors, RestPatterns).
```

To test a descriptor pattern, the predicate *matchesPattern* is used and it has the following clauses:

```
matchesPattern([descriptor(DescriptorName,DescriptorValue)|_],
        PatternDescriptor(DescriptorName,ListofValues)):-
        containsValue(ListofValues,ValueDescriptor).

matchesPattern([_|Descriptors], PatternDescriptor):-
        matchesPattern(Descriptors, PatternDescriptor).
```

5 Databases with Images for Learning and Annotation

The application of the learning results-semantic rules on other images than the ones used in the learning process is rather difficult. In the experiments realized through this study, two databases are used for learning and testing processes. The database used for learning contains 200 images from different nature categories and is used to learn the correlations between images and semantic concepts. All the images from the database have JPEG format and are of different dimensions. The database used in the learning process is categorized into 50 semantic concepts. The system learns each concept by submitting approximately 20 images per category. The table with image categories used for learning is illustrated below:

Table 1. Image categories associated with keywords

ID	Category	Category keywords
1	*Fire*	fire, night, light
2	*Iceberg*	iceberg, landscape, ocean
3	*Tree*	tree, sky, landscape, nature
4	*Sunset*	sunset, nature, landscape, sun, evening, sky
5	*Cliff*	cliff, rock, see, nature, landscape, sky
6	*Desert*	desert, dry, sand
7	*Red Rose*	rose, flower, love
8	*Elephant*	elephant, animal, jungle
9	*Mountain*	mountain, lake, rock, sky, landscape
10	*See*	see, sky, water, cliff, landscape
11	*Flower*	rose, daisy, clover

The keywords are manually added to each category, for describing the images used for learning process. The descriptions of these images are made from simple concepts like *flower*, *mushrooms* to complex ones like *mountains*, *cliff*, *lake*. In average, 3.5 keywords were used to describe each category. The process of manual annotation of images used for learning semantic rules took about 7 hours.

It is considered that an image was correctly classified by the system, if the category predicted by the computer is correct.

The performance metrics, precision and average normalized modified retrieval rate (ANMRR), are computed to evaluate the efficiency and accuracy of the rules generation and annotation methods.

The precision is defined by the following equation:

$$precision = \frac{CC}{TC} \tag{1}$$

where:
CC – is the number of images correctly classified by system, and
TC – is the total number of classified images.

The precision is in the range of [0, 1] and greater values represent a better retrieval performance.

Averaged Normalized Modified Retrieval Rate (ANMRR) is an overall performance calculated by averaging the result from each query [11]. Normalized Modified Retrieval Rate (NMRR) is used to measure the performance of each query. NMRR is defined by:

$$NMRR(q) = \frac{\sum_{k=1}^{NG(q)} \frac{Rank(k)}{NG(q)} - 0.5 - \frac{NG(q)}{2}}{K(q) + 0.5 - 0.5 * NG(q)} \tag{2}$$

where:
NG(q) is the number of relevant images for image *q*, which will be classified,
Rank(k) is the ranking of the correct images classified by the annotation algorithm, and
K(q) specifies the "relevance rank" for each image, which will be classified.

As the size of the relevant images set is normally unequal, a suitable *K(q)* is determined by: *K*(*q*)= min(4**NG*(*q*), 2**GTM*), where *GTM* is the maximum of *NG(q)* for all queries. The NMRR is in the range of [0, 1] and smaller values represent a better retrieval performance. ANMRR is defined as the average NMRR over a range of queries, which is given by:

$$ANMR = \frac{1}{NQ} \sum_{q=1}^{NQ} NMRR(q) \tag{3}$$

where *NQ* is number of query images.

These parameters are computed as average of each image category (Table 2). The results of experiments are very promising, because they show a small average normalized modified retrieval rate and a good precision for the majority of the database categories, making the system more reliable.

Table 2. Image categories associated with keywords

Category	Precision	ANMRR
Fire	0.77	0.39
Iceberg	0.71	0.34
Tree	0.65	0.45
Sunset	0.89	0.14
Cliff	0.93	0.11
Desert	0.89	0.11
Red Rose	0.75	0.20
Elephant	0.65	0.43
Mountain	0.85	0.16
See	0.91	0.09
Flower	0.77	0.31

6 Conclusions

In this study we propose a technique for semantic image annotation based on visual content. For establishing correlations with semantic categories, we experimented and selected some low-level visual characteristics of images. So, each category is translated in visual computable characteristics and in terms of objects that have the great probability to appear in an image category.

On the other hand, images are represented as a single colour regions list and they are mapped to semantic descriptors.

The annotation procedure starts with the semantic rules generation for each image category. The language used for rules representation is Prolog. The advantages of using Prolog are its flexibility and simplicity in representation of rules and it is not a big time consumer.

Our method has the limitation that it can't learn any semantic concept, due to the fact that the segmentation algorithm is not capable to segment images in real objects. Improvements can be brought using a segmentation method with greater semantic accuracy.

References

1. Smeulders, A., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(12), 1349–1380 (2000)
2. Bezdek, J.: Pattern recognition with fuzzy objective function algorithms. Plenum Press, New York (1981)
3. Flickr, http://www.flickr.com/ (accessed April 25, 2009)
4. Mojsilovic, A., Rogowitz, B.: Semantic metric for image library exploration. IEEE Transactions on Multimedia 6(6), 828–838 (2004)
5. Wang, J., Boujemaa, N., Del Bimbo, A., Geman, D., Hauptmann, A., Tesic, J.: Diversity in multimedia information retrieval research. In: Proc. 8th International Workshop on Multimedia Information Retrieval & ACM International Conference on Multimedia, Santa Barbara, CA (2006), http://lastlaugh.inf.cs.cmu.edu/alex/pubs.html (accessed April 25, 2009)
6. Datta, R., Lia, J., Wang, J.: Image retrieval: ideas, influences, and trends of the new age. In: ACM Computing Surveys (2008), http://infolab.stanford.edu/~wangz/project/imsearch/review/JOUR/datta.pdf (accessed April 25, 2009)
7. Ion, A., Stanescu, L., Burdescu, D., Udristoiu, S.: Improving an image retrieval system by integrating semantic features. In: Proc. IEEE Conference on Human System Interaction, Cracow, Poland, pp. 504–509 (2008)
8. Datta, R., Joshi, D., Li, J., Wang, J.: Tagging over time: real-world image annotation by lightweight meta-learning. In: Proc. ACM Multimedia, Augsburg, Germany, pp. 393–402 (2007)
9. Frawley, W., Piatetsky-Shapiro, G., Matheus, C.: Knowledge discovery in databases: an overview (1992), http://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1992.pdf (accessed April 25, 2009)
10. Smith, J.R.: Integrated spatial and feature image systems: retrieval compression and analysis. PhD Thesis, Graduate School of Arts and Sciences, Columbia University, New York, NY (1997)
11. Manjunath, B.S., Salembier, P., Sikora, T.: Introduction to MPEG-7: multimedia content description standard. Wiley & Sons, New York (2001)

Part II
Disabled Persons Helping and Medical H-CSI Applications

A Prototype Mobility and Navigation Tele-Assistance System for Visually Disabled

M. Bujacz, P. Barański, M. Morański, P. Strumiłło, and A. Materka

Institute of Electronics, Łódź University of Technology, Łódź, Poland
{bujaczm,baranski,moranski,pawel.strumillo}@p.lodz.pl,
andrzej.materka@p.lodz.pl

Abstract. The paper presents initial research on the system for remote guidance of the blind. The concept is based on the idea that a blind pedestrian can be aided by spoken instructions from an operator who receives a video stream from a camera and GPS readouts from a sensor carried by a visually impaired user. An early prototype utilizing two laptop PCs and a wireless internet connection is used in indoor orientation and mobility trials, which aim to measure the potential usefulness of the system and discover possible problems with user-operator communication or device design. A second prototype is also constructed using a smaller subnotebook Flybook computer, a GPS sensor and GSM modem with HSDPA technology, and tested in outdoor environments. Test results show a quantitative performance increase when travelling with a remote guide: 15-50% speed increase and halved times of navigational tasks in indoor trials, as well as a significant decrease in the number of missteps and accidental obstacle collisions in outdoor trials. A large part of the success is the supportive feedback from the blind testers, who point out the engendered feeling of safety when assisted.

1 Introduction

Vision loss is a serious impairment that deprives a human of approximately 80-90% perceptual abilities and has a detrimental effect on professional, social and personal quality of life. The European Union surveys report that 4 out of every 1000 Europeans are blind or suffer from a serious visual disability. This number increases yearly due to the aging demographic.

In earlier studies [1], the authors have conducted questionnaires among the blind asking to indicate the main barriers in their everyday activities. The three following ones were pointed out (ranked according to their importance):

1. Lack of the means for safe and independent mobility (e.g., obstacle detection and early warning),
2. Limited capability to navigate in the surrounding environment (route planning and following, identification of whereabouts),

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 97–108.
springerlink.com

3. Difficulties in access to visual or text information (e.g. road signs, bus numbers) and handling of devices (e.g. cellular phones, vending machines).

Overcoming the first two barriers can only partially be achieved by means of the so called primary aids, i.e. the white cane or a guide dog. Much research effort has also been devoted to development of electronic travel aids (ETA) that implement different versions of sensory substitution concepts to compensate for the lost vision [2]. Unfortunately, these devices have not found any widespread acceptance among the blind. The causes include high costs, poor reliability and lack of comfort in using such devices. On the other hand, it is clear that no single device can match the comprehensive help offered by a sighted guide. Obviously, for different reasons, this solution is not always possible or accepted by a blind person and often deepens the feeling of being a social burden.

A compromise concept would be a system that is capable of delivering the assistance of a sighted guide remotely, for example by using existing GSM technologies by building a so called tele-assistance link. The current and upcoming ICTs (Information and Telecommunications Technologies) offer platforms for implementing such systems.

1.1 Existing Tele-Assitance Systems

To the authors' best knowledge, the first reported system for remote guidance of the blind was the system developed at the Brunel University, UK [3]. Three ICT technologies were combined to offer the tele-assistance functionality; namely, GPS (Global Positioning System), GIS (Geographic Information System) and video/voice transmission over the 3G mobile network. The system [3] comprises of two units: the mobile unit that is carried by the blind user, and a stationary PC-based unit. The mobile unit is equipped with a portable camera and an audio headset. The stationary unit, operated by the sighted guide, runs an application displaying what the camera carried on the blind person's chest sees [4]. Video, voice and other data is transmitted over the 3G communication network. The authors have noted successful remote navigated walks of the blind individuals within the campus precinct [3]. However, no commercial deployment of the system was reported.

Although few recent scientific publications in this narrow field can be found, a number of companies are working on very similar commercial projects. One has been announced by a French company Ives Interactivity Video Systemes [5]. The device called Visio.assistant is now in the phase of preindustrial tests using WiFi. In that solution the webcamera and the wireless transmitter are mounted in a case resembling a hand hair dryer.

Another system worthy of mentioning is ***Micro Look***, awarded two medals on the 2007 Brussels Innova Fair. This system is under development by a Polish company Design-Innovation-Integration [6]. ***Micro Look*** differs from the earlier designs by integrating a webcamera with a headset and a mobile telephone platform. The project is at the stage of a prototype under tests.

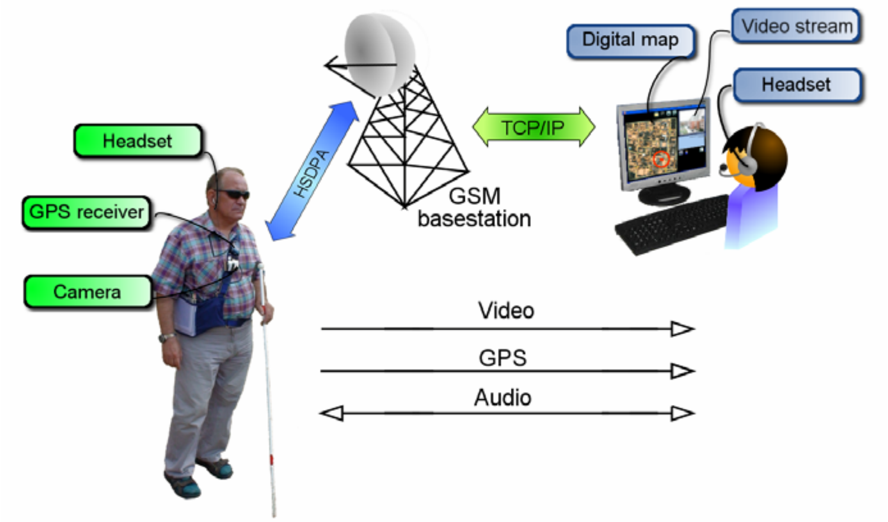


Fig. 1. The blind user of the tele-assistance system can contact a remote operator to request help in mobility and navigation

1.2 The Electronic Travel Aid

For a number of years, the author's group in the Medical Electronics Division of the Łódź University of Technology has been working on a project aimed at development of a prototype ETA system for the blind. The system utilizes stereoscopic cameras for 3D scene reconstruction [1, 7]. Each obstacle that is segmented out from the scene is associated with a unique sound code that warns the blind. Headphones are used for playing the sounds. A special sound lateralization technique implementing the Head Related Transfer Functions (HRTF) is used for "auditory 3D display" of the obstacles [8]. The system module that is currently under tests provides micro-navigation functionality. It is supported by a Symbian OS based smartphone playing the role of the auditory assistant of the blind. Ordinary phone functions are speech synthesized and new software procedures are provided, e.g. internet browser (via RSS feeds), voice-recording, and recognition of colours [9].

Our current goal is to couple the earlier developed system modules i.e. micro-navigation with the macro-navigation functionality, which will be offered by the remote assistant, GPS and digital maps of the urban terrain. In this work we report initial studies of the remote guidance system that underwent preliminarily indoor and outdoor testing.

2 Experimental Setup and Procedures

2.1 The Prototypes

A decision was made to make the first prototypes functional and easily modifiable, without limiting the system to a restrictive platform or a transmission protocol.

The easiest way was to do that was to operate on a PC platform and use existing wireless and GSM technology for development and trial purposes.

The prototype of the guidance system is composed of two notebook computers - one carried by a blind person in a specially constructed, ventilated backpack, the other operated by the guide.

An earphone and microphone headset, and a USB-camera are connected to the notebook PC carried by the blind user. The earphones used are of the "open-air" type in order not to block environmental sounds. The camera is mounted on the chest of the blind person. The initial considered mounting location was on the visually impaired person's head; however, the first trials showed that it was an inconvenient solution both for the operator and the blind user. First, natural movements of the guided person's head provided the operator with erroneous information about the walking direction. Second, instructions from the operator interfered with the way the blind user wanted to move his head in order to hear the sounds of the environment. Despite the usefulness of the operator being able to remotely "look around", due to the aforementioned inconveniences the chest-camera was opted for.

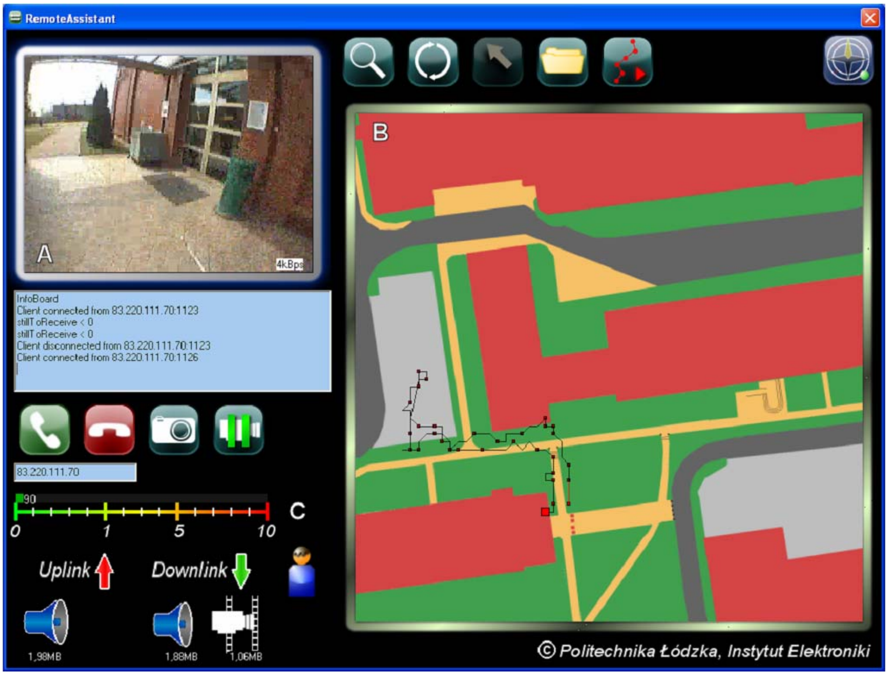
Soon after the first indoor trial the prototype was modified in a number of ways. First of all the link is no longer limited to wireless LAN, but also through a GSM modem using HSDPA technology. The blind users now carry a small Flybook computer instead of a normal laptop. The camera was equipped with a wide-angle fish-eye lens and automatic gain and exposure control. Additionally the prototype was equipped with a GPS receiver, and its readouts are viewed on a digital map by the remote operator.

In the first trials the freeware Skype software had been used, but a dedicated program for audio and video streaming was created soon after. The operator's program is shown in Figure 2. The operator can manually crop the video bandwidth, take high-quality snapshots in JPG format, plot the user's movements on the map.

The software allows both the operator and the user to establish a link through TCP/IP packets that guarantee information delivery. Voice and video samples make their way to the destination by UDP packets that are devoid of any additional wrap-up data (compared to TCP/IP packets). Special tracked packets serve to calculate delays and trigger pauses in transmission if the network is too congested. The status of the connection is accompanied by appropriate auditory icons, i.e. dialing, connection established, weak connection, broken connection.

2.2 Trial Rationale and Goals

Having prepared the early prototype, our goals were to test what problems may arise in further development of the system, study the limitations of user-operator communication, as well as to develop procedures for quantitative review of the system. Among the analyzed aspects were the influence of video quality and delay on the operator's efficiency, what information is necessary to be passed on between the user and the operator, and how does guided and unguided travel and orientation compare.



InfoBoard
Client connected from 83.220.111.70:1123
stillT oReceive < 0
stillT oReceive < 0
Client disconnected from 83.220.111.70:1123
Client connected from 83.220.111.70:1126

83.220.111.70

Fig. 2. Screenshot of the operator's application: A) The video stream, B) Digital map with GPS readouts, with small markers every 5 seconds, C) Connection controls and parameters

2.3 Indoor Trial Description

Three blind volunteers took part in the indoor trials – all male, ages 25-45, two blind since birth, and one for the last 17 years. Two had temporarily successful retina transplantations, which unfortunately regressed leaving them with partial light sensitivity (they were thus blindfolded for the tests). Each of the participants was asked to travel three multi-segment paths and locate a doorway in the corridor.

The multi-segment paths were designated in a 7 by 12 meter hallway. Four cardboard obstacles were deployed at random along the paths and were moved around between each travel attempt, to prevent the participants from memorizing obstacle locations and using them for orientation. One of the experimenters accompanied the blind volunteers at all times, should a surprising or dangerous situation arise. Every blind volunteer used his long cane during the tests.

Each path was completed in four different "modes":

a) *unguided walk* – the participants were told path information only, for example: 10 steps ahead, then turn right and make 6 steps, then make 5 steps and so on. Prior to the tests the number of steps was adjusted to each participant's stride length. In case a participant forgot the route the experimenter reminded him, but nothing else. This run served as a reference, simulating a path known from memory, but with possible unexpected obstacles along the way.

b) *remote guidance without path information* – the blind participant wore the headset, the chest-camera and the backpack with the laptop. The operator provided remote guidance on the base of a real-time video transmission. However, the operator did not provide precise information about the path (number of steps) and only estimated the participant's position from the video feed.
c) *remote guidance with path information* – as in the previous mode, the blind volunteers were guided by the remote operator. This time however, the operator provided them with precise information concerning the path (i.e. the exact number of steps to take).
d) *walk with a human guide* – in this reference run a visually disabled person held a guide's hand who led their way along the precisely defined path. The speed of walking was dictated by the blind participant.

At the end of each trial run the blind participant was asked to point to where he thought the start of the path was located. This was to test whether focusing on the remote guidance influenced the person's orientation skills.

The door-finding task was a more realistic test. The participants were requested to find the sixth doorway on the right in a university corridor. Cardboard obstacles along the route made the task more difficult. Simple as it may appear, all of the participants failed to find the target door with their first attempt due to losing count. This test was subdivided into three categories: unguided, remote guided, and a reference guided run.

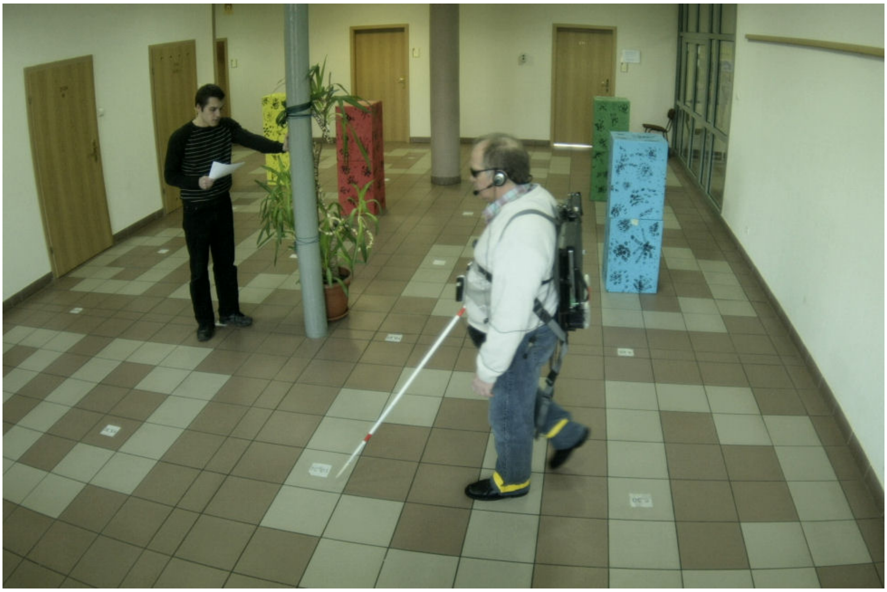


Fig. 3. Navigational trials were automatically recorded. Markings on the participants' pant legs and on the floor allowed precise path reconstruction.

2.4 Data Collection

A wide angle camera automatically took pictures of the trials every 2 seconds. A sample photo is presented in Figure 3. Bright markers on the participants' pant legs allowed quick, half-automated position plotting. The results were interpolated in order to graph the subject's positions with a resolution of one second. Using this technique it was possible to accurately record, time-stamp and measure all the traversed trial paths. The perceived starting point direction was also marked for each trial.

At the end of the experiment, each participant completed a short survey providing feedback about the proposed system, his communication with the operator, and expectations from the developed device.

Table 1. Results from one of the trial paths and the door locating task

	Path wavering* [m]			Path/Task completion time [s]			Error in locating the starting point		
Participant	MM	RP	JM	MM	RP	JM	MM	RP	JM
Path 1 (reference, help of a human guide)	-	-	-	55	38	30	45°	90°	Unable to guess
Path 1 (unguided, path information only)	4,0	9,4	3,9	92	82	82	100°	135°	45°
Path 1 (remote guidance, no path information)	10,0	8,1	8,6	69	55	60	55°	45°	55°
Path 1 (remote guidance with path information)	5,1	9,8	7,5	63	60	58	45°	45°	45°
Door locating task (reference)	-	-	-	30	36	30	-	-	-
Door locating task (unguided)	-	-	-	60	66	94	-	-	-
Door locating task (remote guidance)	-	-	-	34	36	38	-	-	-

*sum of distances from 5 key path points

2.5 Outdoor Trial Description

The second trial, performed with the more portable prototype with an UMTS-/HSDPA modem, consisted of navigating outdoor paths on the university campus. The same three visually impaired volunteers participated in the tests. Trial paths were chosen to provide multiple obstacles, such as stairs, trash cans, park benches, lamp posts and fire hydrants. Every blind volunteer used his long cane during the tests. The paths had only limited turns to be easily remembered. One of the experimenters accompanied the blind volunteers at all times for safety purposes, but provided assistance only in dangerous situations.

In order to prevent path familiarization to influence the results, the three trial participants completed the paths in three different "modes" in the order: a), a), b), c), b), c), b), c), b), c), where:

a) *walk with a human guide* – walking hand-in-hand with a guide, who explained the turns in the path and any dangers along it. All three participants claimed that one run was sufficient to memorize the simple paths, but the process was repeated twice.
b) *remote guidance* – the blind participant wore the earpiece, chest-camera and the portable mini-laptop. The operator provided remote guidance on the base of the real-time video transmission and observed the volunteer's positions thanks to GPS readouts displayed on his digital map.
c) *independent walk* – the participants relied only on themselves to navigate the paths, they were warned when in danger or when they lost their way.

Thanks to such distribution of modes, the familiarization with the path increased steadily and evenly influenced both the remote assisted and independent attempts.

Table 2. Outdoor trial times

		Path I times			**Path II times**		
		JM	MM	RP	JM	MM	RP
Assisted	1	03:27	03:58	04:06	03:21	06:00	04:00
Independent	2	03:31	04:00	03:50	04:02	06:41	04:40
Assisted	3	02:57	03:44	03:18	03:06	05:09	03:50
Independent	4	03:06	03:20	03:53	03:50	05:15	04:09
Assisted	5	02:58	03:06	02:55	03:58	03:58	03:37
Independent	6	03:05	03:30	03:10	03:17	04:41	03:51
Assisted	7	02:55	03:08	02:55	03:40	03:20	03:26
Independent	8	03:00	03:08	03:03	03:20	04:32	03:29

During each path attempt the participants were timed, and the number of missteps, minor collisions, potentially dangerous collisions and path confusions were counted. Missteps were either slight stumbles, or stepping on the grass. Minor collisions were counted when unexpected obstacles were hit strongly with a cane, or a volunteer brushed against an obstacle. Dangerous collisions were always prevented by the assisting experimenter, but included such dangers as tripping on stairs, or walking into an obstacle missed with the cane.

The GPS position readouts were recorded once every second. The experiments were recorded and later replayed so that the participants' positions at 5 second intervals could be plotted. The collected data is presented in Table 2 and the total results, summed for the three participants are shown in Figure 4.

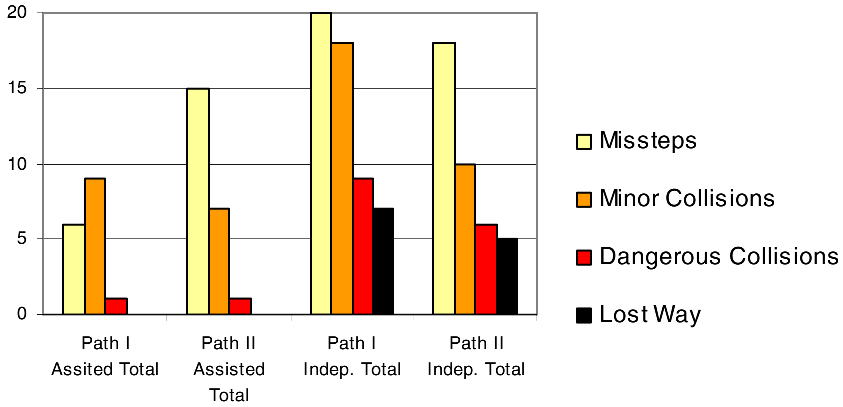


Fig. 4. Comparison of the total number of unfortunate event occurrences for assisted vs. independent trials totalled for all participants and all attempts on two outdoor paths.

3 Trial Results and Analysis

3.1 Performance Review

Comparing the performance between the three participants is difficult as the cause of their blindness and mobility skills are disparate. The results provided in Table 1 are the most representative of the group. Due to space constrains the table shows data for only one of the three test paths.

The detailed trial data shows that the help of a remote guide is a significant improvement over travelling unguided. It is still far from assistance of a human guide present onsite, but the average travel speeds were 20-50% better then when traversing the paths unguided. In outdoor trials the time differences were less significant – 15-20%. This is primarily due to the real-world environment in which the trial participants were experienced in navigating. The trials were constructed in such a way as to quickly familiarize the blind volunteers with the paths, leading them to reach quickly reach a similar safe top speed.

The large path wavering when under remote guidance in indoor trials showed that the operator was not always able to precisely estimate the position of the assisted person. This is improved once he could provide a-priori instructions about the distances in the paths.

Contrary to their expectations, the blind participants retained better orientation when using the remote guide, than when walking unguided or with a human guide, as evidence by better estimation of the starting point at the end of each path.

The task of locating the correct door was failed by all three participants on the first attempt. The successful unguided attempt was on average two times slower than when assisted by a remote guide.

In the outdoor trials, the main improvement was the certainty of navigation and the decrease in missteps, such as walking onto grass or tripping on curbs, and accidental collisions.

3.2 Operator Conclusions

One of the main conclusions about the user-operator communication was that that it is far more effective to inform the guided person about a nearby obstacle and let him pass the obstacle alone than trying to manoeuvre him precisely, e.g. with “step left/right” commands. The operator simply instructs: “obstacle on the left, a doorway on half right”. It requires less effort from both the operator and the blind pedestrian, is more concise and yields better results. After all, the ultimate goal of the system is to complement and not replace the way the blind travel.

The operator’s adjustment to each blind user’s needs seemed to be a key parameter in providing effective navigation. When encountering obstacles, short and simple commands e.g. “park bench on your left, a doorway on half right” were far more welcome to most users than precise manoeuvring instructions. To ensure the blind persons maintain a proper walking direction it was advisable to find a travel boundary, e.g. „a walk along the left curb’’. From the psychological point of view operator-user communication should be settled individually with every blind user. Some participants only wanted short navigating instructions and obstacle warnings; others felt more comfortable and safe after a more detailed description of the environment.

3.3 Survey Results

Both trials were followed by short questionnaires, which revealed some important points for improving the system. All the respondents expressed great excitement about the project. They professed that if the device comes into being, they would use it on daily basis to travel with more confidence and explore new areas. They conceded that the presence of the operator’s voice engendered a feeling of safety, which the ETA [1] all of them had contact with previously could not compare to.

The surveyed participants expressed a general reluctance to hold any extra items such as an electronic compass or a camera, and under no circumstance would they relinquish their canes.

Their three primary predicaments that need to be addressed are as follows:

a) finding a set of traffic lights to safely cross a pedestrian-crossing,
b) finding the button to activate a green light on a pedestrian-crossing, and
c) searching for a quiet street and being guided across it. Traversing a pedestrian-crossing with no traffic lights is the most risky task.

They can be solved by the developed system.

4 Present and Future Work

4.1 Enhancing GPS Navigation

The inclusion of a GPS receiver greatly enhanced the system’s functionality. Availability of precise digital maps enables the visualization of sidewalks, lawns, the outlines of buildings, stairs, traffic lights, zebra crossings, bus and tram stops which are very helpful landmarks for the blind. As the accuracy of commonly used GPS receivers falls short of the precision required to track the exact position of a pedestrian, an electronic compass and inertia sensors are being added to the system.

4.2 Adjustable Video Quality

The tests revealed the necessity to enhance the quality of the video stream to the operator, or if it fails, change the parameters of the video transmission on demand, prioritizing either smooth frame-rate or high resolution. The trial participants expressed the wish to be able to read signs, numbers and names in halls and corridors, the numbers of approaching trams or buses. This can be accomplished by capturing a high resolution video sequence where the frame rate and latency are of secondary importance. On the other hand, crossing streets or moving on busy pavements necessitates a high frame rate and good response that compromises the resolution. These parameters should be regulated by the remote operator according to the needs of the blind user.

4.3 Target Platform

The target platform has yet to be decided upon, but it must be compact and light enough for it to be concealed into a small handbag or waist-pouch. There are several devices considered, all with their pros and cons. Mobile phones have the advantage of being a ready made, and relatively cheap technology; however they lack in computational power and video quality. Small palm-top or ultraportable laptop computers are an expensive solution, but one that meets all the demands and that is currently used in the prototype. The last considered technology is powerful microcontrollers, which would require the design of a custom circuit board and casing.

Interviews with the blind volunteers showed that the size and appearance of any electronic travel aid is of great importance to them. Our target device must be small enough to fit in a pocket or a light shoulder bag. It cannot be overly exposed, as almost half of the interviewed blind users have experienced cell-phone or mp3-player theft at some point of their lives.

5 Conclusions

The trial results and the questionnaire following the tests show that the provision of this “electronic vision”, albeit in its infancy, could elevate both the physical and psychological comfort of the blind. The operator’s voice and “telepresence” causes a strong feeling of safety. Should an unexpected event occur, a blind user of the system can always count on immediate assistance. The authors feel strongly encouraged by the test results and the positive feedback from the trial participants.

Objective test results show that a blind person assisted by a remote guide walks faster, at a steadier pace and is able to more easily navigate inside a building. The trials provided valuable experience to the designers concerning future requirements from the target platform and the communication channel.

Another concept worth considering is that the operators themselves could be disabled individuals. Confined to a wheelchair, they could find employment in aiding the blind.

Acknowledgments

The authors would like to express their deep gratitude to the volunteers from the Łódź Center of the Polish Society of the Blind, for their steadfast cooperation, unabated patience and invaluable feedback in areas for improving the system.

This work has been supported by the Polish Ministry of Science and Higher Education grant no. R02 01303 in years 2007–2010.

References

1. Strumiłło, P., Pełczyński, P., Bujacz, M., Pec, M.: Space perception by means of acoustic images: an electronic travel aid for the blind. In: Acoustics High Tatras 33rd Intern. Acoustical Conference - EAA Symposium, Strbske Pleso, Slovakia, pp. 296–299 (2006)
2. Bourbakis, N.: Sensing surrounding 3-D space for navigation of the blind. In: IEEE Engineering in Medicine and Biology Magazine, pp. 49–55 (Janaury-Feburary 2008)
3. Garaj, V., Jirawimut, R., Ptasiński, P., Cecelja, P., Balachandran, W.: A system for remote sighted guidance of visually impaired pedestrians. British J. Visual Impairment 21, 55–63 (2003)
4. Hunaiti, Z., Garaj, V., Balachandran, W., Cecelja, F.: Use of remote vision in navigation of visually impaired pedestrians. International Congress Series 1282, 1026–1030 (2005)
5. Web page, `http://www.ives.fr/` (accessed April 3, 2009)
6. Web page, `http://www.warsawvoice.pl/view/17989` (accessed April 3, 2009)
7. Web page, `http://www.naviton.pl` (accessed April 3, 2009)
8. Pec, M., Bujacz, M., Strumiłło, P.: Personalized head related transfer function measurement and verification through sound localization resolution. In: Proc. 15th European Signal Processing Conference (EUSIPCO 2007), Poznań, Poland, pp. 2326–2330 (2007)
9. Strumiłło, P., Skulimowski, P., Polańczyk, M.: Programming Symbian smartphones for the blind and visually impaired. In: Łódź (ed.) Intern. Conference on Computers in Medical Activity. LNCS. Springer, Heidelberg (2007) (to be published)

Eye-Mouse for Disabled

T. Kocejko, A. Bujnowski, and J. Wtorek

Department of Biomedical Engineering, Faculty of Electronics,
Telecommunication and Informatics, Gdańsk University of Technology, Gdańsk, Poland
kocejko@gmail.com, {bujnows,jaolel}@biomed.eti.pg.gda.pl

Abstract. This paper describes real-time daylight based eye gaze tracking system. Proposed solution consists of two video cameras, infrared markers and few electronic components. Usage of popular webcams makes the system inexpensive. In dual camera system one camera is used for pupil tracking while second one controls position of the head relative to the screen. Two detection algorithms have been developed and implemented – pupil detection and the screen position detection. Proposed solution is an attempt to create an interface for handicapped users which they could use instead of mouse or keyboard.

1 Introduction

A typical human-computer interface transfers information between a user and a computer. Most commonly used devices for presenting information are screen displays and loudspeakers. The latter one may also be used for audio feedback creation. Traditionally, a keyboard and a mouse are used for communicating with the computer. So called input devices, mentioned above, are useless in case of paralyzed persons. Fortunately, a great amount of persons with motor dysfunctions still may express vocally their feelings and demands. It is extremely important for the comfort of their life. Unfortunately, there are also people vocally and physically disabled which makes the communication with them very difficult. For example, in case of patients with amyotrophic lateral sclerosis (ALS) progressive paralysis of the voluntary muscles develops due to loss of motor neurons. ALS is a severe, progressive disease affecting both the central and peripheral parts of the motor nervous system. In the final stage people cannot control their limbs and facial muscles. Furthermore, this disease may also involve disphonia or aphonia. The computer can be considered as an ideal tool for enabling communication with such people, improving it and making more reliable. Some available and easy measured signals coming out of the patient's body can be utilized to navigate the computer or to express the disabled person itself. There are some possibilities to acquire information starting from electrical signal generated by brain (EEG), muscles (EMG) or from detection of partially active limb micro-movements. A research has been undertaken to use the EEG signal as a control of PC operation or other equipment (like electric wheelchair) which can be combined with PC [1]. There are also some projects involving multiple technologies like EMG and EGT (eye gaze tracking) [2]. Although it is very interesting to operate the machine or computer with thoughts only, this kind of system demands very stable electromagnetic conditions and is very

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 109–122.
springerlink.com

sensitive to noise level and to other interferences. Detecting the eye placement can be use to correct artifacts in electroencephalogram data resulting from eye movement. In this case data coming from the eye tracking device cannot be corrupted by any electrophysiological signal [3].

Detecting the eye position is also used as a factor which allows determination of drivers fatigue. Detected eye is being analyzed (is it open or is it closed) and on this basis drivers fatigue is being defined [4]. Basing on frequency of eyelid blinks the users awareness can be determined [5]. Different methods of eye-gaze and eye-movement tracking have been reported in the literature. A short list includes Electro-Oculography [6-8], Limbus, Pupil and Eye/Eyelid Tracking [9-14] Contact Lens Method, Corneal and Pupil Reflection Relationship [8, 11, 15, 16], Head Movement Measurement [17, 18].

In our research we have focused on the gaze detection as an non-invasive tool for computer navigation. This paper presents preliminary studies on eye pupil position detection and possibility to use this information in computer navigation. This paper is organized as follows. The proposed eye movement tracking algorithm and screen detection algorithm are presented in the Section 2. The Section 3 contains experimental results. Discussion is presented in the Section 4. The Section 5 includes conclusions.

2 Methods

The proposed method relies on simultaneous positioning an eye pupil and a computer screen. The method is composed of four main algorithms: 1.the pupil detection algorithm, 2. the screen detection algorithm, 3. the algorithm for estimation of fixation point, and 4. the algorithm for relating the fixation point with the mouse cursor position. Before the algorithms are used the captured images have to be pre-processed and de-noised. The following operations on images are done: 1. conversion to HSV scale, 2. Gaussian smoothing, 3. morphological (erosion, dilation, and fixing of the threshold). The algorithms are described in the following paragraphs. Description of the above mentioned operations are omitted as they description can be easily found in the literature, e.g. [19].

Algorithm of Eye Pupil Detection

Two different algorithms for eye pupil detection have been proposed: detection of the longest segment algorithm (LSD) and its modified version (MLSD). It is assumed that the searched pupil is represented by black pixels in the captured and processed images. LSD algorithm is based on scanning the resulting, after pre-processing, binary image row by row. The length, starting and ending coordinates are saved for each detected black segment (even one pixel). Then the longest segment is selected and its centre is estimated as the pupil centre.

The horizontal and vertical segments are used by MLSD algorithms. Thus, image is also scanned column by column. The segments parameters are stored in separate buffers. Then, the longest segment is determined and segments shorter than 0,9 and longer than 0,5 of the length of the selected longest one are chosen. In the next step the collection of the beginning and ending points of selected segments is created. To estimate pupil shape the ellipse fitting algorithm is applied using the collected points. The pupil centre is determined by the ellipse centre.

The Screen Detection Algorithm

The algorithm of the screen detection is relatively complicated due to the fact that the captured image contains different information depending on the distance between camera and the screen. In fact, the screen detection algorithm is reduced to detecting position of infra-red (IR) markers. Such markers are attached to each corner of the screen frame. Again LSD algorithm is applied to the pre-processed image, however it is captured by the screen camera. At the beginning the image is divided arbitrarily into four parts, however each part contains one of the screen corner. Then LSD algorithm detects IR markers separately for each part of the image. Finally, an adaptive screen dividing is applied when all four corners are already detected.

The Calibration Procedure

The calibration procedure is used for determination of the fixation point (point on which the vision is fixed). During calibration, the user is asked to look at a number of screen points for which the positions (in the scene image) are known. Three different calibration patterns has been used containing respectively 2, 3 and 9 points.

In case of 2-point calibration the fixation points are estimated in relation to points laying at the beginning and at the end of the screen diagonal. The fixation point is estimated in relation to points corresponding to screen width and height in three points calibration procedure. In case of 2-points calibration the following (1, an 2) equations are used to calculate the fixation point:

$$x_s = f(x_e, y_e) = x_{s1} + \frac{(x_e - x_{e1}) * (x_{s2} - x_{s1})}{(x_{e2} - x_{e1})} \tag{1}$$

$$y_s = f(x_e, y_e) = y_{s1} + \frac{(y_e - y_{e1}) * (y_{s2} - y_{s1})}{(y_{e2} - y_{e1})} \tag{2}$$

Here,

x_s, y_s – are coordinates of the fixation point,

x_e, y_e – current coordinates of the pupil center,

x_{si} y_{si} – coordinates of the point in the captured image corresponding to calibration points, and

x_{ei}, y_{ei} – coordinates of the eye pupil center, related to x_{si} and y_{si}

The 3-points calibration procedure is a compilation of two 2-points calibration made twice (vertical and horizontal arrangements) and uses the same equations. The vertical coordinate is calculated according to calibration points corresponded to screen height. The horizontal coordinate is calculated according to calibration points corresponded to the screen width. In case of the 9-points calibration the fixation point coordinates are approximated using the following (3,4) formulas:

$$x_s = f(x_e, y_e) = a_1 + a_2 x_e + a_3 y_e + a_4 x_e y_e + a_5 x_e^2 + a_6 y_e^2 \tag{3}$$

$$y_s = g(x_e, y_e) = b_1 + b_2 x_e + b_3 y_e + b_4 x_e y_e + b_5 x_e^2 + b_6 y_e^2 \tag{4}$$

Here,
x_s, y_s – are fixation point coordinates,
x_e, y_e – current coordinates of eye pupil center, and
a_i, b_i – are constants determined by the calibration procedure.

Relating the Mouse Cursor Position with Fixation Point
The mapping procedure allows relation mouse cursor position to estimated fixation point. The applied approach relies on trigonometric relationships between coordinates of actual and transferred points. Before mapping procedure can be applied, coordinates of a virtual screen (the screen presented in the captured image) are normalized. Firstly, direction and angle of the virtual screen are calculated. Next, every marked point is rotated by the angle of opposite value than the detected one using appropriate formulas. At last, the coordinates are normalized. After normalization procedure, mouse cursor position can be related to position of the fixation point by means of polynomial mapping.

Experimental Stand
Experimental stand have been performed using a developed experimental stand (Fig. 1). It consisted of two cameras, four infra-red (IR) electro-luminescent diodes (LED) used to mark the screen corners and personal computer. Both cameras were attached to the head of an examined person by means of the glasses-frame. One camera was devoted to observing the eye while the second one was observing the scene in front of the person including computer monitor. The camera that observed the computer screen contained two filters: one protecting from the IR light and the second one to attenuate a day light. The former filter was intended to protect from reflected infra red light. However, the infra red light directed straight to the camera lens was not attenuated completely. The aim of the latter filter was to remove the day light and the light reflected form the screen. As a result only IR LED's were visible. The "scene camera" had resolution equal or above 640x480 dpi (e.g. Creative Labs Webcam Live) while the "eye camera" was a typical low-resolution webcam (e.g. Intuix PC webcam). Both could be easily connected to the PC via USB interface.

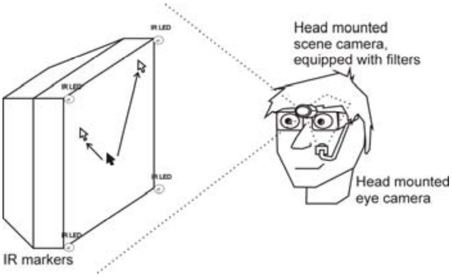


Fig. 1. Hardware setup

Experiments
Experiments were performed on a group of 10 people of different age, sex, and a computer software knowledge. The youngest checked person was 22 years old, the oldest was 79. Various experiments have been performed:

Test 1: The pupil detection algorithms have been compared and tested for correctness of the pupil shape and centre estimation. Firstly, the user was asked to look straight forward and to deflect its eye left and right.

Test 2: MLSD algorithm has been tested for the range of pupil deflection. User was asked to deflect his eye maximally left, right, up and down without changing his head position.

Test 3: The influence of the calibration procedure type on the fixation point estimation was tested. Test executed were to show how many calibration points were needed to keep the balance between "user friendly" calibration procedure and the device correctness. User was asked to look at 9 different points (point by point) of known positions. Before every try user had to go through a calibration procedure. Two, three and nine point calibration patterns were tested. The positions of estimated fixation point and corresponding check point were compared. The test was performed for tree different "user to screen" distances: 40cm, 60cm and 80cm.

Test 4: The influence of "user to screen" distance changes have been tested. The calibration was used for "the user to screen" distance equal to 40cm. User had to gaze on 9 points (point by point) and then the distance from the screen to the user was set 60 and 80cm.

Test 5: The device has been tested for the acceptable head rotation. The user was asked to perform the "test 3" with his head rotated by 30 degrees and "user to screen" distance of 80cm.

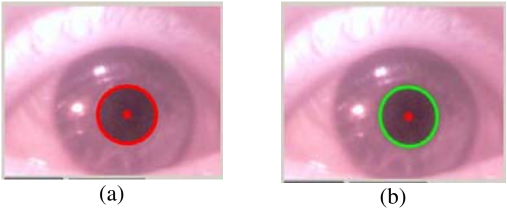
Test 6: The influence of the head slops on device performance. User was asked to perform the "test 3" with his head slope left and right.

Test 7: Hit test has been performed to determinate device resolution. The 15" screen was divided into 16 squares. The user had to concentrate the gaze on the centre of each square several times. All hits were counted and the mean and standard deviation were calculated.

3 Results

Pupil Detection Algorithm Validation

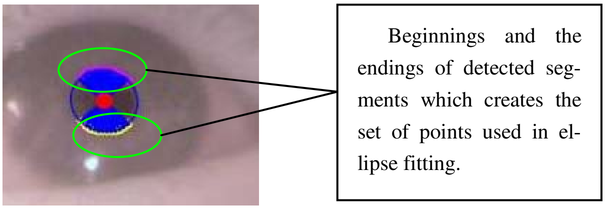
Results of the pupil detection for user looking straight forward (Fig. 2).



(a) (b)

Fig. 2. Results obtained using the pupil detection algorithm a) LSD algorithm result, b) MLSD algorithm result

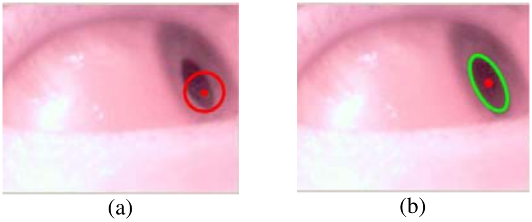
The set of points used to fit the ellipse in case of MLSD algorithm are presented in Fig. 3.



Beginnings and the endings of detected segments which creates the set of points used in ellipse fitting.

Fig. 3. Yellow and purple spots indicates the set of points used to fit the ellipse

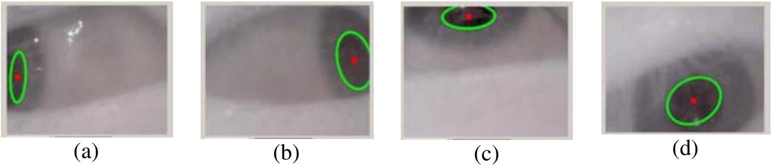
The pupil detection for user extremely deflecting his eye are presented in Fig. 4.



(a) (b)

Fig. 4. Pupil detection algorithm validation a) LSD algorithm result, b) MLSD algorithm result

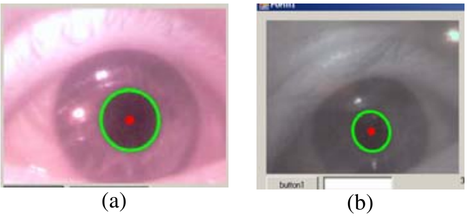
Modified Longest Segment Algorithm Range Validation
User was asked to deflect his eye maximally left, right, up and down without changing his head position. Results are presented in Fig. 5.



(a) (b) (c) (d)

Fig. 5. MLSD algorithm range validation, pupil detection for eye deflected maximally a) left, b) right, c) up, d) down

Results of influence of glasses (Fig. 6b)

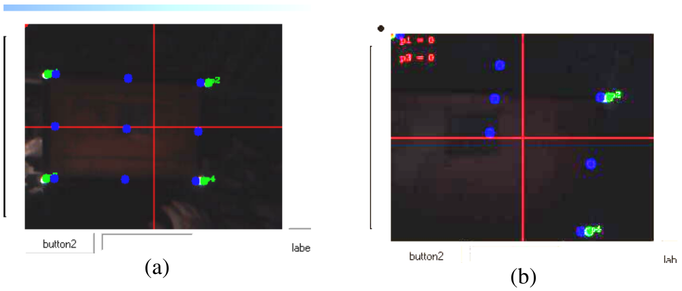


(a) (b)

Fig. 6. Results of pupil detection for user a) without glasses, b) wearing glasses

Screen Detection Algorithm Validation

The green spot indicate which IR markers are detected (fig 7a). When even one marker is not detected (e.g. is out of camera range) user gets visual information which corner is not detected (Fig. 7b). Blue points visible in Fig. 7a refer to points from calibration patterns.

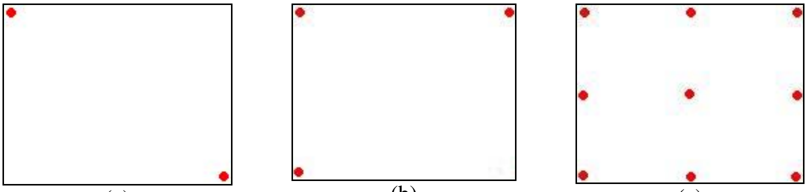


(a) (b)

Fig. 7. Screen detection algorithm, a) correct detection, b) not every corner detected, information set up in left upper corner

Calibration Patterns

Different patterns used in calibration procedure are shown below:

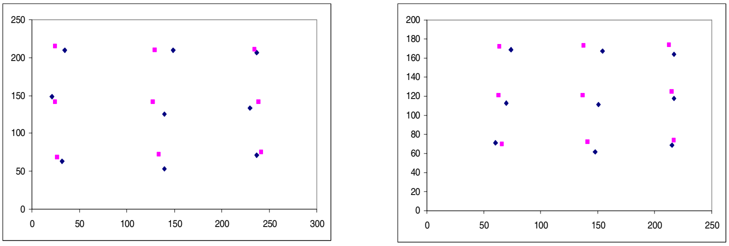


(a) (b) (c)

Fig. 8. Calibration patterns, a) 2-points pattern, b) 3-points pattern, c) 9-points pattern

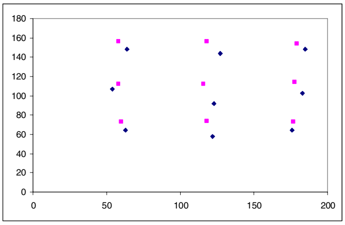
Test of the Influence of the Calibration Procedure on the Estimation of the Fixation Point

The results for 2 points calibration are presented in Fig. 9.



(a) (b)

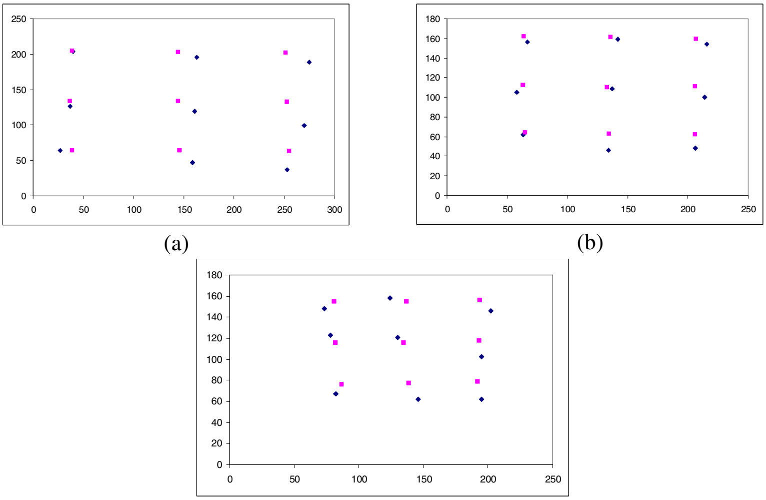
Fig. 9. Calibration procedure validation for different "head to screen" distance, a) 40cm, b) 60cm, c) 90cm. Blue spots refers to fixation point, purple refers to check points



(c)

Fig. 9. (*continued*)

The results for 3 points calibration are presented in Fig. 10.



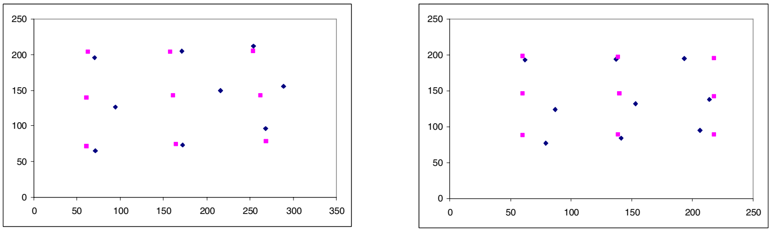
(a)

(b)

(c)

Fig. 10. Calibration procedure validation for different "head to screen" distance, a) 40cm, b) 60cm, c) 90cm. Blue spots refers to fixation point, purple refers to check points

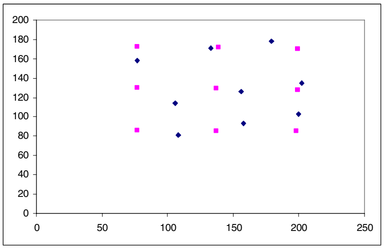
Results for 9 points calibration are presented in Fig. 11.



(a)

(b)

Fig. 11. Calibration procedure validation for different "head to screen" distance, a) 40cm, b) 60cm, c) 90cm. Blue spots refers to fixation point, purple refers to check points



(c)

Fig. 11. (*continued*)

For each test the average, minimum and maximum distance between estimated fixation point and corresponding check point (in pixels) were calculated. The results are gathered in Tables 1-3.

Table 1. Distance between check points and fixation point

Calibration type/distance	**Average**	**Max**	**Min**
2point cal./40cm	12	20	4
2point cal./60cm	10	17	5
2point cal./80cm	12	21	6

Table 2. Distance between check points and fixation point

Calibration type/distance	**Average**	**Max**	**Min**
3point cal./40cm	6	12	1
3point cal./60cm	9	17	2
3point cal./80cm	12	17	7

Table 3. Distance between check points and fixation point

Calibration type/distance	**Average**	**Max**	**Min**
9point cal./40cm	18	32	3
9point cal./60cm	14	34	3
9point cal./80cm	19	33	6

The Influence of "User-to-screen" Distance Changes

Tables 4-6 contain the average, minimum and maximum distance between estimated fixation point and corresponding check point, in pixels.

Table 4. Distance between check points and fixation point

Calibration type/distance	**Average**	**Max**	**Min**
2point cal./40cm	12	20	1
2point cal./60cm	15	29	6
2point cal./80cm	17	27	8

Table 5. Distance between check points and fixation point

Calibration type/distance	**Average**	**Max**	**Min**
3point cal./40cm	6	12	1
3point cal./60cm	8	22	1
3point cal./80cm	11	23	4

Table 6. Distance between check points and fixation point

Calibration type/distance	**Average**	**Max**	**Min**
9point cal./40cm	18	32	3
9point cal./60cm	26	46	4
9point cal./80cm	25	39	13

The Influence of Head Rotation

The head rotation ratio strongly depends on screen camera range. Device estimates fixation point and mouse cursor position only when whole screen is detected. The test results are presented in Fig. 12a.

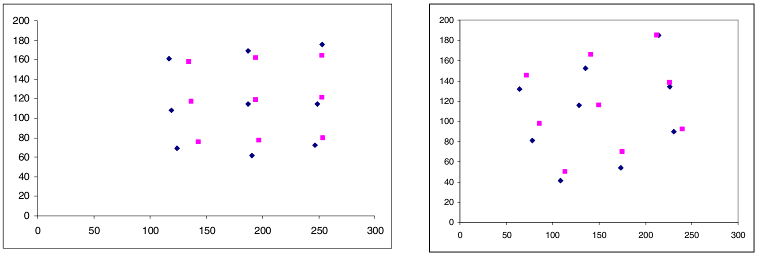
The distance dependence between estimated fixation point and check point is presented in Table 7:

Table 7. Distance between check points and fixation point

Calibration type/distance	**Average**	**Max**	**Min**
3point cal./80cm	13	20	7

The Influence of the Head Slops

The test results are presented in Fig. 12.



(a)

(b)

Fig. 12. a) results for head rotated by the angle, b) results for head sloped left. Blue spots refers to fixation point, purple refers to check points.

The distance dependence between estimated fixation point and check point is presented in Table 8.

Table 8. Distance between check points and fixation point

Calibration type/distance	Average	Max	Min
3point cal./60cm	12	22	2

Hit Test

The examined persons were asked to hit each marked part of screen ten times. All hits were counted and mean and standard deviation were calculated.

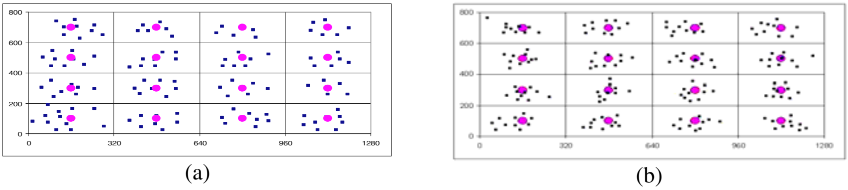


Fig. 13. Hit test, a) results for LSD algorithm, b) results for MLSD algorithm

Two different pupil detection algorithms has been tested. In case of LSD algorithm user had to locate mouse cursor 8 times in every area and 10 times in case of MLSD algorithm.

4 Discussion

Pupil detection algorithms have can be evaluated only subjectively. However, as it can be noticed in Fig. 2, both algorithms work correctly when there are no significant eye deflections. The preponderance of MLSD over LSD algorithm is clearly visible (Fig. 3) when user's eye is deflected. The MLSD algorithm estimates pupil shape correctly in full range of eye movement (Fig. 4). In devices like Eye mouse the balance between user friendly interface and correctness of the device is a very important feature. The test which examined influence of particular calibration pattern on accurateness of fixation point estimation has been performed. Results for different "head to screen" distances are presented in Fig. 9-11. The user had to go through the calibration process every time when the distance between head and screen increased. The optimal pattern used in calibration process has been searched. From Tables 1-3 it is clear that 9-point calibration pattern does not give expected results. Every parameter (average, max. and min. distance) is less satisfied then in case of 2 and 3-point calibration patterns. Next test has clarified which calibration pattern should be used. Comparing data in Tables 1-3 with corresponding ones in Tables 4-6 it can be noticed that they do not differ much only for a three point pattern. In other words, influence of distance between head and screen is not noticeable in this case. Results from Tables 7 and 8 compared with corresponding results from Table 2 show that the developed method allows for head slope and rotation.

In elaborated device we decided to create virtual space in captured by scene-camera image. Virtual space coordinates system is connected to image coordinates system. Every point used in calibration procedure has its representation in virtual space. The procedure of relating the on-screen calibration point with pupil position in fact relates virtual representation of actual calibration points with pupil position. This approach makes the device insensitive to significant head movements and to changes of "user to screen" distance. The user can change the screen placement without necessity of repeating the calibration procedure. However it is true as long as the user's eye, the eye and the screen cameras remain in the same relation as during the calibration procedure. To prove the system performance the hit test procedure has been proposed. The 15" screen has been divided into 16 squares. The user had to concentrate the gaze on the centre of each square several times. All hits were counted and mean and standard deviation were calculated. The standard deviation was calculated from the formula (5):

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}(\bar{x} - x\{n\})^2} \tag{5}$$

where N – is number of all hits directed to each square. Fig. 13a shows calculated results of the single trial for N=8 (LSD algorithm has been used). For each region number of captured gaze positions was counted. This number was from 5 to 12 for each square. The mean value was 8 with standard deviation σ =1,63. Performing the same test using MLSD algorithm, user placed a mouse cursor correctly 10 times in every area (Fig. 13b).

Proposed eye tracking method is fast, reliable and relatively inexpensive. Applying a dual-camera system allows to compensate potential users head movements. The system was tested on healthy subjects, age from 22 to 79, having different iris and skin colour. In all examined cases system appeared to work correctly. All applied algorithms are relatively fast and work in a real-time.

5 Conclusions

Efficient techniques of eye detection in captured images are very important in HCI. Information about the current eye position towards the screen can be used to determinate the mouse cursor position which can give a chance to operate PC by people with serious movement disabilities. We have developed an inexpensive and real-time system using efficient algorithms. System has been tested on 6 persons in different environmental conditions. It appeared to work correctly both in natural and artificial light conditions. Also hit test has shown a relatively good performance. Introducing additional tests, camera enables a correction for the head movement and different position in relation to the screen. It increases a number of correct hits. It is possible to express patient's feelings, and show various information on the screen fitted with specially developed software. The system can operate in real time with modern PC like Pentium 4 CPU with 1GHz clock and 512 MB RAM operating Windows XP system. The source code can be transferred to the other operating systems such as Linux with the Video for Linux layer, but this operation needs more careful webcam selection, as not every web camera has properly working driver for Linux. In this paper we presented solution which could allow the creation of the head's movements free eye tracking device. Although the accuracy of elaborated solution strongly depends on camera's resolution, every attempt of creating this kind of device gives a chance of normal life for suffering people.

References

1. Rebsamen, B., Teo, C.L., Zeng, Q., Ang Jr., M.: Controlling a wheelchair indoors using thought. In: IEEE Intelligent Systems, pp. 18–24 (2007)
2. Chin, C.A.: Enhanced hybrid electromyogram / eye gaze tracking cursor control system for hands-free computer interaction. In: Proc. 28th IEEE EMBS Annual Intern. Conference, New York City, pp. 2296–2299 (2006)
3. Kierkels, J., Riani, J., Bergmans, J.: Using an eye tracker for accurate eye movement artifact correction. IEEE Transactions on Biomedical Engineering 54(7), 1257–1267 (2007)
4. Wang, Q., Yang, W., Wang, H., Guo, Z., Yang, J.: Eye location in face images for driver fatigue monitoring. In: Proc. 6th Intern. Conference on ITS Telecommunications, pp. 322–325 (2006)
5. Xiong, L., Zheng, N., You, Q., Liu, J., Du, S.: Eye synthesis using the eye curve model. In: Proc 19th IEEE Intern. Conference on Tools with Artificial Intelligence, pp. 531–534 (2007)
6. Kaufman, E., Bandopadhay, A., Shaviv, B.D.: An eye tracking computer user interface. In: Proc. Research Frontier in Virtual Reality Workshop, vol. (10), pp. 78–84. IEEE Computer Society Press, Los Alamitos (1993)
7. Hyoki, K., Shigeta, M., Ustno, N., Kawamuro, N.Y., Kinoshita, T.: Quantitative electro-oculography and electroencephalography as indices of alertness. Electroencephalography and Clinical Neurophysiology 106(3), 213–219 (1998)
8. Kocejko, T.: Device which will allow people suffered with lateral amyotrophic sclerosis to communicate with environment. MSc-thesis, University of Technology, Gdańsk, Poland (2008)
9. Myers, G.A., Sherman, K.R., Stark, L.: Eye monitor. IEEE Computer Magazine 3, 14–21 (1991)
10. Collet, C., Finkel, A., Gherbi, R.: A gaze tracking system in man-machine interaction. In: Proc. IEEE Intern. Conference on Intelligent Engineering Systems, vol. (9) (1997)
11. Hu, B., Qiu, M.: A new method for human-computer interaction by using eye-gaze. In: Proc. IEEE Intern. Conference on Systems, Man and Cybernetics, vol. (10) (1994)
12. Ma, Y., Ding, X., Wang, Z., Wang, N.: Robust precise eye location under probabilistic Famework. In: Proc. 6th IEEE Intern. Conference on Automatic Face and Gesture Recognition (2004)

13. Zhu, Z., Ji, Q.: Novel eye gaze tracking techniques under natural head movement. IEEE Transaction on Biomedical Engineering 54(12), 2246–2260 (2007)
14. Miyazaki, S., Takano, H., Makamura, K.: Suitable checkpoints of features surrounding the eye for eye tracking using template matching. In: Proc. SICE Annual Conference, Kagawa University, Japan, pp. 356–360 (2007)
15. Ebisawa, Y.: Improved video-based eye-gaze detection method. In: Proc. IEEE IMTC 1998 Conference, Hamatsu, Japan (1998)
16. Hutchinson, T.E., White Jr., K.P., Martin, W.R., Reichert, K.C., Frey, L.A.: Human-computer interaction using eye-gaze input. IEEE Transactions on Systems, Man and Cybernetics 19(6), 1527–1534 (1998)
17. Ballard, P., Stockman, G.C.: Computer operation via face orientation. In: Proc. 11th IAPR International Conference on Pattern Recognition. Conference A: Computer Vision and Applications, vol. 1, pp. 407–410 (1992)
18. Gee, H., Clipolla, R.: Determining the gaze of faces in images. Technical report CUED/FINFENG/TR 174, University of Cambridge, UK (1994)
19. Web page, `http://www.worldlibrary.net/eBooks/Giveway/Technical_eBooks/OpenCVReferenceManual.pdf` (accessed April 4, 2009)

Eye-Blink Controlled Human-Computer Interface for the Disabled

A. Królak and P. Strumiłło

Institute of Electronics, Technical University of Technology, Łódź, Poland
{aleksandra.krolak,pawel.strumillo}@p.lodz.pl

Abstract. In recent years there has been an increased interest in Human-Computer Interaction Systems allowing for more natural communication with machines. Such systems are especially important for elderly and disabled persons. The paper presents a vision-based system for detection of long voluntary eye blinks and interpretation of blink patterns for communication between man and machine. The blink-controlled applications developed for this system have been described, i.e. the spelling program and the eye-controlled web browser.

1 Introduction

Human-Computer Interface (HCI) can be described as the point of communication between the human user and a computer. Typical input devices used nowadays for communication with the machine are: keyboard, computer mouse, trackball, touchpad and a touch-screen. All these interfaces require manual control and cannot be used by persons impaired in movement capacity. This fact induces the need for development of the alternative method of communication between human and computer that would be suitable for the disabled. Therefore the work on the development of innovative human-computer interfaces attracts so much the attentions of researchers all over the world [1]. For severely paralyzed persons, whose ability of movement is limited to the muscles around the eyes most suitable are systems controlled by eye-blinks since blinking is the last voluntary action the disabled person looses control of [2]. Eye blinks can be classified into three types: voluntary, reflexive and spontaneous. Spontaneous eye blinks are those with no external stimuli specified and they are associated with the psycho-physiological state of the person [3]. Voluntary eye blinks are results of a person's decision to blink and can be used as a method for communication. The eye-movement or eye blink controlled human-computer interface systems are very useful for persons who cannot speak or use hands to communicate (hemiparesis, quadriplegia, ALS). The systems use techniques based mainly on infrared light reflectance or electro-oculography.

The example of the gaze-communication device is Visionboard system [4]. The infrared diodes located in the corners of the monitor allow for detection and tracking of the user's eyes employing the "bright pupil" effect. The system replaces the mouse and the keyboard of a standard computer and provides access to many applications, such as writing messages, drawing, remote control, Internet browsers or

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 123–133.
springerlink.com

electronic mail. However, majority of users were not fully satisfied with this solution and suggested improvements.

More efficient system, based on passive eye and blink detection techniques, was proposed in [5]. The system enables communication using eye blink patterns. The spelling program and two interactive games were prepared for the users of this system.

The application of vision-based analysis of eye blink biometrics was demonstrated in [6]. The authors also examined the influence of ambiguous blink behaviors on specific HCI scenarios. In [7] the authors developed text editor, which is operated using EOG signals. Eye movements are responsible for the movement of the cursor and decision making is simulated by successive blinks. Electro-oculography was also used in [8] to develop a system allowing physically impaired patients to control a computer as assisted communication. The application allowing for entering alphanumeric signs to text editor was controlled by eye blinks. Double eye blinking was used as a decision making signal to avoid errors resulting from physiological blinking. The results confirmed, that eye-movement interface can be used to properly control computer functions and to assist communication of movement-impaired patients.

The vision-based system presented is designed for the disabled users who are capable of blinking voluntarily. The proposed algorithm allows for eye blink detection, estimation of the eye blink duration and interpretation of the sequence of blinks in the real time to control the non-intrusive human-computer interface. The employed image processing techniques are used to measure the area of the visible part of an eye and in this way analyze the eye blink dynamics. Two applications were designed to present the usefulness of the proposed system in human-computer interaction: spelling program named BlinkWriter and BlinkBrowser – software for navigating in the Internet by blinking.

The paper is organized as follows: the next section describes the eye blink monitoring methods, the detailed description of the proposed system and the applications designed for human-computer interaction using the system is given in Sections 3 and 4. Experimental results are presented in Section 5 and Section 6 concludes the paper.

2 Eye-Blink Monitoring

Eye blink detection has a number of applications, like eye blink dynamics analysis, eye detection and tracking or face detection. For this reason a number of eye blink detection techniques have been developed. They can be divided into contact methods requiring direct connection of the measuring equipment to the monitored person, and non-contact methods. Another classification divides eye blink detection techniques into non-vision and vision-based. Non-vision methods include electro-oculography (EOG) and high-frequency transceivers. EOG uses the recordings of the electric skin potential differences collected from the electrodes placed around the eyes. The blinks are identified as the spike waveforms with amplitudes of 0.5÷1mV [9]. Another non-vision-based method for eye blink detection employs the high-frequency (~30GHz) transceivers. The head is illuminated by the transceiver and the analysis of the Doppler components of the reflected signal allow for identification of the eye blinks [10].

Camera based eye blink detection techniques are in general non-contact ones. They make use of some properties or characteristics of an eye that can be detected by the camera. Vision-based methods can be classified into two groups: active and passive. In active eye blink detection additional light sources are required in order to take advantage of the reflective properties of a human eye. In contrast passive eye detection methods do not need special illumination. However, more advanced image processing methods need to be applied on the images taken in natural conditions.

Eye blink detection systems make use of a number of combined image processing methods, usually including active infrared (IR) illumination. The system built by Horng et. al. [11] is based on skin color detection and template matching. Eye blink detection by difference images was designed by Brandt et.al [12]. System proposed by Grauman et. al. [5] employs motion analysis and template matching for eye blink detection and analysis.

The passive vision based system proposed in this paper is constructed in such a way that it is as reliable, safe and user-friendly as possible. The assumptions are:

- nonintrusive system;
- avoid specialized hardware and infrared illumination
- real-time performance;
- use only part of the processing power of the computer;
- run on a consumer grade computer.

3 Eye-Blink Monitoring System

3.1 System Overview

A passive vision-based system for monitoring eye blinks was designed in the Medical Electronics Division in the Technical University of Lodz [13]. The system uses simple commercially available hardware: USB Internet camera and standard personal computer (fig. 1). It was tested on two types of processors: Intel Centrino 1.5GHz and Intel Core2 Quad 2.4GHz. The real-time eye blink detection was achieved with the speed of 30fps., on the images of resolution 320x240. The applications were written using C++ Builder compiler and Intel OpenCV libraries.

The system comprises four main image processing and analysis software modules, as shown in fig. 2. Haar-like face detection [14] is followed by eye localization based on certain geometrical dependencies known for human face. The next step is eye tracking. It is performed using template matching. Eye blinks are detected and analyzed by employing an active contour model [15].

The developed system detects spontaneous and voluntary eye blinks and allows for interpretation of eye blink patterns. The detected eye blinks are classified as short blinks (<200ms) or long blinks (>=200ms). Separate short eye blinks are assumed to be spontaneous and are not included in the designed eye blink code. The example oculogram (plot of eye openness given in percent vs. time) recorded by the system is plotted in fig. 3.

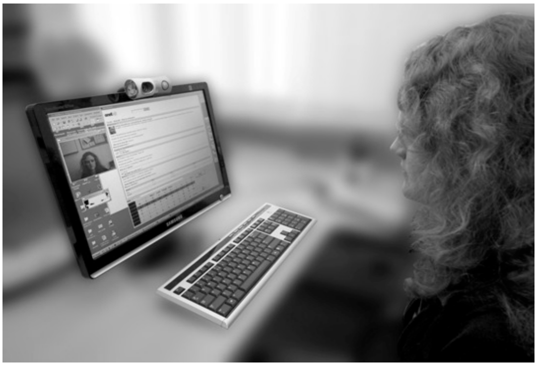


Fig. 1. Test-bench

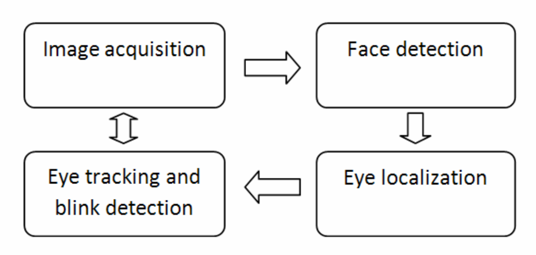


Fig. 2. Block diagram of the system for eye-blink detection

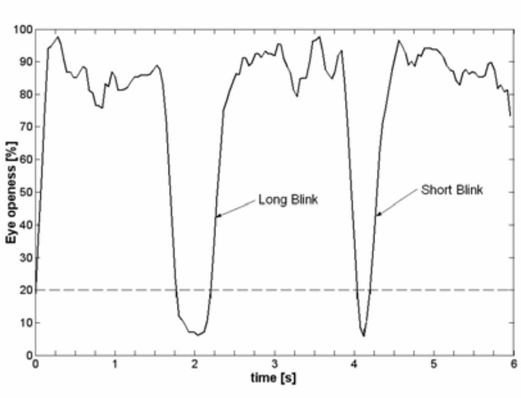


Fig. 3. Example oculogram recorded by the system

3.2 Face Detection and Eye Localization

The first step in the proposed algorithm for eye blink monitoring is face detection. For this purpose the statistical approach using features calculated on the basis of Haar-like masks is employed. The Haar-like features are computed by convolving the image with templates of different size and orientation. The detection decision is carried out by a cascade of boosted tree classifiers. The simple “weak” classifiers are trained on the face images of the fixed size 24×24 pixels. Face detection is done by sliding the search window of the same size as the face images used for training through the test image. The method was tested on the set of 150 face images of different sizes and taken at different lighting conditions. The algorithm recognized 94% of the faces correctly.

The next step of the algorithm is eye localization. The position of the eyes in the face image is found on the basis of certain geometrical dependencies known for human face. The traditional rules of proportion show face divided into six equal squares, two by three. According to these rules the eyes are located about 0.4 of the way from the top of the head to the eyes (fig. 4). The image of the extracted eye region is further preprocessed for performing eye blink detection. Then it is converted from RGB to YCbCr color space and eye regions are extracted by thresholding the image in the red chrominance channel. The threshold level is calculated using the Otsu method [16]. The steps for eye region extraction are presented in fig. 5.

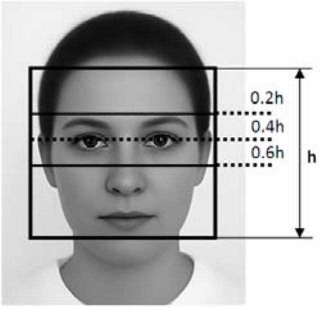


Fig. 4. Eye localization based on face geometry (phantom of average female face from http://graphics.cs.cmu.edu)

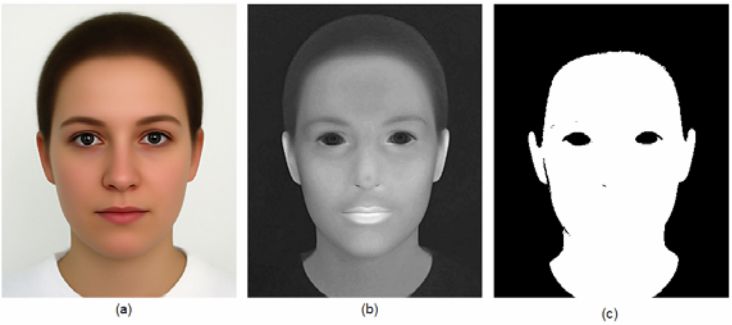


Fig. 5. Eye area detection by skin color segmentation: a) RGB image, b) Cr channel, c) Cr channel image with threshold

3.3 Eye Tracking and Eye Blink Detection

The detected eyes are tracked using normalized cross-correlation method. The template image of the user's eyes is automatically acquired during the initialization of the system. The correlation coefficient calculated is used as a measure of correct detection of the eye region. If it is greater than the predefined threshold value, the algorithm works in a loop of two steps: eye tracking and eye blink detection. If the correlation coefficient is lower than the threshold value, face detection procedure is repeated.

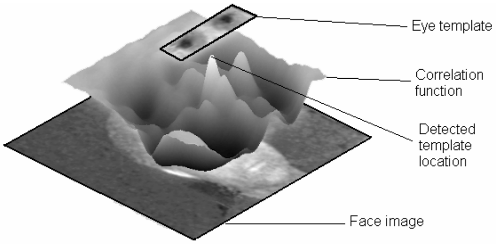


Fig. 6. Eye tracking using template matching

Eye blink detection and analysis is implemented by means of the active contour model. The idea of this technique is to match the computer-generated curve, called a snake, to object boundaries. In the iterative process of energy function minimization the snake becomes deformed and follows the shape of the object's boundary. The energy function E associated with the curve is defined as:

$$E = \alpha E_{cont} + \beta E_{curv} + \gamma E_{img} \quad (1)$$

where E_{cont} is the contour continuity energy, E_{curv} is the contour curvature energy, E_{img} is the image energy and α, β and γ are the weights of the corresponding energies. The role of the contour continuity energy E_{cont} is to make the snake points more equidistant. The curvature energy E_{curv} is responsible for making the contour smoother. Image energy E_{img} depends on the features calculated from the image.

In the proposed algorithm two active contours, one for each eye, are employed for eye blink detection. The initial shape of each snake is an ellipse (fig. 7a). Active contour matching is performed on the binary image of the extracted eye region. The resulting curve (plotted in black in fig. 7c) is approximated to the shape of an ellipse (plotted in white) since the elliptical shape is sufficiently precise model of the eye image. This approximation also facilitates the calculation of the area of the visible part of an eye (eye openness), which is also used for eye closure detection.

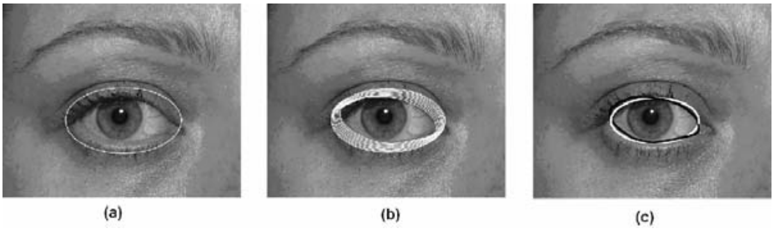


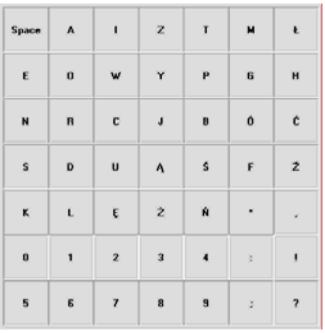
Fig. 7. Eye blink detection using active contour model: (a) snake initialization, (b) iterative deformation of the snake, (c) resulting curve (in black) and approximated ellipse representing the boundaries of the visible part of an eye (in white).

4 The Developed Software

Two applications were created taking advantage of the developed eye-blink detection system: the spelling program named BlinkWriter and a software for load and viewing the web pages called BlinkBrowser.

4.1 BlinkWriter

The BlinkWriter is an application developed for entering alphanumeric signs by blinking. A virtual keyboard containing alphanumeric signs is displayed on the screen. The user accesses the demanded sign by eye-blink-controlled shifting of the active row/column and confirming the selection by performing longer break between the two consecutive eye-blinks (>2s.). The placement of the letters on the screen (fig. 8) was developed using a similar concept to Huffman coding. Thus, letters most often used in Polish language can be selected by shorter sequence of eye blinks.



Space	A	I	Z	T	M	Ł
E	O	W	Y	P	G	H
N	R	C	J	B	Ó	Ć
S	D	U	Ą	Ś	F	Ź
K	L	Ę	Ż	Ń	•	,
0	1	2	3	4	:	!
5	6	7	8	9	;	?

Fig. 8. Letter arrangement on the screen in BlinkWriter application

4.2 BlinkBrowser

The developed system for eye blink detection and analysis was also employed in the application allowing for web browsing by blinking. The graphical user interface of the program is presented in fig. 9.

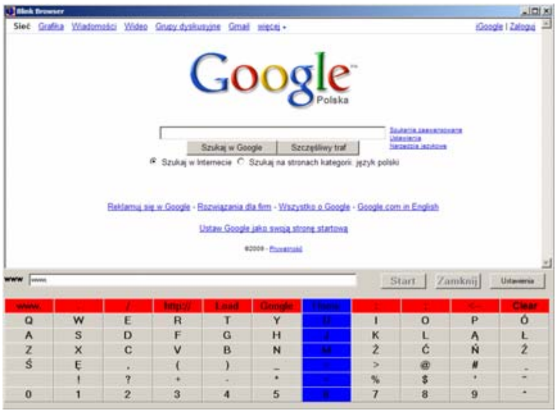
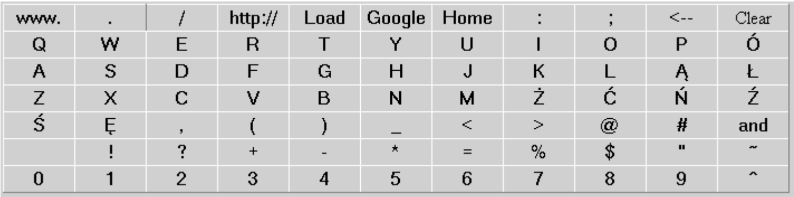


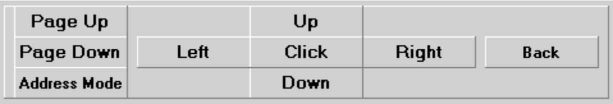
Fig. 9. Graphical user interface of BlinkBrowser

The application works in two modes: "the address mode" and "the view mode". In the address mode the user can enter the URL address from the virtual keyboard (fig. 10) or write a word or phrase to be found on the web. The button "Google" allows for automatic search of the entered phrase by Google search engine. View mode (fig. 11) allows for navigating in the BlinkBrowser window and for controlling the mouse cursor by blinking.



www.	.	/	http://	Load	Google	Home	:	;	<--	Clear
Q	W	E	R	T	Y	U	I	O	P	Ó
A	S	D	F	G	H	J	K	L	Ą	Ł
Z	X	C	V	B	N	M	Ż	Ć	Ń	Ź
Ś	Ę	,	(	)	_	<	>	@	#	and
	!	?	+	-	*	=	%	$	"	~
0	1	2	3	4	5	6	7	8	9	^

Fig. 10. Control panel in the "Address mode" of BlinkBrowser



Page Up		Up		
Page Down	Left	Click	Right	Back
Address Mode		Down		

Fig. 11. Control panel in the "View mode" of BlinkBrowser

5 Results

The developed eye blink monitoring system was tested by 40 healthy persons, aged between 18 and 32 years. Each person was asked to blink 40 times (20 long blinks and 20 short blinks, alternatively). The assessment of the effectiveness of the system was based on three measures: precision, recall and accuracy calculated according to formulas (2), (3) and (4).

$$precision = \frac{TP}{TP + FP} \tag{2}$$

$$recall = \frac{TP}{TP + FN} \tag{3}$$

$$accuracy = \frac{TP}{TP + FP + FN} \tag{4}$$

where:

- TP (True Positives) – number of detected eye blinks when the blink actually appeared,
- FP (False Positives) – number of detected eye blinks when the eye blink did not appeared,
- FN (False Negatives) – number of eye blinks that appear but were not detected by the system.

The overall system precision was equal to 97% and the recall to 98%. The results for short blinks and long blinks detection separately are presented in Table 1.

Table 1. System performance summary

Measure	Precision	Recall	Accuracy
Long eye blink detection	96.91%	98.13%	95.17%
Short eye blink detection	96.99%	98.50%	95.53%
Overall system performance	96.95%	98.31%	95.35%

The functionality of the developed interface was assessed by estimating the time needed to enter the sequence of letters or signs, such as single character, name and surname of the user, given sentence, URL address, but also by assessing time required to move mouse cursor to the desired position and activate the particular link. The results are summarized in Table 2. The average time needed for entering a single character was 10.7s. Moving the mouse cursor from the top left corner of the screen to the bottom right corner took a user 9.5s. in average.

Table 2. Spelling program results

Input sequence	Blink Writer Time [s]	Blink Browser Time [s]
A	6.9	8.1
9	11.8	14.3
CAT	23.1	24.2
MY NAME IS	80.6	91.8
www.yahoo.com	92.4	61.3
Average time needed to enter single character	10.7	12.8

The possibility of head movements while using the system was also tested. The maximum head turn for correct eyen blink detection was equal to ±40°.

6 Conclusions

Obtained results show that the developed algorithm and its software implementation is a viable alternative communication technique suitable for disabled persons. Performed tests demonstrate that the system is able to accurately distinguish between voluntary and involuntary blinks. This is an important aspect as it is used as an interface controlled by facial gestures.

The important advantage of the proposed system is the fact that it does not need prior knowledge of face location or skin color is not required, nor any special lighting. Moreover the system is passive (no IR used) and works in the real time, at a frame rate of 30 fps.

In many human-computer interfaces for the disabled additional hardware is required to be worn by the user, such as special transmitters, sensors, or markers. Since a proposed system uses only the web camera, it is completely non-intrusive, and therefore more user-friendly and easier to configure.

Further testing of the system with users with disabilities is planned to find out what is the most comfortable and effective solution for them. Ideas for the development of the system include work on developing more effective eye blink code and preparing more applications controlled by eye blinks.

Since the eye-blink dynamic changes are reported to be the earliest symptoms of fatigue [17], the developed eye-blink monitoring system is also used as a tool to assess the level of person's workload and drowsiness [18].

References

1. ten Kate, J.H., van der Meer, P.M.: An electro-ocular switch for communication of the speechless. Med. Prog. Technol. 10(3), 135–141 (1983)
2. Stern, J.A., Walrath, L.C., Goldstein, R.: The endogenous eye-blink. Psychophysiology 21(1), 22–33 (1984)
3. Hirokawa, K., Yamada, F., Dohi, I., Miyata, Y.: Effect of gender-types on interpersonal stress measured by blink rate and questionnaires: focusing on stereotypically sex-typed and androgynous types. Social Behavior and Personality (2001), `http://findarticles.com/p/artic-les/mi_qa3852/is_200101/ai_n8941780` (accessed February 26, 2009)
4. Thoumies, P., Charlier, J.R., Alecki, M., d'Erceville, D., Heurtin, A., Mathe, J.F., Nadeau, G., Wiart, L.: Clinical and functional evaluation of a gaze controlled system for the severely handicapped. Spinal Cord 36(2), 104–109 (1998)
5. Grauman, K., Betke, M., Gips, J., Bradski, G.R.: Communication via eye blinks – detection and duration analysis in real time. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 1010–1017 (2001)
6. Heishman, R., Duric, Z., Wechsler, H.: Understanding cognitive and affective states using eyelid movements. In: IEEE Conference on Biometrics: Theory, Applications, and Systems, Washington, DC, pp. 1–6 (2007)

7. Akan, B., Argunsah, A.O.: A human-computer interface (HCI) based on electrooculogram (EOG) for handicapped. In: Proc. IEEE Conference on Signal Processing and Communications Applications, Antalya, Turkey, pp. 1–3 (2007)
8. Borghetti, D., Bruni, A., Fabbrini, M., Murri, L., Sartucci, F.: Low-cost interface for control of computer functions by means of eye movements. Computers in Biology and Medicine 37(12), 1765–1770 (2007)
9. Kong, X., Wilson, G.F.: A new EOG-based eye-blink detection algorithm. Behavior Research Methods, Instruments & Computers 30(4), 713–719 (1998)
10. Colella, T.A.: Drowsy driver detection system (2000), `http://www.jhuapl.edu/ott/technologies/featuredtech/DDDS` (accessed February 27, 2009)
11. Horng, W.B., Chen, C.Y., Chang, Y., Fan, C.H.: Driver fatigue detection based on eye tracking and dynamic template matching. In: Proc. IEEE Conference on Networking, Sensing and Control, Taiwan, pp. 7–12 (2004)
12. Brandt, T., Stemmer, R., Mertsching, B., Rakotonirainy, A.: Affordable visual driver monitoring system for fatigue and monotony. In: Proc. IEEE Conference on Systems, Man and Cybernetics, Hague, the Netherlands, vol. 7, pp. 6451–6456 (2004)
13. Królak, A., Strumiłło, P.: Wizyjny system monitorowania mrugania powiekami w zastosowaniach interfejsów człowiek-komputer. In: Proc 15th Conference Biocybernetyka i inżynieria biomedyczna, Wocław, Poland, p. 165 (2007) (in Polish)
14. Bradski, G., Keahler, A., Pisarevsky, V.: Learning-based computer vision with Intel's open source computer vision library. Intel. Technology Journal 9(2), 119–130 (2005)
15. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: active contour models. International Journal of Computer Vision, 321–331 (1988)
16. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man and Cybernetics 9(1), 62–66 (1979)
17. Lal, S.K., Craig, A.: A Critical Review of the Psychophysiology of Driver Fatigue. Biological Psychology 55(3), 173–194 (2001)
18. Królak, A., Strumiłło, P.: Fatigue Monitoring by Means of Eye Blink Analysis in Image Sequences. In: Proc. Conference on Signals and Electronic Systems, Łódź, Poland, pp. 219–222 (2006)

HMM-Based System for Recognizing Gestures in Image Sequences and Its Application in Continuous Gesture Recognition

A. Wilkowski

Institute of Control and Computation Engineering, Warsaw University of Technology, Poland
a.wilkowski@ia.pw.edu.pl

Abstract. In the article there is presented an efficient system for dynamic gesture recognition in movie sequences based on Hidden Markov Models. The system uses colour-based image segmentation methods and introduces high-dimensional feature vectors to more accurately describe hand shape in the picture. It also utilizes *a-priori* knowledge on gestures structure in order to allow effective dimensionality reduction, hand posture classification and detection schemes. There is also presented a comparison of the algorithm proposed with competitive methods and argued a particular suitability of the system for the situations when only a small amount of training data is available.

1 Introduction

The development of human-computer interaction science requires a new approach in constructing interfaces between human and machine. Today tendency is to promote communication based on natural human means. The most basic ways of inter-human interaction are speech and hand gestures. The methods of speech recognition comprising domains such as signal analysis, feature vector processing and speech flow interpretation has been strongly developed in recent years. Gestures however, are less focused on in research and the problem itself differs much from the previous one at least in terms of methods of data acquisition.

One common approach to the problem of data acquisition in hand gesture recognition is the use of a wearable device for collecting data [1] (data glove). More user transparent approaches base on hand image obtained from a camera. Among those, the methods of image processing vary. Feature vectors can be extracted using e.g. some form of colour analysis [2], detecting local features [3] or differential pictures [4, 5].

The most common solution to interpretation of a gesture as a dynamic sequence is utilization of Hidden Markov Models [6]. In most works no *a-priori* knowledge on gesture structure eg. on trajectory, or hand postures is used. Therefore in order to train and utilize HMMs authors usually resort to continuous state Hidden Markov Models [2] or discreet HMMs supported by the automatic vector quantization algorithm, e.g. LBG [1].

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 135–146.
springerlink.com

In the proposed solution it was assumed that the grammar of available dynamic gestures contains only gestures (which correspond to words) comprised of some sequence of previously defined static hand postures (which correspond to letters or phonems). The proposed constraints give some considerable advantages at the stage of image interpretation and recognition:

- they allow to use feature space transformation methods facilitating discrimination between classes of hand postures,
- they allow to use a wide range of known classifiers including Statistical Classifiers, Neural Networks or Support Vector Machines for hand posture recognition,
- they allow to use model-based methods for object detection and tracking.

In most implementations of HMMs in gesture recognition some arbitrarily chosen, small set of features has been used for recognition. This set could consist for instance of mutual orientation of hands and face, hand size and a few geometrical features [2] or hand position, angle of axis of least inertia and eccentricity of the bounding ellipse [6]. The choice of the best set of features is usually done manually. Results obtained for different sets are compared between one another to choose the best solution. Since the selection of features can be difficult and time-consuming, it seems to be a better idea to, at least partially, automate this process basing on *a-priori* knowledge on the structure of gestures (e.g. the set of hand postures utilized). Therefore, in the proposed solution there was used a large feature vector precisely describing hand shape, and the computational feasibility was ensured by the automatically generated linear projection scheme - LDA aiming at preservation of the most discriminant features for the given set of classes of hand postures.

In the paper there is presented a system for hand gesture recognition implementing both the efficient real-time hand posture recognition as well as recognition of hand posture sequences (dynamic gestures). This article builds upon the previous paper [7]. Therefore, in this article there is given only a shortened overview of the methodology used, so for detailed description please refer to [7]. The main contribution of this work is the evaluation of applicability of the recognition system proposed, in more complex task of *continuous gesture recognition* as well as comparison of the given methodology against several other possible approaches to the problem. The evaluation comprises a comparison of different methods of dimensionality reduction (Linear Discriminant Analysis, Principal Component Analysis and Random Projection) as well as of gesture modeling (discreet-output Hidden Markov Models and continuous-output Hidden Markov Models).

2 Methodology

It is assumed that the source of data for recognition is a single camera. The recognition system is aimed primarily at its utilization in machine control so it uses some artificial vocabulary of sings for communicating commands. The application is able (under some conditions) to automatically isolate a meaningful gesture and to interpret its meaning.

The methodology adopted base on colour analysis of the input pictures, linear projection dimensionality reduction, statistical hand posture classifier and utilization of Hidden Markov Models for dynamic gesture recognition. There is also discussed an adopted method of gesture time segmentation in on-line processing.

2.1 Image Preprocessing and Color Segmentation

Feature vectors used for classification of hand postures are obtained by the geometrical analysis of the area in the image covered by hand. In the proposed approach the segmentation of the hand is based on colour analysis and detection of skin colour in the image.

The colour space used for segmentation in this work was YCbCr. A simple matrix multiplication can be used for transformation between RGB and YCbCr. The resulting colour space can then be a subject to statistical modeling of skin colour distribution based on previously gathered samples [2]. However – in the proposed implementation there has been used static thresholds for skin colour [8]. After image binarization, the morphological operation of n-closure is applied to regularize the resulting shape (in experiments the value n = 3 has been adopted). Then the largest consistent object in the picture is chosen and scaled to fit into the rectangle of 64x64 pixels (starting from the initial resolution of 320x240 pixels).

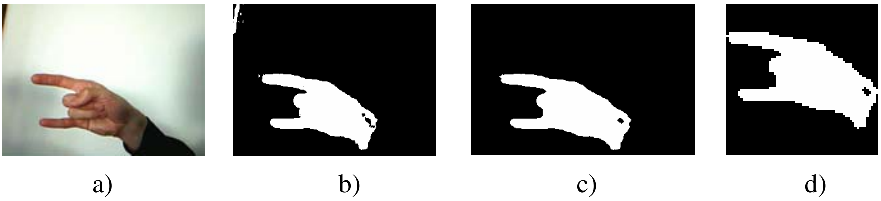


Fig. 1. (a) – input picture, (b) – picture after binarization based on skin colour, (c) – image after applying morphological operations and the choice of the largest object, (d) – hand image after centering and scaling into 64x64 pixels

2.2 Feature Extraction

Feature extraction is an important step in the recognition process. To large extent the quality of classification depends on the choice and representativeness of the extracted features. After binarization and normalization of the hand image described in the previous section the following features, also used in some OCR applications [9] (since we have to do with binary patterns now), are utilized:

- horizontal and vertical projection (histogram) – contains accumulated number of white pixels in single row or column of image,
- image offsets – distances between each edge of the image and the nearest white pixel.

In addition to these 2 types of features there is used another one – which bases on computing of the number of white-black (and black-white) pixel transitions for columns and rows of the image. Some features are extracted with lower resolution, so the total length of feature vectors is 224.

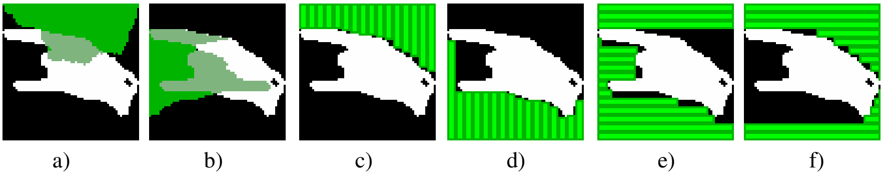


Fig. 2. Example of projection and offset type features (a,b – vertical and horizontal projection, c,d – "north" and "south" offsets, e, f – "west" and "east" offsets)

2.3 Dimensionality Reduction

Modeling probability distributions for large dimensional vectors is usually infeasible due to a fact, that multi-dimensional distributions require large amounts of samples to learn. It is also the case that only a relatively small number of features is required to well characterize classes and the relationships between them.

One way to deal with this problem is to perform a linear projection from a high-dimensional space onto a low-dimensional subspace. The two most widely used in the domain of pattern recognition linear projection methods are based on the Principal Component Analysis and the Linear Discriminant Analysis [10]. Both methods aim at reducing space dimensionality while preserving as much information as possible. With PCA it is possible to preserve to the maximum extent a variance of the data within the remaining dimensions.

As has been already mentioned in the introduction, the recognized gestures are composed of some defined *a-priori* fixed set of hand postures. Thus – the latter method – LDA projection, has been chosen as more suitable solution in this problem. The LDA, unlike PCA, utilizes also information on classes when performing projection. Its goal is to maximize the between-class data scatter $\mathbf{S_B}$ while minimizing the within-class data scatter $\mathbf{S_w}$ in the projected space. The optimization criteria can be defined as the maximization of the ratio given by equation

$$R = \frac{\det(\mathbf{S}_b)}{\det(\mathbf{S}_w)} \tag{1}$$

By solving the associated eigenvalue problem one can obtain the values of linear discriminants and corresponding vectors. The projection matrix is than formed out of the vectors corresponding to the largest linear discriminants.

2.4 Statistical Classifier

In the approach presented in this paper the statistical classifier is used to recognize feature vectors corresponding to different hand postures. The *maximum likelihood* *classifier* was chosen to perform quantization of hand postures. So the classification decision rule is governed be the equation:

$$eval_c(x) = p(x \mid c) \tag{2}$$

where c denotes a class, x is a sample and probability *p(x|c)* depends on the distribution of features over class *c*. The simple procedure for performing classification in this case is to choose the class with the highest value of *p(x|c)*. Since we are dealing with continuous distributions the term *p(x|c),* for the sake of performing classification, can be replaced by the value of probability density function modeling class *c* probed in point *x*. The most widely used solution (that has also been adopted in this work) is to assume the normal distribution of samples within classes and use the multivariate normal distribution probability function as *p(x|c)*. For computability reasons there is actually used a logarithm of the evaluation function (also known as the *log likelihood function*).

Having given *a-priori* knowledge of all allowed classes of hand postures participating in recognition, the classifier can be trained using samples of hand postures labeled as belonging to different classes. Thus the statistical classifier replaces the classic LBG quantizer with an objection that the LBG quantizer is trained using *unsupervised learning* method, while the quantizer proposed uses *supervised learning*.

2.5 Recognition of Gestures

The alphabet of symbols obtained at the hand posture classification stage can be used as an input for higher level recognition processes used for recognition of dynamic gestures. So a discreet-output Hidden Markov Model has been selected to perform the recognition.

In the proposed solution each class of hand posture corresponds to a separate output symbol of the model. Adopted HMM model is the linear left-right HMM model and the number of states is adjusted so each state roughly matches one hand posture (or rather a sequence of consecutive identical postures) that the gesture is composed of. An example of 3-emitting-state Hidden Markov Model modeling a gesture made of 3 different hand postures are given in Fig.3.

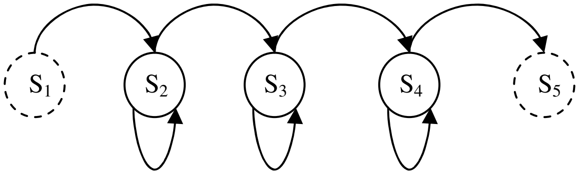


Fig. 3. Example of a hidden Markow model modeling gesture made of a sequence of 3 hand postures. The states marked with dashed line are non-emitting states.

When the structure of the model has been established, the model can be trained using initially a Viterbi training method, and then fine-tuned using Baum-Welsch parameter for re-estimation applied to the training set of gestures.

2.6 Gesture Detection and Segmentation in Time

Recognition of isolated gestures is only a part of the problem to solve when handling the task of using gestures in machine control. Equally important part of this task is the extraction of gestures from a continuous sequence of gestures and random gesticulation (thus reducing the problem of *continuous gesture recognition* into the problem of *isolated gesture recognition*). In the method presented, the problem has been handled by giving special meaning to two of the hand postures and labeling them as <<sil>> (from "silence" in speech recognition) and <<start>>, and assuming that all gestures start with <<start>> posture and ends with <<sil>> posture.

The problem of segmentation can be also addressed in other ways, e.g. as a specific case of *spurious pattern rejection* problem (discussed in [11]), by tracking hand activity (movement) [1] or by HMM filler models [5]. Gesture segmentation in time is not necessary for efficient gesture recognition though it can improve recognition results. Later in this work there will be presented experiments with *continuous gesture recognition* in which gestures are not treated separately but as a part of the whole sentence.

3 Experimental Results

3.1 Recognition of Isolated Gestures

The methodology described, was used to recognize gestures in testing sequences [7]. For the recognition of hand postures there were utilized images of hand postures divided into 9 different classes. The examples from each of the class are presented in Fig. 4.

Each gesture was composed of 3 static hand postures, began with <<start>> hand posture and ended with <<sil>> hand posture (which did not belong to the gesture itself). Training and evaluation settings are described in details in [7]. The gesture segmentation and recognition scheme gave very good results for the test sequence of 25 gestures (all of which were extracted and recognized correctly). In terms of hand-posture recognition there was obtained an accuracy of 92.33% and even 99.8% if the moments of transition between different postures were not taken into account. The "R" package [12] and Hidden Markov Toolkit [13] were used for feature projection and HMM modeling. An efficient implementation ensured also more than real-time algorithm efficiency.

3.2 Comparison with Other Methods

Selection of Methods

The key point of the algorithm proposed is the utilization of *a-priori* knowledge on gesture structure (such as the sequence of hand postures that it is composed of) in the task of dimensionality reduction and statistical modeling using Hidden Markov Models. This information is utilized two times, firstly when parameters of LDA projection scheme are established and secondly, when the statistical classifier performs quantization of hand postures. However, it is quite hard to compare the proposed feature acquisition scheme (feature generation + dimensionality reduction) with other methods (including manual selection of features) due to enormous number of possibilities.

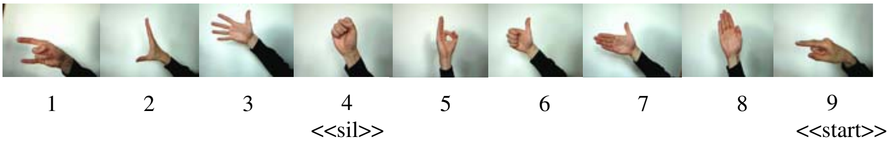


Fig. 4. Examples from classes of recognized hand postures

Therefore for subsequent experiments it was decided to use a single scheme of feature generation (described before) and run the test on a selection of methods for dimensionality reduction and generation of input data for Hidden Markov Models. The proposed solution was compared against two methods of dimensionality reduction which do not take into account the *a-priori* knowledge of gesture structure. Two projection methods fulfilling this requirement have been selected, the Principal Component Analysis [10] projection and the Random Projection [14]. The PCA projection scheme uses some training set to “blindly” retrieve information on the structure of feature space. It analyzes the training set and struggles to find the linear projection matrix which, while reducing the dimension, maintains as much variability of the data set as possible. On the other side - the Random Projection scheme does not use any training at all and produces a random projection matrix W, which elements have been generated from the Gaussian distribution $w_{ij} \sim N(0, 1)$. It is important to underline that neither of these projection schemes use the information of the classes’ structure that the space of hand postures is divided into, however the PCA tries to establish the best (most discriminative) projection directions basing on the structure of the dataset as a whole.

The dynamic statistical model adopted in this work was based on the discreet output HMM model. In this approach the information on classes of hand postures was used to perform vector quantization on data before supplying it into the HMM. So another goal of the evaluation was to confront the proposed method based on discrete output HMM against HMM modeling schemes that do not perform vector quantization and are not dependant on *a-priori* knowledge on classes of hand postures. For this task the continuous-output HMMs have been utilized which model feature vectors distributions in states as multi-dimensional gaussians. Finally, the following test configuration schemes were selected for comparison:

- Random Projection dimensionality reduction + continuous output HMM,
- PCA projection dimensionality reduction + continuous output HMM,

- LDA projection dimensionality reduction + continuous output HMM,
- LDA projection dimensionality reduction + discrete output HMM (the original method).

Testing Environment

Tests for comparing the approaches described above were performed off-line with the use of HTK toolkit [13]. Since the original method gave very good results for testing sequences, some changes to the testing environment were made to increase the difficulty level of recognition. First of all, the gesture segmentation scheme has been turned off, so no hand postures were treated as special articulation signs (such as <<start>> and <<sil>>). The extended set of gestures (which we will call 'words' further on) to recognize consisted of five different symbols "one", "two", "three", "four", "five" (in opposition to three in previously described experiments) and each symbol was composed of 4 hand postures (this is given in Fig. 5).

The number of words recorded amounted to 318 elements and was formed into long sentences of 14-30 words. The input for the recognition process were whole sequences, so it was in fact an example of *continuous gesture recognition* task.

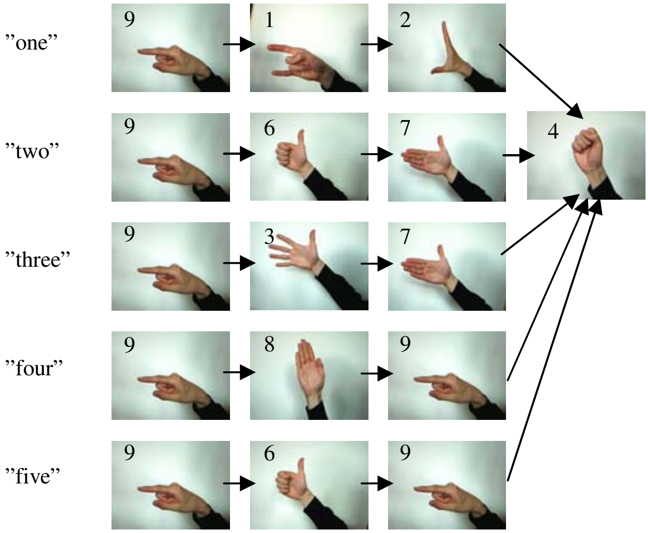


Fig. 5. Recognized dynamic gestures “one”, “two”, “three”, “four”, “five”

In all approaches the same method was used to obtain feature vectors, and then some projection scheme (PCA, LDA or Random Projection) was used to reduce vectors' sizes. For more reliable comparison, the vectors' dimensions in all cases were reduced to 8 (since it is the maximum dimension for the LDA projection in our case) and the gaussian distributions in HMM states were also 8-dimensional. Each dynamic gesture (word) was associated with exactly one HMM. The data processing structure for different evaluations performed is given in Fig. 6.

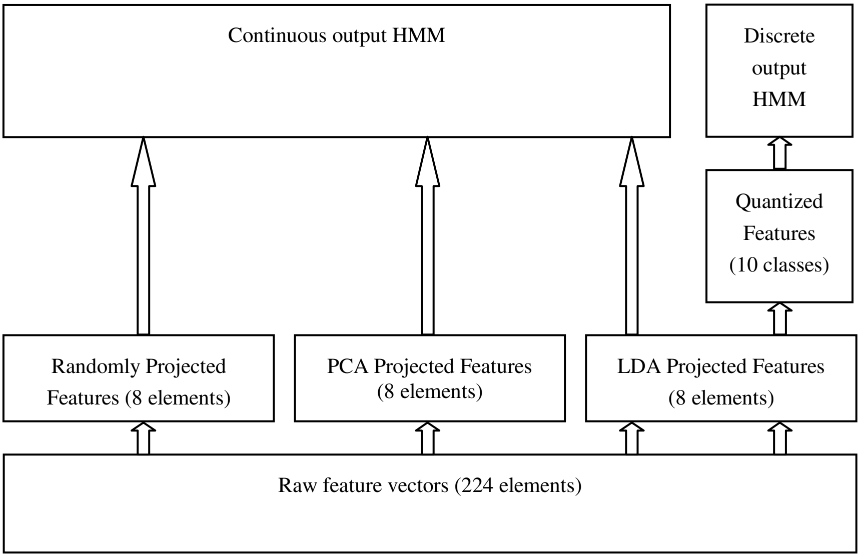

Continuous output HMM
Discrete output HMM
Quantized Features (10 classes)
Randomly Projected Features (8 elements)
PCA Projected Features (8 elements)
LDA Projected Features (8 elements)
Raw feature vectors (224 elements)

Fig. 6. Data processing stages for 4 recognition schemes tested

At the training stage the Viterbi and Baum-Welsch algorithm were used to initialize each model using a few learning samples. This methods, however, require manual segmentation of particular gestures and are cumbersome for large training sets. So, the method of Embedded Training [13] was used to fine tune model parameters. In the case of Embedded Training, the learning process is sentence-based, not word-based. The 'word' HMM models already initialized by the means described above were joined together to form one large HMM model for the whole sentence basing on the word transcription provided by the user. Then the large model was trained using the image sequence containing the whole sentence.

The gesture set was made up of 318 samples. 123 samples were divided into five distinct sentences containing only repeated gestures of the same kind (e.g. only words "one", "two"......). These sentences formed a core of the training set. The rest of the samples were divided between the training and validation set. Behavior of algorithms for different proportions between training and testing data was verified. In order to improve statistical properties of results (especially in case of smaller training sets), the remaining samples were being added either to the training or validation set using the cross-validation principle. The experiments were performed for the following divisions between training and validation sets: 39% - 61%, 69% - 31%, 80% - 20%, 85% - 15%.

In the recognition phase the manually prepared sentence transcriptions were used to evaluate recognition results. The sentences and their transcriptions were compared using the optimal matching algorithm. The results given further on, refer to successful and bad matches between particular words in sentences.

Recognition Results

The recognition rate obtained during the experimental procedure described in the previous section is given in Table 1 and Fig.7. The accuracy computed comprises all types of recognition errors detected when comparing original transcription with recognition results (the errors are comprised of mismatches, omissions and insertions).

Table 1. Recognition rates for different sizes of the training set

	Rand. proj.	PCA	LDA	LDA
	cont. HMM	cont. HMM	cont. HMM	discr. HMM
39% of samples in the training set	12.3%	32.8%	70.3%	88.2%
69% of samples in the training set	65.1%	83.6%	82.6%	88.7%
80% of samples in the training set	76.9%	88.7%	84.6%	88.2%
85% of samples in the training set	72.8%	86.2%	85.1%	88.2%

The most striking difference between 4 methods tested can be observed in case when the smallest number of samples (123) was used in training. In this case, the two methods not dependant on the *a-priori* knowledge on gestures' structure scored below 50% in accuracy (with as little as 12.3% for the random projection). Whereas the method utilizing LDA projection with continuous-output HMMs obtained about 70% of accuracy, and the corresponding method utilizing discrete-output HMMs practically reached a peak of its performance! Together with increase of the number of samples in the training set the results of all methods converge, however (as could be expected), the results for the simplest projection scheme Random Projection still lags more than 10% behind the rest of methods.

With extending of the training set the method based on PCA projection quickly makes up for bad results obtained for the smallest training set and reaches and even surpasses the recognition level of LDA-based projection method with continuous-output HMMs. For all proportions between the training and validation set the results for the LDA-based projection method with discrete-output HMMs were very good and typically better than for other methods by 3-4%. The results of all three algorithms (PCA-based, LDA-based with continuous and dicrete-output HMMs) for larger training sets can be described as satisfying (figures about 80%-90% of recognition rate).

Experimental figures stress the importance of additional knowledge on gesture structures (inherent to the methods based on LDA projection and discreet-output HMM) when the number of training data is severely limited. The overall better performance of discreet-output HMMs can be attributed either to general steeper learning curve of discreet-output HMMs in comparison with the continuous counterpart or to a specific set of tested gestures which were highly "posture"-oriented (the gestures under test were in general constructed of long –lasting known hand postures and quick transitions in between).

Separating the influence of LDA projection and the influence of utilization of statistical classifier together with discreet-output HMM is not straightforward basing on the data obtained. It can be argued however, that both factors have significant

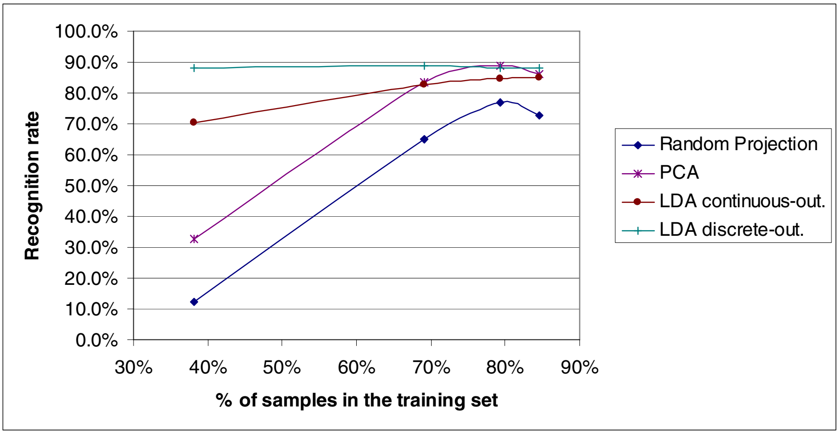


Fig. 7. Comparison of recognition accuracy for different projection schemes and HMM output models

positive impact on performance for small training sets (for the smallest training set they are responsible for 70.3% and 88.2% of accuracy correspondingly vs. 32.8% in case of PCA projection).

4 Conclusions

In the paper there was described an efficient system for both hand posture and dynamic gesture recognition. This paper proves that the concept of using high-dimensional feature vectors in the problem of gesture recognition, can give very good results not only in terms of quality of the recognition (which is more obvious) but also in terms of efficiency. This is mainly thanks to the fact of using the efficient linear projection scheme, which reduces dimensionality of the input space at early stage of processing.

What is more, the use of *a-priori* knowledge regarding hand postures used in dynamic gestures, creates a new possibilities of vector quantization (including use of all modern classifiers trained in a supervised mode), however at the expense of less automatic training stage. As it has been shown – it can also facilitate a solution to the problem of gesture segmentation in time.

Experiments, during which the method proposed was compared against other related methods of recognition, showed that incorporating *a-priori* knowledge of gesture structure in recognition schemes gives largest benefits in case of very small training sets and the results of methods become comparable for larger training sets. This property indicates some possible applications of the method proposed. For instance it could be used in adaptive systems, where the recognizer should learn new gestures very quickly under an assumption that they are based on a pre-specified alphabet of hand postures.

Acknowledgements

The support by the research grant N514 1287 33 from Polish Ministry of Science and Higher Education is gratefully acknowledged.

References

1. Lee, C., Xu, Y.: Online, Interactive learning of gestures for human/robot interfaces. In: Proc. IEEE Conference on Robotics and Automation, Minneapolis, MN, pp. 2982–2987 (1996)
2. Kapuściński, T.: Recognition of the polish sign language in the vision system. PhD dissertation, University of Zielona Góra (2006) (in Polish)
3. Marnik, J.: Polish finger alphabet signs recognition using mathematical morphology and neural networks (2003) (in Polish), `http://www.statsoft.pl/czytelnia/badanianauko-we/d0ogol/marnik.pdf` (accessed April 27, 2009)
4. Rigoll, G., Kosmala, A., Eickeler, S.: High performance real-time gesture recognition using hidden Markov models. Technical report, Gerhard-Mercator-University Duisburg (1998)
5. Eickeler, S., Kosmala, A., Rigoll, G.: Hidden Markov model based continuous online gesture recognition. In: Proc. Conference on Pattern Recognition, Brisbane, Australia (1998)
6. Starner, T., Pentland, A.: Visual recognition of American sign language using hidden Markov models. In: Proc. Intern. Workshop on Automatic Face- and Gesture-Recognition, Zurich, Switherland (1995)
7. Wilkowski, A.: A HMM-Based System for Real-Time Gesture Recognition in Movie Sequences. In: Proc. IEEE Conference on Human System Interaction, Cracow, Poland, pp. 737–742 (2008)
8. Kukharev, G., Nowosielski, A.: Visitor identification – elaborating real time face recognition system. In: Proc. conference on computer graphics, visualization and computer vision, Plzen, Czech Republic, pp. 157–164 (2004)
9. Luckner, M.: Automatic identification of selected symbols of musical notation. Master's Thesis, Warsaw University of Technology (2003) (in Polish)
10. Swets, D.L., Weng, J.Y.: Using discriminant eigenfeatures for image retrieval. IEEE Trans Pattern Analysis and Machine Intelligence 18(8), 831–836 (1996)
11. Wilkowski, A.: An efficient system for continuous hand posture recognition in video sequences. Computational Intelligence: Methods and Applications, EXIT, Warsaw, pp. 411–422 (2008)
12. The R project for statistical computing, `http://www.r-project.org/` (accessed April 27, 2009)
13. Young, S., et al.: The HTK bok: Microsoft Corporation (2000), `http://htk.eng.cam.ac.uk/docs/docs.shtml` (accessed April 27, 2009)
14. Kohonen, T., Kaski, S., Lagus, K., Salojärvi, J., Honkela, J., Paatero, V., Saarela, A.: Self organization of a massive text document collection. IEEE Transactions on Neural Networks, Spec. Issue on NN for Data Mining and Knowledge Discovery 11(3), 574–585 (2000)

Machine Learning of Melanocytic Skin Lesion Images

G. Surówka

Faculty of Physics, Astronomy and Applied Computer Science, Jagiellonian University,
Kraków, Poland
grzegorz.surowka@uj.edu.pl

Abstract. We use some machine learning methods to build classifiers of pigmented skin lesion images. We take advantage of natural induction methods based on the attributional calculus (AQ21) and MLP and SVM supervised methods to discover patterns in the melanocytic skin lesion images. This methodology can be treated as a non-invasive approach to early diagnosis of melanoma. Our feature set is composed of wavelet-based multi-resolution filters of the dermatoscopic images. Our classifiers show good efficiency and may potentially be important diagnostic aids.

1 Introduction

The death rate of patients suffering from malignant melanoma is very high [1]. This is due to inefficient early detection of this malignancy. Even experienced specialists have problems with visual identification of the initial melanoma progression [2].

There are broad attempts to support medical diagnosis with computational intelligence systems. In this work we present machine learning methods for non-invasive dermatological diagnosis of melanoma using multi-resolution wavelet-based decomposition of the pigmented skin lesion images. This approach assumes that some neighborhood properties of pixels in dermatoscopy images can be a sensitive probe of different pigmented skin lesion types and the melanoma progression.

Progression of a pigmented mole consists of some transformations of melanocytes (pigment cells) in the epidermis. There are three steps leading from benign to malignant lesion: melanocytic nevus, dysplastic nevus which reveals some geometrical and cytologic atypia and then some phases of malignant melanoma. The latter comprises the radial growth phase where the lesion expands horizontally within the epidermis and the vertical growth phase where the melanocytic cells expand to the dermis [3].

Dermatologists have measures that help diagnose pigmented skin lesions. The most popular are: ABCDE [4, 5], 7-Point Checklist [6, 7], and the Menzies scale [10]. Those descriptive measures have limited sensitivity especially in the early stages of melanoma development.

Recently a lot of articles on the traditional ABCD rule and the melanoma indicator TDS (Total Dermatoscopic Score) were issued [40] (see references therein). They report on the optimized version of the ABCD rule, which is concluded from

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 147–159.
springerlink.com

melanocytic data analyzed in various machine learning systems (LEM2, LERS, IRIM, C4.5). Machine learning methods (NN, LDA, k-NN) supporting classification of dermatoscopy images can also become a diagnostic aid [35, 36].

Dermatoscopy [9-11, 39] consists in visual examination of skin lesions that are optically enlarged and illuminated by halogen light. The magnified field of the lesion can be digitally photographed and stored by a computer acquisition system [12, 13, 37, 38].

The fully reliable method to identify nevi and melanoma lesions is biopsy. The histologic criteria of the resected lesion, 'Clark' and 'Breslow', assume, however, invasive treatment of the suspicious pigmented spots which is not feasible especially for patients with Atypical Nevus Syndrome (ANS). Development of non-invasive methods for dermatological diagnosis is of key importance [8, 35].

Since the dermatoscopic images can be taken under different conditions (illumination, optical magnification, skin complexion) we construct a feature set from different frequency filters of the skin texture. Those filters are found by means of wavelet transforms which have been widely studied as tools for a multi-scale pattern recognition analysis [14, 15, 18].

Performance of the wavelet analysis of images depends on various factors [16]: recursive (pyramidal) decomposition of the low-frequency branch [17] vs. selective tree-structured analysis [15, 18], order of the transform [17, 18], the wavelet base [19, 20], and 2D [22, 23] vs. 1Dx1D filtering [21]. For that analysis we have chosen a full 3-fold tree-structured decomposition with the 1Dx1D Daubechies algorithm of order 3 [24].

Having assumed that the multi-resolution analysis is well suited to determine distinctive signals characterizing the class of the dermatoscopy image, the wavelet-based features of the dermatoscopy images are then attributes to machine learning classification.

Taking into account the great variety of learning paradigms [25, 26], we decided to probe the performance of the Attributional Calculus (AC) [41], multilayer perceptron (MLP), and the support vector machine (SVM), to make a comparison with the wavelet-based results from the literature.

AC is a logic system that combines elements of propositional calculus, predicate calculus, and multi-valued logic for the purpose of facilitating natural induction, a method of machine learning whose goal is to induce hypotheses from data in forms close to natural language descriptions. We used the AQ21 multitask learning and knowledge mining program [42] whose learning module is based on the concept of a star (a set of maximal generalizations of a given positive example) and its generation.

The choice for the MLP neural network classifier was done due to its ease in implementation and the possibility to test the stability of the solutions [26]. The topology of the network was subject to tests to determine an optimal configuration for maximizing its performance (both quality of the classifier and minimal training time). Evaluation of the features according to their rank was crucial in effective teaching of the neural network.

The Support Vector Machine (SVM) technique maps input vectors to a higher dimensional space to build an optimally separating hyperplane to maximize the margin between the two classes of data [27, 28]. There are a few reasonable kernels that may be used in the SVM method [29]. A linear, polynomial, and a radial basis kernels were taken into account.

In this paper we do not discuss the methodological background of knowledge representation by those machine learning methods used [26, 27, 43, 44], and concentrate on the performance of the constructed classifiers towards their application in the computer-aided diagnosis of melanoma.

The following sections explain our experimental procedure and present the results.

2 Procedure

We have collected anonymous data from 19 unique dermatology patients suffering from malignant melanoma (acquisition time: 18 months). We have selected also 20 cases of dysplastic nevus (available with a higher rate). For all 39 patients both the dermatoscopy images of the pigmented skin lesions and the histologic examinations were available.

The images were collected using a Minolta Dimage Z5 digital camera equipped with an epiluminescence lens with white halogen lighting. The settings of the camera were fixed on resolution of 2272x1704 pixels and quantization depth of 24 bits (RGB-8).

Three ways of extracting the numerical values of pixels are usually possible:

(normal): binary values of the three channels R, G, B are put together (in the presented order) to compose one 24-bit long binary integer. This value is subtracted from the 24-bit long all-‘1’ binary number yielding a negative integer.
(average): an average value of R, G and B is calculated and stored as a floating point number.
(RGB): value of R, G, and B is stored independently in separate matrices. This approach assumes independent processing of the three channels.

The most promising extraction method, measured in terms of the classification performance, was (RGB). We have chosen this data set with help of the Ridge linear model with 40 penalty vectors (see further).

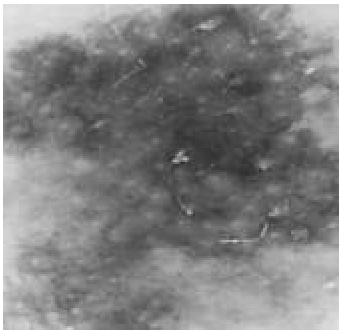


Fig. 1. Texture of one of the analyzed anonymous skin lesions

1.1.1	**1.1.2**	**1.2.1**	**1.2.2**	**2.1.1**	**2.1.2**	**2.2.1**	**2.2.2**
1.1.3	**1.1.4**	**1.2.3**	**1.2.4**	**2.1.3**	**2.1.4**	**2.2.3**	**2.2.4**
1.3.1	**1.3.2**	**1.4.1**	**1.4.2**	**2.3.1**	**2.3.2**	**2.4.1**	**2.4.2**
1.3.3	**1.3.4**	**1.4.3**	**1.4.4**	**2.3.3**	**2.3.4**	**2.4.3**	**2.4.4**
3.1.1	**3.1.2**	**3.2.1**	**3.2.2**	**4.1.1**	**4.1.2**	**4.2.1**	**4.2.2**
3.1.3	**3.1.4**	**3.2.3**	**3.2.4**	**4.1.3**	**4.1.4**	**4.2.3**	**4.2.4**
3.3.1	**3.3.2**	**3.4.1**	**3.4.2**	**4.3.1**	**4.3.2**	**4.4.1**	**4.4.2**
3.3.3	**3.3.4**	**3.4.3**	**3.4.4**	**4.3.3**	**4.3.4**	**4.4.3**	**4.4.4**

Fig. 2. Decomposition process. (Sub-)Numbers denote row/column respective filters: 1=low/low, 2=low/high, 3=high/low and 4=high/high

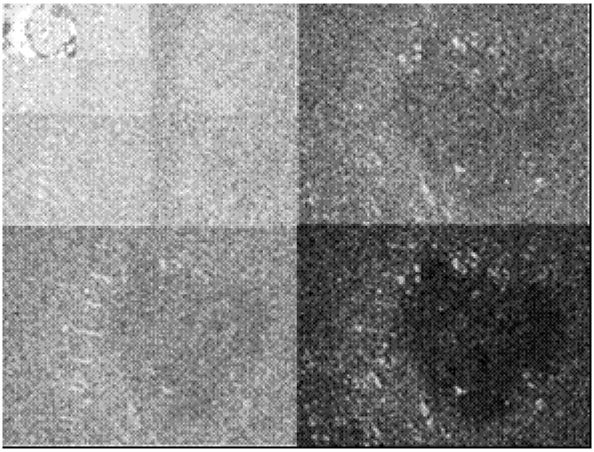


Fig. 3. Recursive decomposition of the low-pass filter (luminosity of the image is offset for this presentation). Decomposition of other branches is not shown here to clearly present the three stages of filtering.

To the three sub-channels of each image a 2D=1D*1D wavelet transform was applied. The class of the filter was Daubechies 3 [24] and its efficient algorithm was taken from [31]. The choice for the filter was made concerning simplicity, performance and possible comparisons with references [14, 15].

An iteration of the wavelet algorithm produces 4 sub-images which can be considered as LL, LH, HL and HH filters, where L and H denote the respective low-pass and high-pass filters. One sub-image is a product of the wavelet transform acting on each row and then on each column of the parent image. Any iteration reduces half of the rows and half of the columns from the parent image. This procedure is depicted in Fig. 2.

In each iteration (resulting in 4 sub-images) 11 coefficients were calculated: energies of the sub-images (e_1, e_2, e_3, e_4), maximum energy ratios (e_i/e_{max}, i=any three out of four except the sub-image with the maximum energy), and the fractional energy ratios ($e_1/(e_2+e_3+e_4)$ + its three permutations), where the term energy means the sum of the absolute values of the pixels normalized by the (sub-)image dimension [14, 15]. The total number of coefficients, i.e. our feature set, had the size of (1+4+16)x11=231 numbers, where 1, 4 and 16 mean here the number of sub-images to filter in each iteration.

Training outputs were coded as (1=melanoma) and (0=dysplastic nevus). Prior to the model construction, input variables were normalized by removing the mean and dividing by the standard deviation for each variable separately.

Two preprocessing methodologies have been used:

1. All the attributes were accepted as potentially significant discriminating signals.
2. Since the number of patterns taken to our analysis was limited to only 39 (19 cases of melanoma + 20 cases of dysplastic nevus), we tried to pre-select the attributes to avoid the overtraining effect.

In case of the inductive learning for feature selection we used the simplified 'promise' method [45] available in AQ21. The 'Attribute_selection_threshold' parameter defining the minimum attribute discriminatory power was experimentally set to 0.9. This value limited our feature set to 4 attributes out of 231 (our initial requirement was arbitrarily set to 'less than 10 attributes'). Since the AQ21 program implements methods introduced in the Attributional Calculus, before selecting the relevant features we decided to digitize continuous domains of the coefficients by invoking the built-in ChiMerge algorithm [46].

For the other two methods for feature selection a Matlab toolbox Entool has been used [30]. Entool is a statistical learning package that has a set of tools for classification and regression analysis. Its performance is relatively high due to ensembling methods. In this analysis the Ridge linear model was used. Ridge regression constructs a model $\hat{y} = X\beta + \beta_0$ (X-data matrix, y-vector of categories), but instead of minimizing the sum of squared residual $(y - X\beta - \beta_0)^T(y - X\beta - \beta_0)$, it minimizes the regularized loss function (Tikhonov regularization):

$$\mathrm{RSS}_{\mathrm{pen.}} = (y - X\beta - \beta_0)^T(y - X\beta - \beta_0) + \lambda\beta^T\beta\,.$$

The additional penalty $\lambda\beta^T\beta$ increases bias moderately whereas the variance of the constructed model is decreased considerably. The penalty parameter $\lambda \geq 0$ can be used to fine tune the bias-variance tradeoff. For this study, the optimal ridge penalty was automatically determined by Leave-One-Out Cross Validation on each training fold individually. We applied an ensemble of one hundred Ridge models with 60 penalty vectors each, to select the most significant features.

Due to small statistics of individuals (39) the classification accuracy was tested by the n-fold cross validation method [32]. We randomly divided the set of all available patterns into 19 subsets and 19 different models were trained using data from 18 sets and validated using the remaining one (the holdout fold).

Table 1. Crucial settings of the AQ21 program (the full index and meaning of parameters can be found in [42])

FEATURE SELECTION
Attribute_selection_method = promise
Attribute_selection_threshold = 0.9

LEARNING
Consequent = [class = *]
Mode = TF / ATF
w=0.3 (for ATF)
MaxRule=1
Cross_validation = 19

TESTING
Method = atest
Threshold = 0.5
Eval_of_conjunction = min
Eval_of_disjunction = max
Eval_of_selector = strict

The generalization step of our classification system was performed by means of three different supervised learning paradigms:

1. **AQ21:** The AQ21 program was run in two different learning modes [42], namely Theory Formation (TF) and Approximate Theory Formation (ATF). In the TF mode, learned rules are complete and consistent, whereas the ATF mode rules consist of the TF rules later optimized according to the given measure Q(w) that reads: $\mathrm{Q(R,w)} = compl(\mathrm{R})^{\mathrm{w}} \cdot consig(\mathrm{R})^{(1\text{-w})}$. This optimization (parameter w) may cause a loss of completeness (compl) and/or consistency (consig) but may increase the predictive power of the learned rule (R) [43]. For our experiments we used settings presented in Table 1.
2. **MLP:** The neural network consisted of three layers. The input layer was composed of ten linear-output neurons with constant unity weights. The hidden layer made of ten neurons and an output layer formed by one neuron had logistic-like activation functions. The topology of the network was subject to tests to determine an optimal configuration for maximizing its performance (both quality of the classifier and minimal training time). At the beginning the weights were set to values randomly chosen from the range of [-1,1] and then modified in the learning process. In one training cycle some 150 000 iterations were performed until the network could classify the input with a defined precision.
3. **SVM:** We used the C-SVM algorithm with the RBF kernel implemented in the osu-svm-3.0 toolbox for Matlab [33]. The C-SVM adds the class mean information into the standard SVM which makes the decision function less sensitive to 'fuzziness' of data or outliers. A grid search over the full parameter space was done to select the most efficient learning parameters C=512 and γ=2.44e-4.

3 Results and Discussion

Numerical results for the AQ21 image classification are presented in Table 2. We use the following notation:

ERROR: the total number of misclassified patterns (misclassified positive patterns / misclassified negative patterns)
AMBIG: the number of ambiguities, i.e. patterns classified to both classes (ambiguous positive patterns / ambiguous negative patterns)
OTHER: the number of patters classified to none of both classes (from positives patterns / from negative patterns),
ε_{AVG}^{PA} : average predictive accuracy for all classes computed separately
=1- (ERROR+OTHER)/(2*#ITERATIONS),
ε_{POS} : average predictive accuracy when ambiguous patterns are classified as positive cases (melanoma) = 1- ($\text{ERROR}_{\text{POS}}$+$\text{OTHER}_{\text{POS}}$)/(#ITERATIONS),
ε_{NEG} : average predictive accuracy when ambiguous patterns are classified as negative cases (dysplastic naevus)
= 1- ($\text{ERROR}_{\text{NEG}}$+$\text{OTHER}_{\text{NEG}}$)/(#ITERATIONS),

The acronym P0.9 states for using only the coefficients selected by the ‘promise’ algorithm.

According to Table 2 the methodology of Approximate Theory Formation seems to fit the best to the inherent fuzzy nature of benign and malignant skin lesion images.

Table 2. Results for the AQ21 image classification

	ATF	**ATF promise**	**TF**	**TF promise**
ERROR	0	2 (1/1)	3 (1/2)	3 (1/2)
AMBIG.	4 (2/2)	2 (1/1)	2 (1/1)	0
OTHER	0	0	4 (2/2)	0
ε_{AVG}^{PA}	100%	95%	82%	92%
ε_{POS}	100%	95%	84%	95%
ε_{NEG}	100%	95%	79%	89%

Results for the neural network classifier and support vector machine are presented below. Accuracy of the Ridge regression and that of the classification model is graphically presented by means of the Receiver Operating Characteristics (ROC) [34]. ROC shows the sensitivity i.e. ratio of correctly identified melanoma lesions (TPF-true positive fraction) in relation to the specificity i.e. ratio of correctly identified dysplastic nevi (1-FPF, FPF-false positive fraction). Area under ROC (AUC) is a numerical measure of the performance.

The first preprocessing step with the Ridge models determined the optimal numerical representation of pixels from the raster. The (arbitrary) criterion was to maximize the classification performance of 20 most significant features. The following values of AUC were obtained: (normal)=85%, (average)=90.8% and (RGB)=93.2%. Separate wavelet processing of R, G, B color channels produced three different values of the channel energy. They had to be summed up before the 11 (per iteration) coefficients were calculated.

Table 3. Ten most significant features according to the Ridge pre-selection

average energy	$e_{4.1.2}$, $e_{2.1.1}$, $e_{3.3.2}$, $e_{3.3.1}$
maximum energy ratio	$e_{1.2.2}/e_{1.2.1}$, $e_{4.2}/e_{4.1}$, $e_{4.3}/e_{4.1}$
fractional energy ratio	$e_{2.3.3}/(e_{2.3.1}+e_{2.3.2}+e_{2.3.4})$, $e_{3.1.1}/(e_{3.1.2}+e_{3.1.3}+e_{3.1.4})$ $e_{3.1.1}/(e_{3.1.2}+e_{3.1.3}+e_{3.1.4})$

The second step was limited to the analysis of the most efficient representation. We applied an ensemble of 100 Ridge models with 60 penalty vectors each to the full (RGB) data. Now the number of features was reduced to only ten (out of 231) most discriminating. Those features are presented in Table 3. For that feature set AUC was increased to 97.4%. ROC is shown in Fig. 4.

In the generalization step the selected feature set was used to teach a three-layer back-propagated neural network and the C-SVM machine. The cross validation set of one melanoma and one dysplastic nevus was shifted between the training runs. The MLP classification was done for 10 most discriminating features to suppress the overtraining effect.

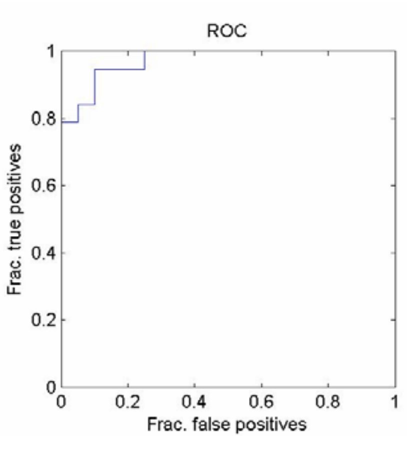


Fig. 4. ROC i.e. (sensitivity) vs. (1-specificity) for the ten most significant features derived with help of 100 ridge regression models. AUC is 97.4%

One of the results is presented in Fig. 5. As we can see, correctly identified lesions are those distributed around 1 on the ordinate for the first 19 indices (images) and around 0 for the other ones. On average two dysplastic nevi and two melanoma images were misclassified yielding the sensitivity (true positive fraction) of 89.2% and specificity (1-false positive fraction) of 90%.

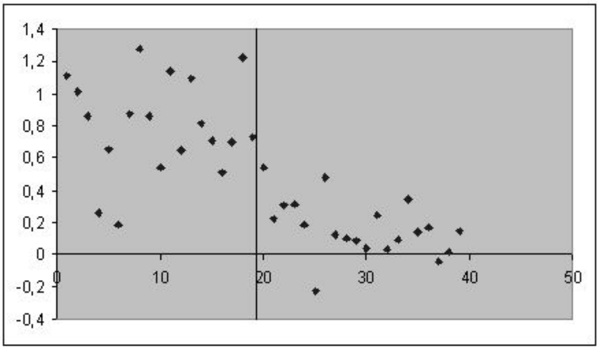


Fig. 5. MLP binary classification of the investigated lesions: the abscissa is the image number (1-19 melanoma, 20-39 dysplastic nevus), the ordinate shows the classification result (1=melanoma, 0=dysplastic nevus)

Due to small statistics (39 skin lesion images) we decided to perform additional tests with the oversampling technique. The (RGB) images limited to the pigmented spot of the mole (background-free) were divided into 64x64 blocks of pixels. Each block was decomposed with the Daubechies-3 wavelets for all possible sub-bands of frequency (3 iterations). Segments grouped to their parent images yielded AUC=92.4%. This method appeared to improve the accuracy of the MLP classifier but taking into account the burden for the processing, it was not satisfactory.

For the SVM classifier, on the other hand, feature selection reduced the information content available for learning from TPF=94.7%, (1-FPF)=95% (231 attributes) to TPF=84.2%, (1-FPF)=85% (10 strongest attributes). In principle the SVM results are more intuitive due to clear concept of the separating hyperplane in contrast to the MLP 'black box' approach.

Since development of pigmented atypia may reveal some transformation stages between benign and malicious lesions the main impact on the classification performance has the inherent fuzzy nature of both classes, 'melanoma' and 'dysplastic nevus'.

MLP and SVM numerical results are gathered in Table 4.

Table 4. Performance of the tested MLP and SVM classifiers

	MLP	SVM RBS	
#attributes	10	231	10
TPF	89.2%	94.7%	84.2%
1-FPF	90%	95%	85%

Classification models of the wavelet-based feature sets of melanocytic skin lesion images were studied in references [18], [15] and [14]. In reference [18] wavelet channel decomposition is controlled by the threshold ratio of the channel average energy to the highest channel energy at the same level of decomposition. In the learning phase (15 melanomas+15 nevi) tree-like topologies of both classes are produced. In the testing phase (10 melanomas+15 nevi) unknown images are decomposed with the thresholds found in the learning phase to build tree structures that are classified due to the Mahanalobis distance from the known patterns. This procedure yields TPF=70% and FPF=20%. The arbitrary selection of the maximum energy thresholds is the main drawback of this method.

In the ADWAT method [15] (like in this work) different combinations of channel energy ratios (not only maximum energy thresholds) are features analyzed to be linearly separable or bimodally distributed between both classes. This histogram-based statistical analysis of features is used to find optimal thresholds for development of the tree-structured wavelet decomposition. In the testing phase unknown skin lesion images are semantically compared with the tree structure of the melanoma and the dysplastic nevus class. Results from [14] for the Adwat method read (learning: 15M+15D, testing: 15M+45D): TPF=86.66%, FPF=11.11%.

In [14] an extended Adwat method is presented. In this approach an additional ‘suspicion lesion’ class is defined. Unknown images are classified as melanoma/dysplastic nevus only if the tree structure completely matches the pattern assigned to one of those classes. Incomplete tree structures are assigned to the suspicious lesion class and their trans-illumination images (580nm and 680nm) are analyzed in the second step. For the learning/testing scheme of 15M+15D and 15M+45D this method yields TPF=93.33% and FPF=8.88%.

When in addition fuzzy membership partitions of the melanoma and dysplastic lesion images by the Bell or Gaussian membership functions are applied, better results are obtained: TPF=100% and FPF=4.44% (FPF=17.77%) for the Gauss (Bell) partition functions.

4 Conclusions and Outlook

19 (20) dermatoscopy images of the melanoma (dysplastic) lesions, all confirmed by histologic examinations, have been classified using a wavelet-based set of features. Discriminant power of those features has been determined by either Ridge regression models, or the ‘promise’ algorithm, and generalized in a three-layer back-propagated neural network/support vector machine, and by the Attributional Calculus.

Our results, presented in Table 2 and Table 4, confirm that neighborhood properties of pixels in dermatoscopy images record the melanoma progression and together with the selected machine learning methods may be important diagnostic aids. Especially the inductive learning approach (AQ21, Table 2) shows great classification potential.

The most critical factor in this paper is the statistics of the melanoma training samples. This is due to low rate (1-2%) of melanoma patients in the population (~1500) examined for the sake of this study (the statistics will be enlarged tenfold by 2009). Nevertheless, the present accuracy of the classifiers is a promising factor for the further research in the field, especially considering the fact that all dermatoscopy images were taken in white light without any specialized instruments (like in [14]).

The constructed model of features for melanoma and dysplastic lesions can evolve, after some fine-tuning and improvements, to a computer-aided diagnostic system.

Acknowledgments

G. Surówka thanks dr Piotr Kanas (GChD) for the database of dermatoscopic images.

References

1. Marks, R.: Epidemiology of melanoma. Clinical and Experimental Dermatology 25(6), 459–463 (2000)
2. Westerhoff, K., McCarthy, W.H., Menzies, S.W.: Increase in the sensitivity for melanoma diagnosis by primary care physicians using skin surface microscopy. British J. Dermatology 143(5), 1016–1020 (2000)
3. Odom, R.B., James, W.H., Berger, T.G.: Melanocytic nevi and neoplasms. In: Andrews' Diseases of the Skin, 9th edn., Philadelphia, pp. 881–889 (2000)
4. Dial, W.F.: ABCD rule aids in preoperative diagnosis of malignant melanoma. Cosmetic Dermatol 8(3), 32–34 (1995)
5. Carli, P., De Giorgi, V., Palli, D., et al.: Preoperative assessment of melanoma thickness by ABCD score of dermatoscopy. J. American Academy Dermatology 43(3), 459–466 (2000)
6. Johr, R.H.: Dermatoscopy: alternative melanocytic algorithms - the ABCD rule of dermatoscopy, menzies scoring method, and 7-point checklist. Clinics in Dermatology 20(3), 240–247 (2002)
7. Argenziano, G., Fabbrocini, G., Carli, P., et al.: Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions: Comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis. Archives of Dermatology 134(12), 1563–1570 (1998)
8. Żabińska-Płazak, E., Wojas-Pelc, A., Dyduch, G.: Videodermatoscopy in the diagnosis of melanocytic skin lesions. Bio-Algorithms and Med-Systems 1, 333–338 (2005)
9. Carli, P., De Giorgi, V., Gianotti, B., et al.: Dermatoscopy and early diagnosis of melanoma. Archives of Dermatology 137, 1641–1644 (2001)
10. Menzies, S.W.: Automated epiluminescence microscopy: human vs machine in the diagnosis of melanoma. Archives of Dermatology 135(12), 1538–1540 (1999)
11. `http://www.dermogenius.com`,
    `http://www.dermamedicalsystems.com` (accessed April 24, 2009)
12. Piccolo, D., Smolle, J., Argenziano, G., et al.: Teledermatoscopy - results of a multicentre study on 43 pigmented skin lesions. J. Telemedicine and Telecare 6(3), 132–137 (2000)
13. Robinson, J.K., Nickoloff, B.J.: Digital epiluminescence microscopy monitoring of high-risk patients. Archives of Dermatology 140(1), 49–56 (2004)

14. Patwardhan, S.V., Dai, S., Dhawan, A.P.: Multi-spectral image analysis and classification of melanoma using fuzzy membership based partitions. Computerized Medical Imaging and Graphics 29(4), 287–296 (2005)
15. Patwardhan, S.V., Dhawan, A.P., Relue, P.A.: Classification of melanoma using tree structured wavelet transforms. Computer Methods and Programs in Biomedicine 72(3), 223–239 (2003)
16. An analysis of wavelet characteristics in image compression. In: Proc. SPIE Conference on Wavelets: Applications in Signal and Image Processing (2003), `http://spiedigitallibrary.aip.org/browse/vol_level.jsp?scode=EBO03&bproc=symp` (accessed March 20, 2009)
17. Mallat, S.G.: A theory for multiresolution signal decomposition: the wavelet representation. IEEE Transactions on pattern analysis and machine intelligence 11(7), 674–693 (1989)
18. Chang, T., Kuo, C.C.J.: Texture analysis and classification with tree-structured wavelet transform. IEEE Transactions on Image Processing 2(40), 429–441 (1993)
19. Kadiyala, M., DeBrunner, V.: Effect of wavelet bases in texture classification using a tree structured wavelet transform. In: Proc. Conference on Signals, Systems, and Computers, vol. 2, pp. 1292–1296 (1999)
20. Mojsilovic, A., Popovic, M.V., Rackov, D.M.: On the selection of an optimal wavelet basis for texture characterization. IEEE Transactions on Image Processing 9(12), 2043–2050 (2000)
21. Porter, R., Canagarajah, N.: A robust automatic clustering scheme for image segmentation using wavelets. IEEE Transactions on Image Processing 5(4), 662–665 (1996)
22. Wang, J.W., Chen, C.H., Chien, W.M., Tsai, C.M.: Texture classification using nonseparable two-dimensional wavelets. Pattern Recognition Letters 19(13), 1225–1234 (1998)
23. Kovacevic, J., Vatterli, M.: Nonseparable multidimensional perfect reconstruction filter banks and wavelet bases for R^n. IEEE Transactions on Information Theory 38(2), 533–555 (1992)
24. Daubechies, I.: Ten lectures on wavelets. Society for Industrial and Applied Mathematics, Philadelphia, PA (1992)
25. Michie, D., Spiegelharter, D.J., Tylor, C.C.: Machine learning, neural and statistical classification. Ellis Horwood, Upper Saddle River (1994)
26. Kuncheva, L.I.: Combining pattern classifiers: methods and algorithms. Wiley & Sons Inc., New York (2004)
27. Vapnik, V.: Statistical learning theory. Springer, Heidelberg (1998)
28. Vapnik, V., Golowich, S., Smola, A.: Support vector method for function approximation, regression estimation, and signal processing. In: Mozer, M., Jordan, M., Petsche, T. (eds.) Advances in neural information processing systems 9, pp. 281–287. MIT Press, Cambridge (1997)
29. Hsu, C.W., Lin, C.J.: A comparison of methods for multi-class support vector machines. IEEE Transactions on Neural Networks 13(2), 415–425 (2002)
30. `http://zti.if.uj.edu.pl/~merkwirth/entool.htm` (accessed April 24, 2009)
31. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical recipes electronic edition, 3rd edn. (2007), `http://www.amazon.com/Numerical-Recipes-Source-Code-CD-ROM/dp/0521706858` (accessed March 20, 2009)
32. Webb, A.: Statistical pattern recognition. Wiley & Sons Inc., New York (2007)

33. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines (2009), `http://www.csie.ntu.edu.tw/~jlin/libsvm` (accessed April 24, 2009)
34. van Erkel, A.R., Pattynama, P.M.: Receiver operating characteristic (ROC) analysis in basic principles and applications in radiology. European Journal of Radiology 27(2), 88–94 (1998)
35. Sboner, A., et al.: A multiple classifier for early melanoma diagnosis. Artificial Intelligence in Medicine 27(1), 29–44 (2003)
36. Hoffman, K., et al.: Diagnostic and neural analysis of skin cancer (DANAOS). British J. of Dermatology 149(4), 801–809 (2003)
37. Kittler, H., et al.: Identification of clinically featureless incipient melanoma using sequential dermatoscopy imaging. Archives of Dermatology 142(9), 1113–1119 (2006)
38. Malvehy, J., Puig, S.: Follow-up of melanocytic skin lesions with digital total-body photography and digital dermatoscopy: a two-step method. Clinics in Dermatology 20(3), 297–304 (2002)
39. Roesch, A., Burgdorf, W., Stolz, W., Landthaler, M., Vogt, T.: Dermatoscopy of 'dysplastic nevi': A beacon in diagnostic darkness. European J. of Dermatology 16(5), 479–493 (2006)
40. Grzymała-Busse, J.W., Hippe, Z.S., Knap, M., Paja, W.: Infoscience technology: the impact of internet accessible melanoid data on health issues. Data Science J. 4, 77–81 (2005)
41. Michalski, R.S.: ATTRIBUTIONAL CALCULUS: a logic and representation language for natural induction. Reports of the Machine Learning and Inference Laboratory, MLI 04-2, George Mason University, Fairfax,VA (2004)
42. Wojtusiak, J.: AQ21 User's Guide. Reports of the Machine Learning and Inference Laboratory, MLI 04-3, George Mason University, Fairfax, VA (2005)
43. Michalski, R.S., Kaufman, K., Pietrzykowski, J., Wojtusiak, J., Mitchell, S., Seeman, W.D.: Natural induction and conceptual clustering: a review of applications. Reports of the Machine Learning and Inference Laboratory, MLI 06-3, George Mason University, Fairfax, VA (2006)
44. Kaufman, K.: INLEN: A methodology and integrated system for knowledge discovery in databases. Ph.D. dissertation, School of Information Technology and Engineering (1997)
45. Baim, P.: The PROMISE method for selecting most relevant attributes for inductive learning systems. Reports of the Intelligent Systems Group, ISG 82-1, UIUCDCS-F-82-898, Department of Computer Science, University of Illinois, Urbana (1982)
46. Kerber, R.: Chimerge: discretization for numeric attributes. In: Proc. 10th Conference on Artificial Intelligence, pp. 123–128. AAAI Press, Menlo Park (1992)

A Preliminary Attempt to Validation of Glasgow Outcome Scale for Describing Severe Brain Damages

J.W. Grzymała-Busse[1,2], Z.S. Hippe[2], T. Mroczek[2], W. Paja[2], and A. Bucinski[3]

[1] Department of Computer Science, University of Kansas, Lawrence, KS
jerzy@eecs.ku.edu

[2] Institute of Biomedical Informatics,
University of Information Technology and Management, Rzeszów, Poland
{zhippe,tmroczek}@wsiz.rzeszow.pl

[3] Department of Biopharmacy, Faculty of Pharmacy,
Collegium Medicum of Nicolaus Copernicus University, Bydgoszcz, Poland
kizbiofarmacji@cm.umk.pl

Abstract. The main goal of our research was to investigate the Glasgow Outcome Scale (GOS), one of several measures applied to the evaluation of patient's functional agility and his/her condition after brain damage therapy. In the first stream of the research, our attention was focused on the question of the importance of particular parameters, used by medical staff for the description of the patients situation after the stroke. Our current research was based on the application of own, in-house developed data mining systems to the same dataset describing the GOS. Results of our experiments shows that from 42 descriptive attributes, just 5 from them were never accessed in the decision rules induction process. On the other hand parameters having the larger frequency quotient (importance of the parameter in diagnostic process) should be diagnosed very carefully.

1 Introduction

According to many sources [1-3] the brain damages seems to be one of very widespread civilization illnesses, occurring at various levels of severity, usually described by means of various measures (scales) [4]. However, it is worth to emphasize that no single outcome measure can describe or predict all dimensions of recovery and disability after acute stroke. Several scales have proven reliability and validity in stroke trials [5], including the *National Institutes of Health Stroke Scale* (called NIHSS), the *modified Rankin Scale* (mRS, patient's functional agility), the *Barthel Index* (BI), the *Glasgow Outcome Scale* (GOS, patient's condition after therapy), and the *Stroke Impact Scale* (SIS). Several scales have been combined in stroke trials to derive a global statistic to better define the effect of acute interventions, although this composite statistic is not clinically tenable. In practice, the NIHSS is useful for early prognostication and serial assessment, whereas the BI is useful for

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 161–170.
springerlink.com

planning rehabilitative strategies. The mRS and GOS provide summary measures of outcome and might be most relevant to clinicians and patients considering early intervention, whereas the SIS was designed to measure the patient's perspective on the effect of stroke. Familiarity with these scales could improve clinicians' interpretation of stroke research and their clinical decision-making.

2 The Main Goal of the Research

Recently, a comparison of Rankin and GOS scales applied to the evaluation of a real database of anonymous patients with severe brain damage was released [6]. It was found, that both scales produce large and unpredictable errors, possibly caused by the fact that descriptive parameters in both scales are ambiguous, complicated, and difficult to set correctly even for experienced medical staff. This finding inspired us to investigate in more details the GOS scale, trying to estimate which of the scale parameters (descriptive attributes) play the most important role in the evaluation of patient's condition after therapy. On the other hand, the research should provide information which parameters are of low significance, hence, can be avoided or even removed. This approach can generally alert the medical personnel and focus the attention on the most important parameters of the GOS scale, and clarify their influence on the final assignment to a given class of patient's condition after therapy. The basic idea of the research was based on the use of specialized data mining systems to search for the deep knowledge, hidden in the database. This knowledge was revealed in the form of various sets of belief rules of the type IF ... THEN. These set of rules (called by us also *models*) build some relations between the description of an anonymous patient and his/her state, according to real (assigned by an medical expert) class in *Glasgow Outcome Scale.*

3 General Methodology of the Research

The main objective of our current research was based on the application of own, in-house developed data mining systems (*Belief*SEEKER [7], *New*GTS [8], *Rule*-SEEKER [9], and LERS [10] to the same data set describing the *Glasgow Outcome Scale* for 162 anonymous patients. In this data set 42 descriptive attributes were used (names of attributes and their allotted values are gathered in the Table 1), while cases were divided into five classes (concepts), accordingly to scores of the *Glasgow Outcome Scale*: **1** (death), **2** (persistent vegetative state), **3** (severe disability), **4** (moderate disability) and **5** (good recovery). As it was stated in Section 2, data mining systems mentioned were used to generate learning models in the form of sets of production rules. Additionally, we used the LERS data mining system for discretization of investigated data, based on divisive cluster analysis [11]. These sets of rules were then analyzed following two separate approaches. In the first one we apply some research schemes, described by us in [12], to evaluate the change of the error rate caused by optimization of learning models (particularly, by the change the number of rules, from 16 to 41). In the second approach, the set of rules were analyzed with the aim to find the most important descriptive attributes

Table 1. Variables (descriptive attributes) applied in validation of the *Glasgow Outcome Scale*

	Variable		
Context	**Code**	**Name**	**Allowed values**
General data about patient	***A1***	**Gender**	*<Male>* ; *<Female>*
	A2	**Admission_diagnosis** (Acc. to ICD-10 classification)	*<Subarachnoid_hemorrhage>* *<Intracerebral_hemorrhage>* *<Cerebral_infarction>* *<Stroke>* *<Other_cerebrovascular_diseases>*
	A3	**Final_diagnosis** (Acc. to ICD-10 classification)	*<Subarachnoid_hemorrhage>* *<Intracerebral_hemorrhage>* *<Cerebral_infarction>* *<Stroke>* *<Other_cerebrovascular_diseases>*
	A4	**Body_temperature** [^{0}C]	Discrete variable
	A5	**Age** [years]	Discrete variable
	A6	**Abode**	*<Town>* ; *<Village>*
	A7	**Time spent in hospital** [days]	Discrete variable
	A8	**Time_elapsed** (*from* observation of illness occurrence *to* hospital admission)	*<Less_than_1_hour>* *<Less_than_3_hours>* *<3-6_hours>* *<6-12_hours>* *<12-14_hours>* *<2-3_days>* *<More_than_3_days>*
	A9	**Patient_cure_location**	*<Stroke_ward>* *<Neurology_ward>*
Patient's specific features	***B1***	**Arterial_hypertension**	*<Present>* ; *<Absent>*
	B2	**Ischemic_heart_disease**	*<Present>* ; *<Absent>*
	B3	**Past_cardiac_infarct**	*<Present>* ; *<Absent>*
	B4	**Atrial_fibrillation**	*<Present>* ; *<Absent>*
	B5	**Organic_heart_disease**	*<Present>* ; *<Absent>*
	B6	**Circulatory_insufficiency**	*<Present>* ; *<Absent>*
	B7	**Diabetes**	*<Present>* ; *<Absent>*
	B8	**Hypercholesterolemia**	*<Present>* ; *<Absent>*
	B9	**Obesity**	*<Present>* ; *<Absent>*
	B10	**Transient_ischemic_attack**	*<Present>* ; *<Absent>*
	B11	**Past_stroke**	*<Present>* ; *<Absent>*
	B12	**Infection_in_a_week_to_stroke**	*<Present>* ; *<Absent>*
	B13	**Alcohol_addiction**	*<Present>* ; *<Absent>*
	B14	**Nicotine_addiction**	*<Present> ; <Absent>*
Condition of health	***C1***	**Systolic_pressure**	*<Present> ; <Absent>*
	C2	**Diastolic_pressure**	*<Present> ; <Absent>*
	C3	**Pulse**	Discrete variable
	C4	**Heart_action**	*<Normal_rythm>* *<Atrial_fibrylation>* *<Other_dysrythmia>*

Table 1. (*continued*)

	C5	**General_state_at_admission**	<*Getting_up_alone*> <*Staying_in_bed_consc.*> <*Consciousness_disturbances*>
	C6	**Consciousness_at_admission**	<*Conscious*> <*Coma*> <*Consciousness_disturbances*>
	C7	**Stroke_type*** (Acc. to Oxford classification, OCSP)	<*LACS*> <*PACS*> <*POCS*> <*TACS*> <*Hard_to_class*>
Disorders	*D1*	**Consciousness_disorders** (during cure)	<*Present*> ; <*Absent*>
	D2	**Speech_disorders** (during cure)	<*Present*> ; <*Absent*>
	D3	**Swallowing_disorders** (during cure)	<*Present*> ; <*Absent*>
Treatment	*E1*	**Aspirine_treatment**	<*Present*> ; <*Absent*>
	E2	**Anticoagulants**	<*Present*> ; <*Absent*>
	E3	**Antibiotics**	<*Present*> ; <*Absent*>
	E4	**Antihypertensives**	<*Present*> ; <*Absent*>
	E5	**Anti-edematous_agents**	<*Present*> ; <*Absent*>
	E6	**Neuroprotective_agents**	<*Present*> ; <*Absent*>
Rehabilitation	*F1*	**Exercise_therapy**	<*Present*> ; <*Absent*>
	F2	**Speech_therapy**	<*Present*> ; <*Absent*>
	F3	**Occupational_therapy**	<*Present*> ; <*Absent*>
Decision		**Glasgow_Outcome_Scale**	**1** (death) **2** (persistent vegetative state) **3** (severe disability) **4** (moderate disability) **5** (good recovery)

applied in the *Glasgow Outcome Scale*, and to validate their overall importance. At this point of the discussion is worth to stress that the original data set was incomplete: 72 attribute values were missing. First, the missing attribute values were replaced by the most common values typical for a given class. Using this method, for any case *x* and attribute *a* with a missing attribute value, we restricted our attention to all cases from the same class as *x*, and the missing attribute value for the attribute *a,* was replaced by the most frequent value of *a* restricted to the given class.

4 Results of Experiments

An outcome of *Belief*SEEKER, a Bayesian network, was converted into a set of rules. The rule set produced by *Belief*SEEKER contains only sixteen rules, all of them have five conditions; they are shown (in alphabetic order) in Table 3. These rules classified unseen cases (unknown patients) with the error rate on the level of roughly 38%.

In the next step, the initial set of 16 rules was improved and optimized, using successively *New*GTS and *Rule*SEEKER. The resulting learning model (Table 2) contains 41 rules. This improved model classified unseen cases (unknown patients) with a very low error rate about 3%. Other results, shown in Table 3-Table 5, are discussed in Section 5.

Table 2. Initial learning model (in the form of production rules)

RULE 1
IF State_at_admission IS(ARE) coma AND Body_temperature IS(ARE) 36..37.6[C] AND Anti-edematous_agents IS(ARE) present AND Exercise_therapy IS(ARE) absent AND Occupational_therapy IS absent THEN GLASGOW_OUTCOME_SCALE IS **1**

RULE 2
IF State_at_admission IS(ARE) coma AND Body_temperature IS(ARE) 37.6..39[C] AND Anti-edematous_agents IS(ARE) present AND Exercise_therapy IS(ARE) absent AND Occupational_therapy IS(ARE) absent THEN GLASGOW_OUTCOME_SCALE IS **1**

RULE 3
IF State_at_admission IS(ARE) consciousness_disorders AND Body_temperature IS(ARE) 36..37.6[C] AND Anti-edematous_agents IS(ARE) present AND Exercise_therapy IS(ARE) absent AND Occupational_therapy IS(ARE) absent THEN GLASGOW_OUTCOME_SCALE IS **1**

RULE 4
IF State_at_admission IS(ARE) consciousness_disorders AND Body_temperature IS(ARE) 36..37.6[C] AND Anti-edematous_agents IS(ARE) absent AND Exercise_therapy IS(ARE) absent AND Occupational_therapy IS(ARE) absent THEN GLASGOW_OUTCOME_SCALE IS **1**

RULE 5
IF State_at_admission IS(ARE) conscious AND Body_temperature IS(ARE) 37.6..39[C] AND Anti-edematous_agents IS(ARE) absent AND Exercise_therapy IS(ARE) absent AND Occupational_therapy IS(ARE) absent THEN GLASGOW_OUTCOME_SCALE IS **1**

RULE 6
IF State_at_admission IS(ARE) conscious AND Body_temperature IS(ARE) 37.6..39[C] AND Anti-edematous_agents IS(ARE) present AND Exercise_therapy IS(ARE) present AND Occupational_therapy IS(ARE) absent THEN GLASGOW_OUTCOME_SCALE IS **3**

RULE 7
IF State_at_admission IS(ARE) conscious AND Body_temperature IS(ARE) 37.6..39[C] AND Anti-edematous_agents IS(ARE) absent AND Exercise_therapy IS(ARE) present AND Occupational_therapy IS(ARE) absent THEN GLASGOW_OUTCOME_SCALE IS **3**

RULE 8
IF State_at_admission IS(ARE) coma AND Body_temperature IS(ARE) 36..37.6[C] AND Anti-edematous_agents IS(ARE) present AND Exercise_therapy IS(ARE) present AND Occupational_therapy IS(ARE) absent THEN GLASGOW_OUTCOME_SCALE IS **3**

RULE 9
IF State_at_admission IS(ARE) coma AND Body_temperature IS(ARE) 36..37.6[C] AND Anti-edematous_agents IS(ARE) absent AND Exercise_therapy IS(ARE) present AND Occupational_therapy IS(ARE) absent THEN GLASGOW_OUTCOME_SCALE IS **3**

RULE 10
IF State_at_admission IS(ARE) consciousness_disorders AND Body_temperature IS(ARE) 36..37.6[C] AND Anti-edematous_agents IS(ARE) absent AND Exercise_therapy IS(ARE) present AND Occupational_therapy IS(ARE) absent THEN GLASGOW_OUTCOME_SCALE IS **3**

RULE 11
IF State_at_admission IS(ARE) consciousness_disorders AND Body_temperature IS(ARE) 36..37.6[C] AND Anti-edematous_agents IS(ARE) present AND Exercise_therapy IS(ARE) present AND Occupational_therapy IS(ARE) present THEN GLASGOW_OUTCOME_SCALE IS **4**

RULE 12
IF State_at_admission IS(ARE) conscious AND Body_temperature IS(ARE) 36..37.6[C] AND Anti-edematous_agents IS(ARE) absent AND Exercise_therapy IS(ARE) present AND Occupational_therapy IS(IS) present THEN GLASGOW_OUTCOME_SCALE IS **4**

RULE 13
IF State_at_admission IS(ARE) conscious AND Body_temperature IS(ARE) 36..37.6[C] AND Anti-edematous_agents IS(ARE) absent AND Exercise_therapy IS(ARE) absent AND Occupational_therapy IS(ARE) present THEN GLASGOW_OUTCOME_SCALE IS **5**

RULE 14
IF State_at_admission IS(ARE) conscious AND Body_temperature IS(ARE) 36..37.6[C] AND Anti-edematous_agents IS(ARE) absent AND Exercise_therapy IS(ARE) absent AND Occupational_therapy IS(ARE) absent THEN GLASGOW_OUTCOME_SCALE IS **5**

RULE 15
IF State_at_admission IS(ARE) consciousness_disorders AND Body_temperature IS(ARE) 36..37.6[C] AND Anti-edematous_agents IS(ARE) present AND Exercise_therapy IS(ARE) present AND Occupational_therapy IS(ARE) absent THEN GLASGOW_OUTCOME_SCALE IS **5**

RULE 16
IF State_at_admission IS(ARE) consciousness_disorders AND Body_temperature IS(ARE) 37.6..39[C] AND Anti-edematous_agents IS(ARE) absent AND Exercise_therapy IS(ARE) present AND Occupational_therapy IS(ARE) absent THEN GLASGOW_OUTCOME_SCALE IS **5**

Table 3. GOS parameters applied in the initial learning model

Parameter	Frequency quotient
Anti-edematous_agents	16/16
Body_temperature	16/16
Exercise_therapy	16/16
Occupational_therapy	16/16
State_at_admission	16/16

Table 4. Optimized learning model (in the form of production rules)

RULE 1
IF State_at_admission IS(are) coma AND Exercise_therapy IS(ARE) absent THEN GLASGOW_OUTCOME_SCALE IS **1**

RULE 2
IF State_at_admission IS(ARE) consciousness_disturbances AND Anti-edematous_agents IS(ARE) present AND Exercise_therapy IS(ARE) absent THEN GLASGOW_OUTCOME_SCALE IS **1**

RULE 3
IF State_at_admission IS(ARE) consciousness_disturbances AND Anti-edemaTHENus_agents IS(ARE) absent AND Exercise_therapy IS(ARE) absent THEN GLASGOW_OUTCOME_SCALE IS **1**

RULE 4
IF Body_temperature IS(ARE) 37.6..39[C] AND Exercise_therapy IS(ARE) absent THEN GLASGOW_OUTCOME_SCALE IS **1**

RULE 5
IF Body_temperature IS(ARE) 36..37.6[C] AND Past_stroke IS(ARE) absent AND Swallowing_disorders IS(ARE) present AND Anti-edematous_agents IS(ARE) present AND Occupational_therapy IS(ARE) absent THEN GLASGOW_OUTCOME_SCALE IS **1**

RULE 6
IF Time_elapsed IS(ARE) less_than_3_hours AND Circulatory_insufficiency IS(ARE) absent AND Hypercholesterolemia IS(ARE) absent AND Consciousness_disorders IS(ARE) present THEN GLASGOW_OUTCOME_SCALE IS **1**

RULE 7
IF Atrial_fibrilation IS(ARE) absent AND Nicotine_addiction IS(ARE) present AND Stroke_type IS(ARE) TACS THEN GLASGOW_OUTCOME_SCALE IS **1**

RULE 8
IF Stroke_type IS(ARE) POCS AND Anti-edematous_agents IS(ARE) present THEN GLASGOW_OUTCOME_SCALE IS **2**

RULE 9
IF Admission_diagnosis IS(ARE) stroke AND Past_stroke IS(ARE) absent AND Heart_action IS(ARE) normal_rythm AND Consciousness_disorders IS(ARE) present AND Antibiotics IS(ARE) absent THEN GLASGOW_OUTCOME_SCALE IS **2**

RULE 10
IF Body_temperature IS(ARE) 36..37.6[C] AND State_at_admission IS(ARE) consciousness_disturbances AND Anti-edematous_agents IS(ARE) absent AND Exercise_therapy IS(ARE) present THEN GLASGOW_OUTCOME_SCALE IS **3**

RULE 11
IF Body_temperature IS(ARE) 37.6..39[C] AND State_at_admission IS(ARE) conscious AND Exercise_therapy IS(ARE) present THEN GLASGOW_OUTCOME_SCALE IS **3**

RULE 12
IF Transient_ischemic_attack IS(ARE) absent AND Past_stroke IS(ARE) absent AND Heart_action IS(ARE) normal_rythm AND Occupational_therapy IS(ARE) present THEN GLASGOW_OUTCOME_SCALE IS **3**

RULE 13
IF Diabetes IS(ARE) absent AND Hypercholesterolemia IS(ARE) absent AND Obesity IS(ARE) absent AND Transient_ischemic_attack IS(ARE) absent AND Heart_action IS(ARE) atrial_fibrillation AND Anticoagulants IS(ARE) absent THEN GLASGOW_OUTCOME_SCALE IS **3**

RULE 14
IF Admission_diagnosis IS(ARE) subarachnoid_hemorrhage AND Age IS(ARE) 17..67 AND State_at_admission IS(ARE) consciousness_disturbances THEN GLASGOW_OUTCOME_SCALE IS **3**

RULE 15
IF Age IS(ARE) 17..67 AND Stroke_type IS(ARE) LACS AND Speech_disorders IS(ARE) present THEN GLASGOW_OUTCOME_SCALE IS **3**

Table 4. (*continued*)

RULE 16
IF Gender IS(ARE) m AND Atrial_fibrilation IS(ARE) absent AND Alcohol_addiction IS(ARE) absent AND Heart_action IS(ARE) normal_rythm AND Speech_disorders IS(ARE) present AND Swallowing_disorders IS(ARE) absent AND Occupational_therapy IS(ARE) absent THEN GLASGOW_OUTCOME_SCALE IS **3**

RULE 17
IF Organic_heart_disease IS(ARE) present AND Circulatory_insufficiency IS(ARE) absent THEN GLASGOW_OUTCOME_SCALE IS **4**

RULE 18
IF Time_spent_in_hospital IS(ARE) 3..37[days] AND A11 IS(ARE) sale_udarowe AND State_at_admission IS(ARE) conscious THEN GLASGOW_OUTCOME_SCALE IS **4**

RULE 19
IF Time_elapsed IS(ARE) less_than_1_hour AND Swallowing_disorders IS(ARE) absent THEN GLASGOW_OUTCOME_SCALE IS **4**

RULE 20
IF Heart_action IS(ARE) other_dysrythmia AND Swallowing_disorders IS(ARE) absent THEN GLASGOW_OUTCOME_SCALE IS **4**

RULE 21
IF Atrial_fibrilation IS(ARE) absent AND Diabetes IS(ARE) present AND Obesity IS(ARE) absent AND General_state_at_admission IS(ARE) staying_in_bed_conscious THEN GLASGOW_OUTCOME_SCALE IS **4**

RULE 22
IF Final_diagnosis IS(ARE) cerebral_infarction AND Time_elapsed IS(ARE) 2-3_days AND Alcohol_addiction IS(ARE) absent AND Consciousness_disorders IS(ARE) absent THEN GLASGOW_OUTCOME_SCALE IS **4**

RULE 23
IF Final_diagnosis IS(ARE) cerebral_infarction AND Past_stroke IS(ARE) present AND Consciousness_disorders IS(ARE) absent AND Anticoagulants IS(ARE) absent AND Occupational_therapy IS(ARE) absent THEN GLASGOW_OUTCOME_SCALE IS **4**

RULE 24
IF Final_diagnosis IS(ARE) cerebral_infarction AND Abode IS(ARE) town AND Ischemic_heart_disease IS(ARE) absent AND Transient_ischemic_attack IS(ARE) absent AND Stroke_type IS(ARE) PACS THEN GLASGOW_OUTCOME_SCALE IS **4**

RULE 25
IF Final_diagnosis IS(ARE) cerebral_infarction AND Organic_heart_disease IS(ARE) absent AND Swallowing_disorders IS(ARE) absent AND Exercise_therapy IS(ARE) absent THEN GLASGOW_OUTCOME_SCALE IS **4**

RULE 26
IF Abode IS(ARE) miasTHEN AND Time_elapsed IS(ARE) 3-6_hours AND Obesity IS(ARE) absent AND Past_stroke IS(ARE) absent AND Pulse IS(ARE) 55..100 AND General_state_at_admission IS(ARE) staying_in_bed_conscious AND Antibiotics IS(ARE) absent THEN GLASGOW_OUTCOME_SCALE IS **4**

RULE 27
IF Admission_diagnosis IS(ARE) intracerebral_hemorrhage THEN GLASGOW_OUTCOME_SCALE IS **5**

RULE 28
IF State_at_admission IS(ARE) consciousness_disturbances AND Anticoagulants IS(ARE) absent AND Anti-edematous_agents IS(ARE) present AND Exercise_therapy IS(ARE) present THEN GLASGOW_OUTCOME_SCALE IS **5**

RULE 29
IF Body_temperature IS(ARE) 36..37.6[C] AND State_at_admission IS(ARE) conscious AND Anti-edematous_agents IS(ARE) absent AND Exercise_therapy IS(ARE) absent THEN GLASGOW_OUTCOME_SCALE IS **5**

RULE 30
IF Patient_cure_location IS(ARE) neurology_ward AND Pulse IS(ARE) 100..180 AND Anticoagulants IS(ARE) absent THEN GLASGOW_OUTCOME_SCALE IS **5**

RULE 31
IF Time_elapsed IS(ARE) 2-3_days AND Past_cardiac_infarct IS(ARE) absent AND State_at_admission IS(ARE) conscious THEN GLASGOW_OUTCOME_SCALE IS **5**

RULE 32
IF Abode IS(ARE) wieś AND Time_spent_in_hospital IS(ARE) 3..37[days] AND Past_cardiac_infarct IS(ARE) absent AND Organic_heart_disease IS(ARE) absent AND Past_stroke IS(ARE) absent AND State_at_admission IS(ARE) conscious AND Consciousness_disorders IS(ARE) absent THEN GLASGOW_OUTCOME_SCALE IS **5**

RULE 33
IF Atrial_fibrilation IS(ARE) present AND Organic_heart_disease IS(ARE) absent AND Past_stroke IS(ARE) absent AND Pulse IS(ARE) 55..100 AND Consciousness_disorders IS(ARE) absent THEN GLASGOW_OUTCOME_SCALE IS **5**

Table 4. (*continued*)

RULE 34
IF Heart_action IS(ARE) normal_rythm AND Speech_disorders IS(ARE) absent AND Occupational_therapy IS(ARE) present THEN GLASGOW_OUTCOME_SCALE IS **5**

RULE 35
IF Diabetes IS(ARE) absent AND Obesity IS(ARE) present AND Consciousness_disorders IS(ARE) absent AND Speech_therapy IS(ARE) absent THEN GLASGOW_OUTCOME_SCALE IS **5**

RULE 36
IF Nicotine_addicti IS(ARE) absent AND State_at_admission IS(ARE) conscious AND Consciousness_disorders IS(ARE) present AND Anti-edematous_agents IS(ARE) absent AND Occupational_therapy IS(ARE) absent THEN GLASGOW_OUTCOME_SCALE IS **5**

RULE 37
IF Time_spent_in_hospital IS(ARE) 3..37[days] AND Diabetes IS(ARE) absent AND Past_stroke IS(ARE) absent AND Systolic_pressure IS 90..200 AND Diastolic_pressure IS(ARE) 95..180 AND Heart_action IS(ARE) normal_rythm AND State_at_admission IS(ARE) conscious THEN GLASGOW_OUTCOME_SCALE IS(ARE) **5**

RULE 38
IF Diabetes IS(ARE) absent AND Past_stroke IS(ARE) absent AND Alcohol_addiction IS(ARE) absent AND SysTHENlic_pressure IS(ARE) 90..200 AND Heart_action IS(ARE) normal_rythm AND Consciousness_disorders IS(ARE) absent AND Speech_therapy IS(ARE) present AND Occupational_therapy IS(ARE) absent THEN GLASGOW_OUTCOME_SCALE IS **5**

RULE 39
IF Diabetes IS(ARE) absent AND Stroke_type IS(ARE) POCS AND Consciousness_disorders IS(ARE) absent THEN GLASGOW_OUTCOME_SCALE IS **5**

RULE 40
IF Ischemic_heart_disease IS(ARE) absent AND Obesity IS(ARE) present AND Past_stroke IS(ARE) absent AND Heart_action IS(ARE) normal_rythm AND Consciousness_disorders IS(ARE) absent AND Swallowing_disorders IS(ARE) absent THEN GLASGOW_OUTCOME_SCALE IS **5**

RULE 41
IF Final_diagnosis IS(ARE) other_cerebrovascular_diseases AND Speech_disorders IS(ARE) absent THEN GLASGOW_OUTCOME_SCALE IS **5**

Table 5. GOS parameters applied in the final (optimized) learning model

Parameter	Frequency quotient
State_at_admission	13/41
Consciousness_disorders	12/41
Heart_action	11/41
Exercise_therapy	9/41
Anti-edematous_agents	8/41
Occupational_therapy	7/41
Past_stroke	7/41
Diabetes	6/41
Swallowing_disorders	6/41
Body_temperature	5/41
Final_diagnosis	5/41
Obesity	5/41
Stroke_type	5/41
Time_elapsed	5/41
Anticoagulants	4/41
Atrial_fibrilation	4/41

Table 5. (*continued*)

Parameter	Frequency quotient
Organic_heart_disease	4/41
Speech_disorders	4/41
Abode	3/41
Admission_diagnosis	3/41
Alcohol_addiction	3/41
Pulse	3/41
Time_spent_in_hospital	3/41
Transient_ischemic_attack	3/41
Age	2/41
Antibiotics	2/41
Circulatory_insufficiency	2/41
General_state_at_admission	2/41
Hypercholesterolemia	2/41
Ischemic_heart_disease	2/41
Nicotine_addiction	2/41
Past_cardiac_infarct	2/41
Patient_cure_location	2/41
Speech_therapy	2/41
Systolic_pressure	2/41
Diastolic_pressure	1/41
Gender	1/41

5 Discussion and Conclusions

In the Table 3 five most important descriptive parameters of the *Glasgow Outcome Scale* are presented. All of them (*Anti-edematous_agents, Body_temperature, Exercise_therapy, Occupational_therapy, State_at_admission*), seem to be equally important in relation to the investigated database. The entry, say *Body_temperature* 16/16, informs that in the set of 16 rules this parameter was included 16 times. However, the error rate for the discussed set of rules is not acceptable.

More important and interesting is the improved set of 41 rules. Having extremely small error rate (≈3%), it shows that from 42 descriptive parameters used in the investigated database, just 5 from them were never accessed in the process of inductive generating of rules. These unnecessary parameters not needed from the point of view of classification (at least, within a group of investigated patients) are: *Arterial_hypertension, Infection_in_a_week_to_stroke, Diastolic_pressure, Antihypertensives, Neuroprotective_agents.* Further inspection of Table 5 pointed out that parameters having the larger frequency quotient (# of entries / # of rules, the rightmost column in Table 5), say for example 13/41, should be diagnosed very carefully. On the other hand, it is to note that the *Age* and *Gender* of a patient seem to have rather minor influence on the outcome of Glasgow Scale.

The future research should be oriented towards simplification of the discussed descriptive parameters, looking for the compromise of a low error rate and high effectiveness of the Glasgow Outcome Scale.

Acknowledgments

Authors thank Dr. A. Strepikowska and Dr. A. Tutaj for their valuable help and consultation in this research and preparation of the paper.

References

1. Gorelick, P.B., Atler, M.: The prevention of stroke. CRC Press, Boca Raton (2002)
2. Diener, H.C., Forsting, M.: Udar mózgu. Elsevier-Urban & Partner, Wrocław (2004) (in Polish)
3. Barnett, H.J.M., Bogousslavsky, J., Meldrum, J.H.: Advances in neurology; vol 92 ischemic stroke. Lippincott Williams & Wilkins, Philadelphia (2003)
4. Jennet, B., Bond, M.: Assessment of outcome after severe brain damage: A practical scale. Lancet 1, 480–484 (1975)
5. Kasner, S.E.: Clinical interpretation and use of stroke scales. The Lancet Neurology 5(7), 603–612 (2006)
6. Mroczek, T., Grzymała-Busse, J.W., Hippe, Z.S., Paja, W., Buciński, A., Strepikowska, A., Tutaj, A.: Informational database on brain strokes: validation of the Glasgow outcome scale and Rankin scale. In: Proc. 5th Conference on Databases in Research: Bases-Systems-Applications, Sopot, Poland, pp. 127–131 (2008) (in Polish)
7. Mroczek, T., Grzymała-Busse, J.W., Hippe, Z.S.: Rules from belief networks: A rough set approach. In: Tsumoto, S., Słowiński, R., Komorowski, J., Grzymała-Busse, J.W. (eds.) RSCTC 2004. LNCS (LNAI), vol. 3066, pp. 483–487. Springer, Heidelberg (2004)
8. Hippe, Z.S., et al.: A new version of a data mining system powered by an efficient recursive covering algorithm. UITM internal research report (2008) (in Polish)
9. Paja, W.: Design of optimum learning models using secondary source of knowledge. PhD dissertation, AGH University of Science and Technology, Cracow (2008) (in Polish)
10. Grzymała-Busse, J.W.: A new version of the rule induction system LERS. Fundamenta Informaticae 31, 27–39 (1997)
11. Pawlak, Z.: Rough Sets. Intern. J. Comp. Inf. Sci. 11, 341–356 (1982)
12. Grzymała-Busse, J.W., Hippe, Z.S., Mroczek, T., Buciński, A., Strepikowska, A., Tutaj, A.: Prediction of severe brain damage outcome using two data mining methods. In: Proc. IEEE Conference on Human System Interaction, Cracow, Poland, pp. 585–590 (2008)

Segmentation of Anatomical Structure by Using a Local Classifier Derived from Neighborhood Information

S. Takemoto[1], H. Yokota[1,2], T. Mishima[3], and R. Himeno[2]

[1] Bio-research Infrastructure Construction Team, VCAD System Research Program, RIKEN, Saitama, Japan
{satoko-t,hyokota}@riken.jp

[2] Living Matter Simulation Research Team, RIKEN, Saitama, Japan
himeno@riken.jp

[3] Department of Information and Computer Sciences, Saitama University, Saitama, Japan
mishima@me.ics.saitama-u.ac.jp

Abstract. Rapid advances in imaging modalities have increased the importance of image segmentation techniques. Image segmentation is a process that divides an image into regions based on the image's internal components to distinguish between the component of interest and other components. We use this process to analyze the region of the component of interest and acquire more detailed quantitative data about the component. However, almost all processes of segmentation of anatomical structures have inherent problems such as the presence of image artifacts and the need for complex parameter settings. Here, we present a framework for a semi-automatic segmentation technique that incorporates a local classifier derived from a neighboring image. By using the local classifier, we were able to consider otherwise challenging cases of segmentation merely as two-class classifications without any complicated parameters. Our method is simple to implement and easy to operate. We successfully tested the method on computed tomography images.

1 Introduction

Modern imaging modalities such as computed tomography (CT) and magnetic resonance imaging (MRI) readily yield detailed information on anatomical structures such as organs. Many scientific disciplines rely not merely on image observation but also on the quantification of various characteristics of organs and other regions of interest. Image segmentation [7, 13] can play an important role in the quantification of imaging data. Image segmentation is a process that divides an image into regions based on the image's internal components to distinguish between the component of interest and other components.

Segmentation of anatomical structures is, however, frequently complicated by image artifacts such as noise and partial-volume effects [3] and by the diverse geometry of the segmentation target. Parametric deformable models such as snakes

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 171–180.
springerlink.com

[6, 8, 14, 15] and geometric deformable models such as the level-set method [2, 4, 6, 9], and the graph-cut algorithm [1] are widely used and yield robust segmentation results under extreme conditions. However, one drawback of these methods is that they require complicated parameter settings to deliver the high performance needed to acquire highly accurate segmentation results.

To overcome these drawbacks, we proposed a new framework of anatomical structure segmentation that is simple to implement and useful. Our method offers semi-automatic segmentation of the target structure from 3D volume images composed of sequential 2D images (Fig. 1). Our method requires only a seed region, which indicates the region of the segmentation target and is present in at least one image. After defining the seed, we can detect the target region in each subsequent image by using a local classifier that has been trained by the seed image, or the already segmented neighboring image that has been yielded by the seed image. Another key advantage is that use of the local classifier makes our method very competitive in terms of computational cost.

In Section 2, we briefly describe the proposed framework and its methodology. Experimental results and an evaluation of the proposed framework on test CT images with ground truth are presented in Section 3. Finally, a conclusion is offered in Section 4.

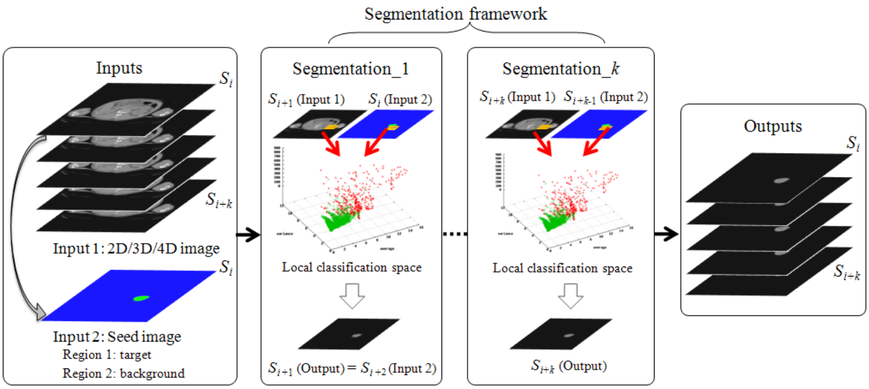


Fig. 1. Brief overview of our segmentation framework: one seed image showing the region of the segmentation target is used for segmentation of the entire region from the sequential images

2 Segmentation Framework with a Local Classifier

The framework we present here uses a simple and robust semi-automatic technique to segment the target anatomical structure present on sequential 2D images (i.e., a 3D image). First, we define one seed region that represents the target on one slice (Fig. 1). After the seed region has been defined, the class to which each pixel belongs – either “target” or “background” – is calculated automatically from all sequential images, in sequential order. The calculation used to classify each pixel is performed with our proposed local classifier. The classification procedure used is as follows:

1. Manually set the target region R_i as the seed in the starting image S_i.
2. Copy R_i into the neighboring image S_{i+1}. Call the copied region R'_{i+1}.
3. Detect the boundary pixels inside R'_{i+1}.
4. Store the detected pixels in the transient data queue (TDQ).
5. Retrieve one boundary pixel P_n from the TDQ and determine the class of P_n by using the local classifier (described later).
6. Update the boundary of R'_{i+1} according to the result of classification of P_n and store the newly detected boundary pixel in the TDQ.
7. Repeat steps 5 and 6 until the TDQ becomes empty.
8. Let $R_{i+1} \leftarrow R'_{i+1}$. Then, R_{i+1} represents the target region in S_{i+1}.
9. Let $i \leftarrow i+1$.
10. Repeat steps 2 through 9 as long as the target region exists in the processing image.

Because of introduction of the TDQ into our framework, no end condition is needed to judge whether the segmentation process is finished with their calculation in each image. The empty TDQ means that there is no pixel that needs to be classified in each image. This is an effective way of improving the segmentation accuracy. This is because we consider that minimizing the arbitrary conditions applied is an effective way of achieving robust segmentation. Note that the information on the images where the segmentation target exists is defined manually. In the above procedure, P_n is classified in step 5 by using the local classifier as follows:

I. Define a local region $L^i_{P_n}$ enclosing P'_n in the neighboring image S_i. P'_n has the same (x, y) coordinates as does P_n.
II. Train a classifier by using the segmentation results of all the pixels inside $L^i_{P_n}$.
III. Assign the class of P_n according to the results of classification with the trained classifier.

Figure 2 is a schematic diagram of the steps involved in defining the training region for one local classifier. We predefine a window size of $L^i_{P_n}$ according to the size of the target, so as to produce the situation of the two classes' existence inside $L^i_{P_n}$. The classifier is generated by use of a pattern classification technique (e.g., k-nearest neighbor algorithm, support vector machine [11]). The features used for training the classifier can be related to intensity, texture, or other image properties.

The main advantage of our method is that it constructs one classifier at every pixel near the boundary of the target structure. Because the association between

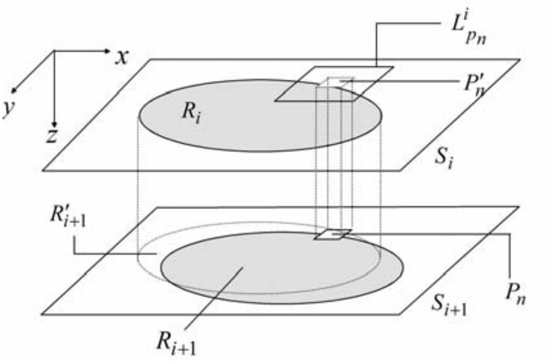


Fig. 2. Schematic diagram of definition of a training region $L^i_{P_n}$ for a local classifier

features and class has diversity at each position, use of the classifier generated by using neighboring image's information in each case is likely to improve the accuracy of segmentation. In addition, since the local classifier is generated only near the boundary, we can reduce the computational cost.

Note that our framework requires a high z-resolution to prevent existence of an excessive bias between the two classes in each local classifier. That is, if the z-resolution is too low, the movement of a target region between neighboring images may become too large. In that case, users should set a larger window size to maximize the reliability of the segmentation results. A flexible window size allow us to avoid generating biased distribution of the two classes inside each $L^i_{P_n}$. As a result, we can reduce the possibility that each local classifier produces a misclassified result.

3 Experimental Analyses and Discussion

3.1 Advantages of Our Local Classifier

Generally, the geometric diversity of anatomical structures complicates their segmentation. Our method is able to overcome this difficulty simply by approaching the segmentation problem as a two-class classification according to the framework described in Section 2. The reason for using the neighboring segmented results as a training dataset for the classifier is to avoid making assumptions in terms of feature distribution. Frequently, image features such as pixel intensity are assumed to be derived from a mixture of probability distributions (usually Gaussian). In the case of multi-class segmentation, the Gaussian mixture model (GMM) [5] is often used to represent feature distributions. However, in many cases, because of the influence of imaging factors such as noise, the GMM does not quite correspond to the actual feature distribution; its application therefore results in declining segmentation performance. In contrast, use of our local classifier achieves more robust segmentation than that obtained with the GMM because we make no assumptions as to feature distribution. That is, our local classifier can be regarded as a nonparametric classifier.

Figure 3 compares the classification results obtained by using the GMM and a method without any assumptions—that is, a nonparametric classifier. The objective of this experiment was to classify the test image shown in Figure 3 (a) into two classes according to internal components, as for the neighboring image (b) that was manually provided. The test image is an 8-bit grayscale CT image. As a set of training data, we used intensity values from the whole image. In the case of the GMM, to estimate parameters to fit a Gaussian distribution, we used the expectation–maximization (EM) algorithm [5, 10]. The EM is a well-established maximum likelihood algorithm used to fit a mixture model to a set of training data.

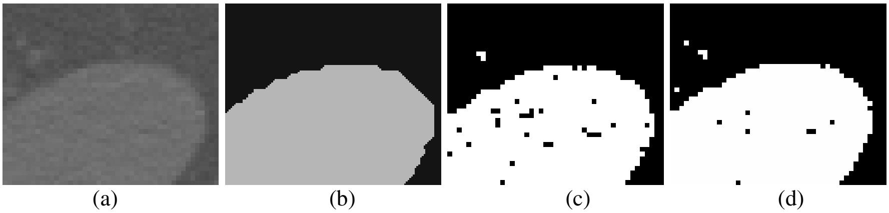


Fig. 3. The test image (a) and the trained image (b): images resulting from the GMM (c) and ANN (d)

In the case of the nonparametric classifier, we used the approximate nearest neighbors (ANN) method [12], a kind of k-nearest neighbor method. Because ANN requires pre-defined training data, we performed a classification by using the trained image, which had already been classified (Fig. 3(b)). Figures 3(c) and 3(d) show the results of classification by the GMM and ANN, respectively. It is clear that under-segmentation has occurred in the gray region in Figure 3(c). These results indicate that the criterion derived by using ANN achieved a more accurate classification than did that derived with the GMM. That is, even if the target image has the simple components as shown in Fig. 3(a), the distribution of image features does not always fit a Gaussian distribution. Therefore, we consider that our method will perform better with the nonparametric classifier, which represented the real image features well and without any assumptions in terms of the two-class classification.

3.2 Segmentation on Test CT Images

We implemented our segmentation framework and tested it on CT images of the human abdomen. The goal of segmentation was to extract only the kidney region from the sequential 2D images. The kidney region was extracted after we had defined a seed from a single image (Fig. 4). In this test, we used the ANN classifier as the local classifier trained by the intensity value from each 7×7 pixel local region. The number of nearest neighbors referenced for the classification was 11. For comparison, we also tested another segmentation method with the GMM classifier. Figure 5 shows the segmentation results. In each resulting image, a black line indicates the boundary of the extracted region. With our proposed method, each target region in each image was segmented within 20 s on a personal computer (Intel Pentium 4, 2.53-GHz) with 2 GB of memory.

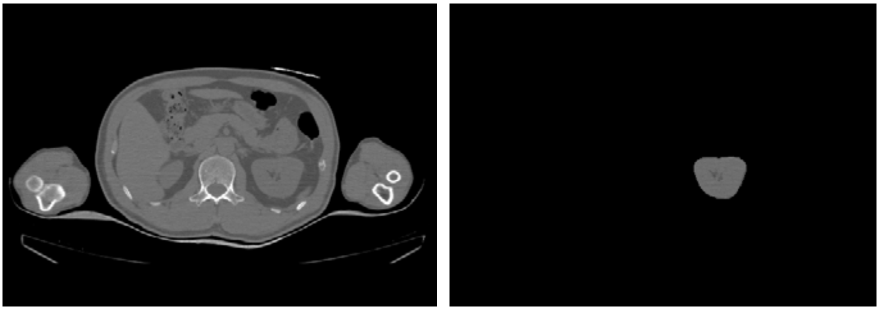


Fig. 4. The test image (left) and the seed image (right): in the image at right, the non-black region indicates the kidney

On reviewing the segmentation results shown in Figure 5, we noted that a few regions were segmented incorrectly with our classifier, whereas many regions were segmented incorrectly with the GMM classifier, especially in the later images. To analyze this phenomenon quantitatively, we calculated the accurate segmentation ratio (ASR), as follows:

$$ASR\,(\%) = \frac{region_{seg} \cap region_{true}}{region_{seg} \cup region_{true}} \times 100 \tag{1}$$

where $region_{seg}$ represents the region extracted by using our segmentation framework, and $region_{true}$ represents the true kidney region extracted manually. According to equation (1), a perfect segmentation result has an ASR of 100%. We show the ASR results in Figure 6. In both cases, the ASR of segmentation dropped rapidly in later images because there was an excessive movement between neighboring images. This degradation was due to the propagation of error occurring in the neighboring image; the error spread gradually through the sequential images. However, it should be noted that, with our classifier, segmentation accuracy peaked at approximately 90% for many images and at about 80% overall throughout almost the entire image sequence. In comparison, use of the GMM classifier led to a peak accuracy of about 70% and about 50% overall accuracy. Therefore, the local classifier that had avoided predefined assumptions regarding distribution was effective in robust segmentation.

Our segmentation framework enables the user to detect false segmentation clearly and directly on the computer screen. If the user detects false segmentation, he or she can simply set a new seed image. This resetting of the seed improves the accuracy of segmentation.

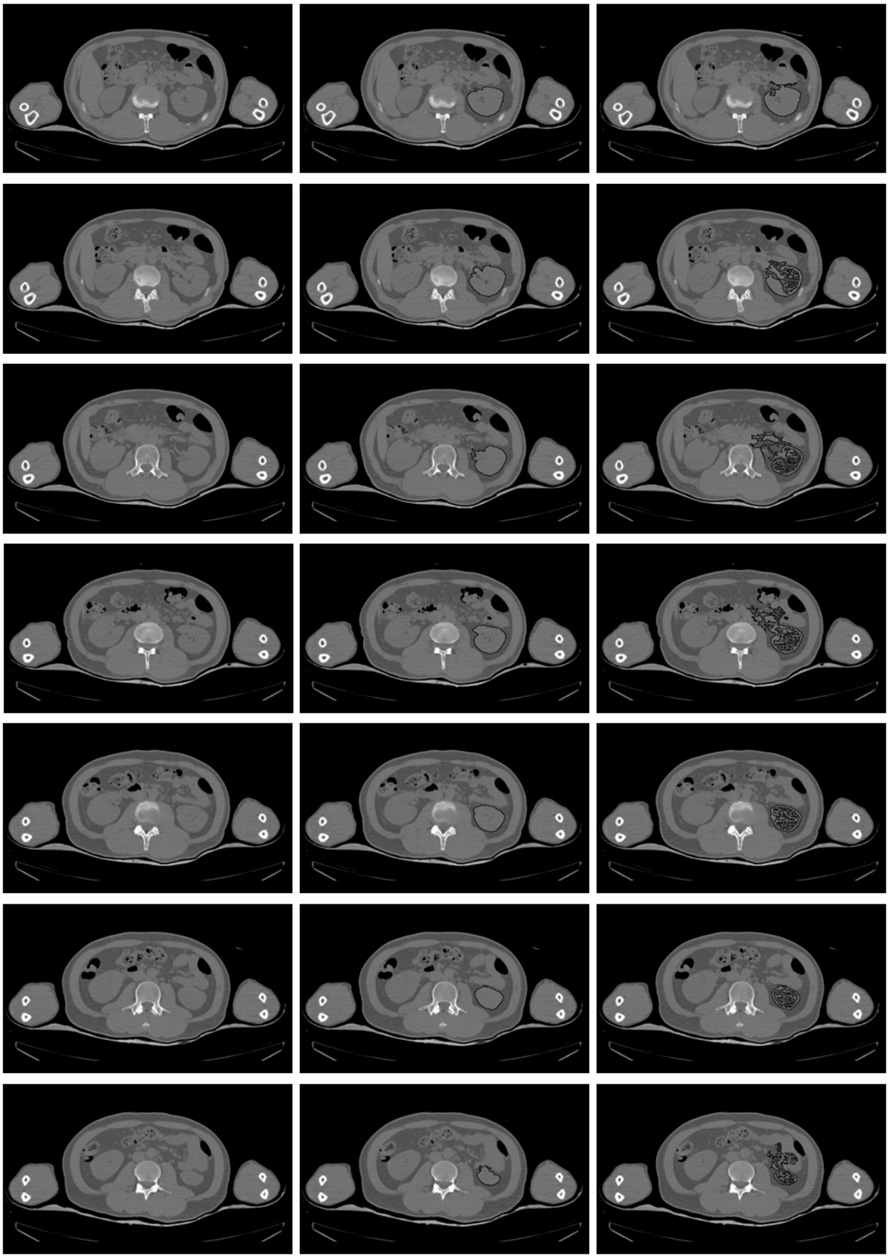


Fig. 5. Test abdominal images (left column). Images generated by using our proposed framework (center column). Images generated by using the GMM classifier (right column). Every fifth image is shown.

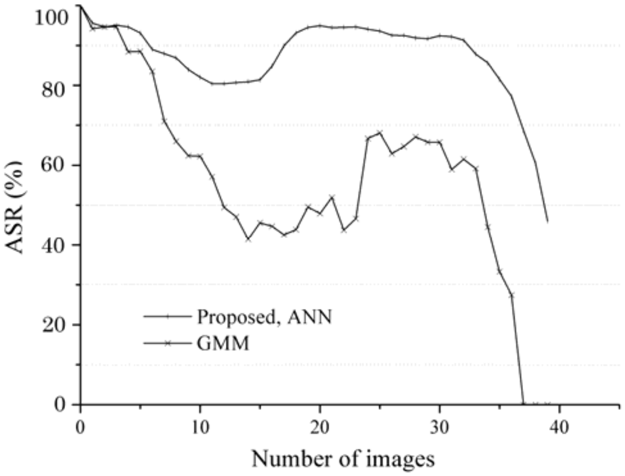


Fig. 6. Plot of accurate segmentation ratios of 40 segmented regions in the test images

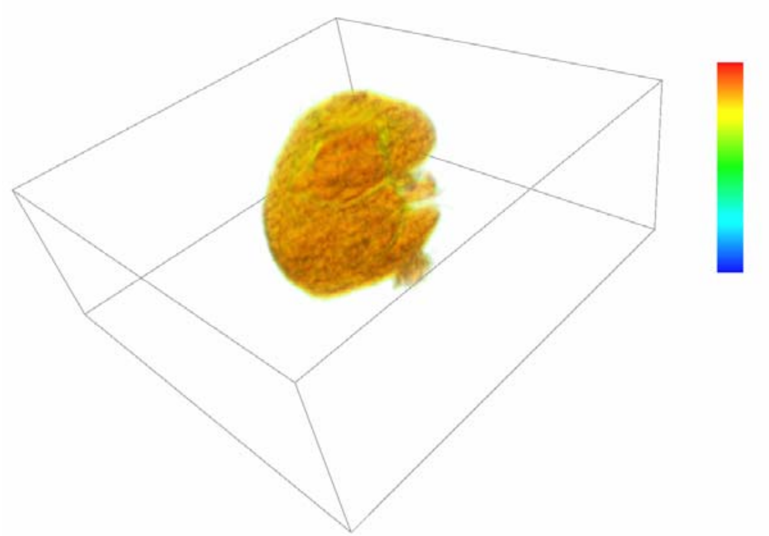


Fig. 7. Three-dimensional kidney model reconstructed by using the volume-rendering method which is equipped in the Volume CAD (VCAD) tools [16]. The color bar means degree of 8-bit pixel intensity.

3.3 Deliverables of Successful Segmentation

As we mentioned in Section 1, our segmentation results allow us to analyze a structure of interest, such as an organ, from various perspectives. That is because we were able to distinguish one region of interest from the others by segmentation. For example, we can use the segmentation results to show a 3D kidney image. Figure 7 is a 3D reconstruction model of the segmented kidney generated by the volume-rendering technique. By using this product, we were able to observe the kidney from many viewing directions on the computer screen. A surface model of the segmented kidney is also shown (Fig. 8). These models may be helpful in future biological simulations.

In addition, we were able to calculate the surface area, volume, and surface-to-volume ratio of the segmented kidney; the segmented kidney was 256.81 cm^2 in surface and 106.37 cm^3 in volume, and the calculated surface-to-volume ratio was 2.41. These data were easily calculated by counting the number of the segmented pixels. We therefore consider that successful segmentation of anatomical structures will yield new insights into biological studies, such as functional anatomy and anatomical physiology.

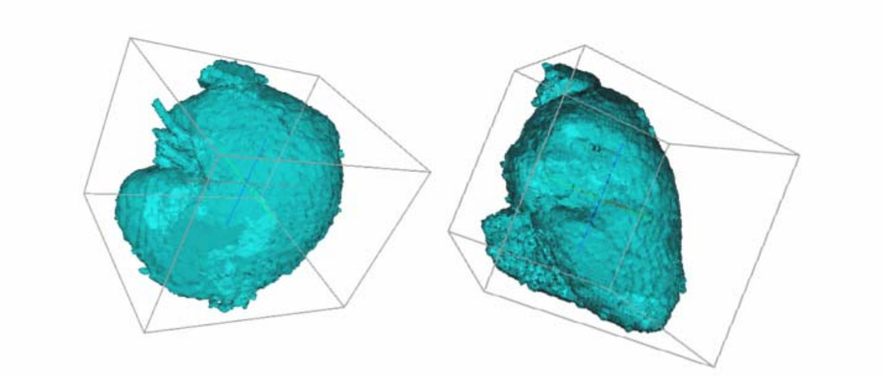


Fig. 8. Surface model of the segmented kidney represented by using the surface-rendering method which is equipped in the VCAD tools

4 Conclusions

We developed a new segmentation framework based on the use of a local classifier trained by segmentation of previous regions. The local classifier plays an important role in simplifying a challenging segmentation task into a simple classification task. Our framework enables users to segment anatomical structures after the seed region has been defined from only a single image. In addition, our framework is simple to implement and easy to operate because it avoids reliance on predefined parameters.

We tested our framework only on human abdominal CT images, but it is not limited to these. Note that it is preferable not to have a branched structure as a segmentation target; if the target is a branched structure, it will be necessary to devise measures for defining multiple seed regions. Although we used only intensity value as an image feature here, the method could be improved by incorporating additional features or higher level classifiers (e.g., texture information in multi-scale space or support vector machines).

Acknowledgments

This work was supported by Grants-in-Aid for Scientific Research of Japan (Young Researcher, Start-up: 19800062).

References

1. Boykov, Y., Funka-Lea, G.: Graph cuts and efficient N-D image segmentation. Intern. J. Computer Vision 70(2), 109–131 (2006)
2. Yan, P., Shen, W., Kassim, A.A., Shah, M.: Segmentation of neighboring organs in medical image with model competition. In: Proc. Medical Image Computing and Computer-Assisted Intervention Conference, pp. 270–277 (2005)
3. Pham, D.L., Bazin, P.L.: Simultaneous boundary and partial volume estimation in medical images. In: Proc. Medical Image Computing and Computer-Assisted Intervention Conference, pp. 119–126 (2004)
4. Qu, Y., Chen, Q., Heng, P.A., Wong, T.T.: Segmentation of left ventricle via level-set method based on enriched speed term. In: Proc. Medical Image Computing and Computer-Assisted Intervention Conference, pp. 435–442 (2004)
5. Carson, C., Belongie, S., Greenspan, H., Malik, J.: Blobworld: image segmentation using expectation-maximization and its application to image querying. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(8), 1026–1038 (2002)
6. Xu, C., Pham, D., Prince, J.: Image segmentation using deformable models. In: Handbook of Medical Imaging. Medical Image Processing and Analysis, vol. 2, pp. 175–272 (2000)
7. Lakere, S., Kaufman, A.: 3D segmentation techniques for medical volumes. Technical Report, State University of New York, New York (2000)
8. McInerney, T., Terzopoulos, D.: T-snakes: Topology adaptive snakes. Medical Image Analysis 4(2), 73–91 (2000)
9. Malladi, R., Sethian, J.A., Vermuri, B.C.: Shape modeling with front propagation: a level set approach. IEEE Transactions on Pattern Analysis and Machine Intelligence 17, 158–174 (1995)
10. Akaho, S.: The EM algorithm for multiple object recognition. In: Proc. Internat. Conference on Neural Network, pp. 2426–2431 (1995)
11. Vapnik, V.N.: The nature of statistical learning theory. Springer, New York (1995)
12. Arya, S., Mount, D.M., Netanyahu, N., Silverman, R., Wu, A.Y.: An optimal algorithm for approximate nearest neighbor searching in fixed dimensions. In: Proc. 5th ACM-SIAM Symposium on Discrete Algorithms, pp. 573–582 (1994)
13. Pal, N.R., Pal, S.K.: A review of image segmentation techniques. Pattern Recognition 26(9), 1277–1294 (1993)
14. Cohen, L.D.: On active contour models and ballons. Comp. Vision, Graphics, and Image Processing: Image Understanding 53(2), 211–218 (1991)
15. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: active contour models. Computer Vision, 321–331 (1988)
16. `http://vcad-hpsv.riken.jp/en/` (accessed April 29, 2009)

An Application of Detection Function for the Eye Blinking Detection

T. Pander, T. Przybyła, and R. Czabański

Division of Biomedical Electronics, Institute of Electronics,
Silesian University of Technology 44-100 Gliwice, Poland
Tomasz.Pander@polsl.pl

Abstract. The electrooculogram represents the electrical activity of muscles that control movements of eyes. The eye blinking is a natural protection system which defends the eye from environmental exposure. The spontaneous eye blink is considered to be a suitable indicator for fatigue diagnostics during many, different tasks of human being activity. The action of eye blinks covers a specific range of frequency and for that reason it is possible to construct a function which processes the signal and generates an artificial peak when blink occurs. This function is called the detection function. This function is used to detect the spontaneous eye blink action. Nonlinear and linear signal processing methods are applied to obtain the detection function waveform. On this base the position of an eye blink is estimated. The results demonstrate that the measurement of an eye blink parameter provides reliable information for eye-controlled systems from human-machine interface.

1 Introduction

Communication between people seems to be much more simple than between a man and a computer machine. This difficulty increases when a person is disabled. The eye movement can be applied as a mean to communicate with a computer. The idea of such interface is the following. The human-machine interface device (one of the important components is a digital signal processor - DSP) records the electrooculographic (EOG) signal and then EOG is processed in DSP. And in the end, the device generates steering signals for a computer.

The blinking can be applied to a human-machine interface as an indicator of some typical signaling methods. Such system may be used for detecting basic commands to control some instruments, robots or any other applications by people with limited upper body mobility [3]. The most typical situation is a control of applications during working with a computer.

The EOG signal is based on electrical measurement of the potential difference between the cornea and the retina. The cornea-retinal potential creates an electrical field in the front of a head [15]. This field changes in orientation as the eyeballs

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 181–191.
springerlink.com

rotate. The amplitude of EOG signal varies from 50 to 3500 μV with a frequency range of about DC-100 Hz. Its behavior is practically linear for gaze angles of ±30° [3]. It should be pointed out here that the variables measured in the human body (any biopotentials) are almost always recorded with a noise and often have non-stationary features. Their magnitude varies with time, even when all possible variables are under control. This means that the variability of the EOG signals depend on many factors that are difficult to determine [3, 15].

Eye blinks and eyes movement are "windows" across a man can understand an surrounding world. Eye blinking is the contraction of sets of muscles of eye and produces an electrical activation of eyelid's muscles. Durations of such signals last for a fraction of a second [1]. The eye blink mechanism has several effects. The eye blinking can be divided into reflex blink (in response to something invading in the eye, this type of blink is instinctive response that protects the eye against air puffs and dust, this is also part of scared response to loud noises), voluntary blink (as a result of a decision to blink) and involuntary (spontaneous without external stimuli, probably controlled by a blink generator in the brain) [1, 14]. The spontaneous eye blink is considered to be a suitable ocular indicator for fatigue diagnostics [4, 6, 8]. An eyelid movement (blink) introduces a change in the potential distribution around the eye [14]. Spontaneous blinks are typically of a shorter duration than reflexive and voluntary blinks, and voluntary blinks show the greatest amplitude in EOG waveform [7]. Recent studies demonstrate that the analysis of spontaneous blinks may provide substantial information concerning nervous activation process and fatigue [4].

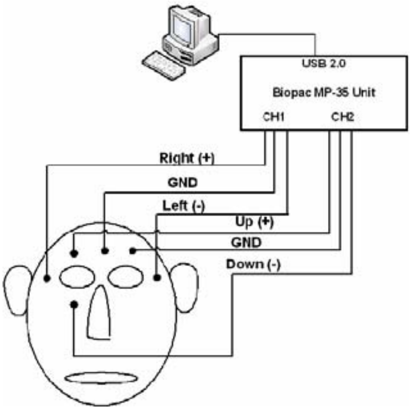


Fig. 1. Bipolar electrode placement and the measuring system connection

The EOG signal can be recorded in a horizontal and a vertical direction of eye's movements. This requires six electrodes (standard Ag/AgCl electrodes attached by means of rings of double adhesive) which are placed in the front of a human face.

In this work applied measurement system is presented in the Fig. 1. This system is based on the Biopac MP-35 unit. The sampling frequency for that work is 100 Hz.

The recognition of eye movements and its behavior can be classified in two categories: invasive methods which require direct contact with the eyes and non-invasive method which avoid any physical contact with the user. Such methods are more comfortable and safety for the user than the first group of methods [2].

The simplest method of eye blink detection is manual detection based on the direct observation of a investigated man. This method is time-consuming and is typically limited to short time intervals, however, due to the natural variability in blinking, measurements should generally last at least 5 min [14]. This method can be a reference method.

Another way of recording an eye's movements signal and eye blinks is an application based on the different reflection of the emitted infrared light from eyelid and eyeball [4, 11, 13, 14].

The goal of this paper is to employ the method of a detection function to raise the precision of detection of eye blinks and detection of single or double blinks. The rest of this paper is organized as follows: Section 2 describes the proposed method of the detection function; Section 3 presents the method of localization of blinks from the detection function waveform in the time domain and Section 4 is devoted to final conclusions.

2 The Detection Function for Eye-Blink Detection

The EOG signal can be very noisy and is non-stationary in its nature. An example of EOG signal recorded in vertical and horizontal plane is presented in the Fig. 2.

The main source of a noise in EOG signal is a signal of electrical activity of face's muscles. A movement of a head or speaking can also disturb EOG signal. The rapid change of amplitude of EOG signal (it is shown in the Fig. 2) is caused by the saccadic movement of eye's ball. Three spikes which appeared in a vertical plane are eye's blinks. Blinks of eye are very well seen in a signal which is recorded in a vertical plane. A high change in the potential distribution around the eye is a result of movements of muscles of upper and lower eyelid. This potential is greater in a vertical plane than in a horizontal plane.

An idea of the detection function comes from ECG signal processing [5, 9, 12]. The concept of such device consists of two steps. The first step is a filtering process (including robust nonlinear filtering and traditional linear filtering) and the second step is nonlinear operation (using a square or an absolute value function). Then the “description” function is made. Because EOG signal contains signal components from DC-100 Hz, it is possible to construct a cascade of digital filters to detect an eye's blink. But the range of frequency for blinks is from 1Hz to 10Hz. The purpose of the signal filtering is to attenuate noise and enhance those features of the signal used for detection as this leads to an increased probability for correct detection of blinks. The detection function is created by the device shown in the Fig. 3.

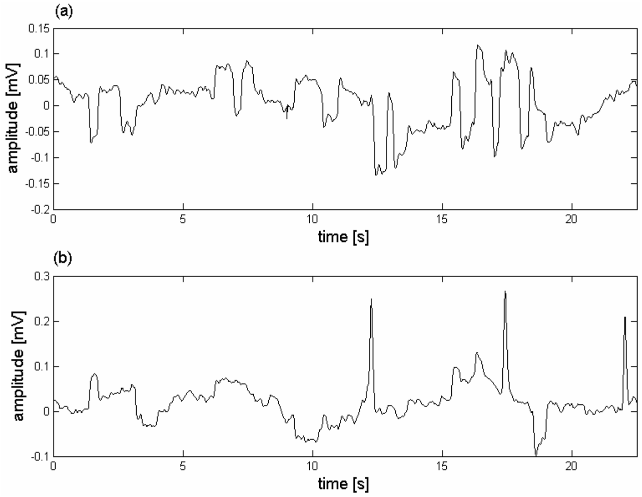


Fig. 2. The EOG signal in horizontal (a) and vertical plane (b)

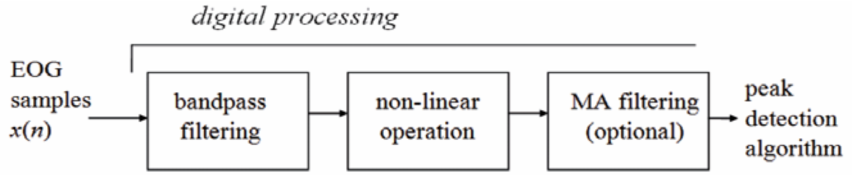


Fig. 3. A structure of device which creates the detection function

The detection function is implemented in the following way. The first step is rejection of any outliers from a signal by an application of nonlinear median filter. The length of the window of a median filter is 10 samples. The second step is the bandpass filtering realized by serial joined of three FIR filters (applying Matlab procedures):

1. the low-pass filtering is realized with the 32 order low-pass FIR filter and the cut-off frequency is 30 Hz.
2. the high-pass FIR filter and the cut-off frequency is 1.5 Hz. The Chebyshev window with 20 decibels of relative sidelobe attenuation is also used. The order of the filter is 62.
3. the low-pass FIR filtering is implemented, the order of the filter is 60, the cut-off frequency is 8 Hz, and the Chebyshev window with 20 decibels of relative sidelobe attenuation is also used.

The frequency characteristic of the proposed band-pass filter is presented in the Fig. 4.

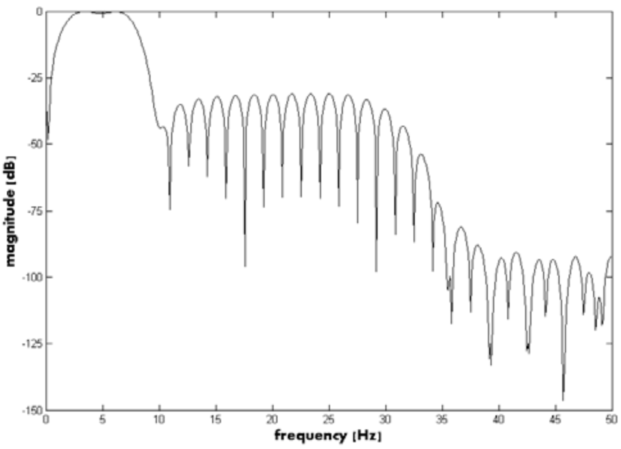


Fig. 4. The frequency characteristic of the bandpass filtering of the proposed detection function

The next step is a nonlinear operation and smoothing of a obtained signal with a moving average filter. There are two kinds of nonlinear operation which can be applied. The first one is the square function and the second is the absolute value function. These operations can be described in the following way:

$$y(n) = \frac{1}{2N+1}\sum_{i=-N}^{N}(x_f(n+i))^2 \tag{1}$$

$$y(n) = \frac{1}{2N+1}\sum_{i=-N}^{N}\left|x_f(n+i)\right| \tag{2}$$

where: $x_f(n)$ is the output signal of bandpass FIR filtering, $2N$+1 is a length of the moving average filter. In this work N=20. Such value of N guaranties that detection function has only one peak. The example of the detection function (with square and absolute value functions as a nonlinear operation) waveform in respect to EOG signal which contains blinks is presented in the Fig. 5 and in the Fig. 6.

The *AmpThreshold* level which is presented in the Fig. 5 and 6 determines the threshold level for the detection of peaks in the detection function waveform. Value of *AmpThreshold* from these figures is the same and equals 0.3. When the absolute value function is applied as the nonlinear operation (Eq. 2), then peaks generated by the detection function have higher amplitude than peaks generated by using the square function for the same value of *AmpThreshold*. The square function more effectively suppress small components in the detection function waveform and attenuate higher components, but requires smaller value of *AmpThreshold* for proper recognition of blinks than the absolute value function.

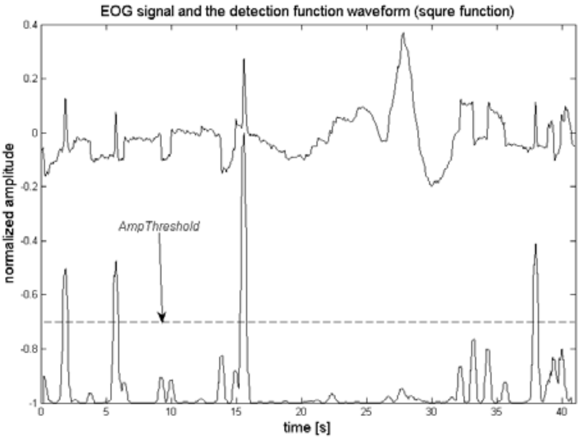


Fig. 5. The detection function waveform obtained with the square function as a nonlinear operation (lower plot) and the corresponding EOG signal (upper plot)

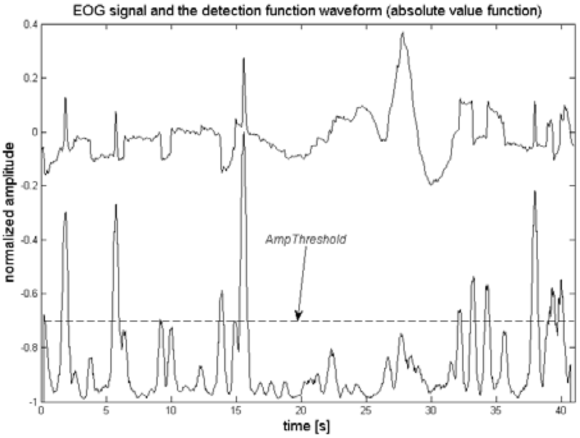


Fig. 6. The detection function waveform obtained with the absolute value function as a nonlinear operation (lower plot) and the corresponding EOG signal (upper plot)

Peaks which occur in the shape of the detection function correspond to time moments when blinks of eyelids appear. This method operates proper only when one blink occur in short time interval. In the case of appearance of two or few, fast blinks, the detection function can detect only the first blink. This situation is not an important problem, because the location of first blink is detected, and in order to detect second blink (in short time interval) another method can be used. An example of such situation is presented in the Fig. 7.

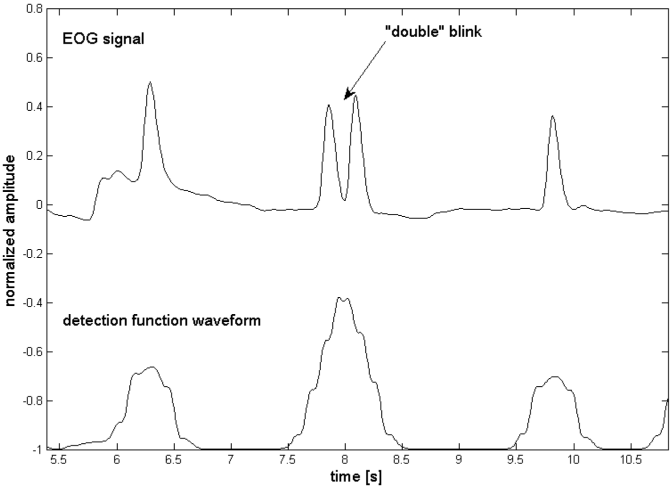


Fig. 7. Problem of detection of "double" blink in EOG signal (eyes make two fast blinks in short time interval)

3 Location of Blinks in Time Domain

The second stage of detection of blinks is a localization process in the time domain on the base of the shape of the detection function waveform. In order to make this task the method described in [10] is applied. This method allows to locate and measure the positive peaks in a noisy data sets. It detects peaks by looking for downward zero-crossings in the smoothed first derivative that exceeds some a slope of threshold and peak amplitudes that exceed given amplitude threshold, and determines the position [10]. An example of the peak detection on the base of the first derivative is presented in the Fig. 8.

Adjustable parameters allow to discriminate of blinks signal peaks from a noise or saccadic eye movements which can appear in the detection function waveform. The pseudocode of the applied algorithm is the following:

```
for i=1:length(y)
 if sign(d(i)) > sign(d(i+1))
  if d(i)-d(i+1) > SlopeThreshold*y(i)
    if y(i) > AmpThreshold
      begin
        gather points around peak
        find maximum of points
      end
    endif
  endif
 endif
end
```

where: **y** is the vector variable which represents the detection function waveform from eq. 1, **d** is the vector variable which denotes the first derivative of the detection function waveform, ***SlopeThreshold*** and ***AmpThreshold*** are adjustable parameters which values depends on the values of the detection function. When the sampling frequency equals 100 Hz, the average width of peaks in the detection function waveform ***width***=60, then

$$SlopeThreshold = 0.5 \cdot width^{(-2)} \tag{3}$$

and value of ***AmpThreshold*** depends on the kind of applied nonlinear function used in detection function waveform creation. In this work values of this parameter are equaled 0.2 for the square function or 0.3 for the absolute value function.

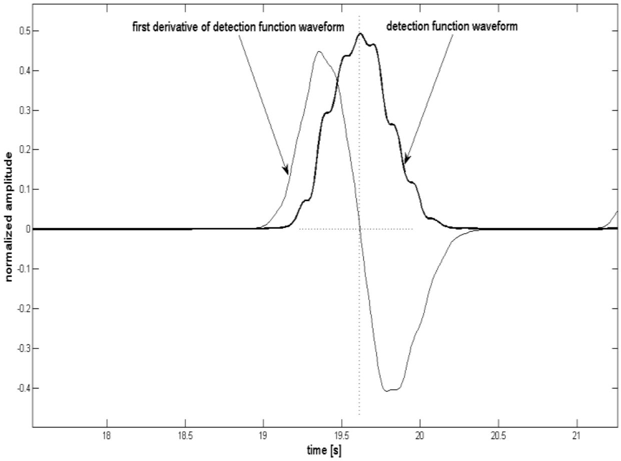


Fig. 8. The example of the detection function waveform and its corresponding the first derivative in the moment when eye blink occurs

The first derivative (**d**) calculated from the detection function waveform has many local fluctuations which can make difficult to detect peaks of the detection function waveform. This problem is presented in upper plot of the Fig. 9. In order to avoid such situation the first derivative of the detection function waveform was smoothed by MA filter. The order of this filter is equaled half of the width of a peak from detection function waveform. In this work the width of a peak is equaled 40 samples.

The detection of double blinks is performed in the following way: creation of the detection function waveform on the base of EOG signal, location of blinks in time domain and then checking does in short time interval (0.6 sec.) exist one blink or two (or more), fast blinks.

These steps are sufficient to distinguish single or double blink. The example of proper recognition of single and double blinks with presented method is shown in

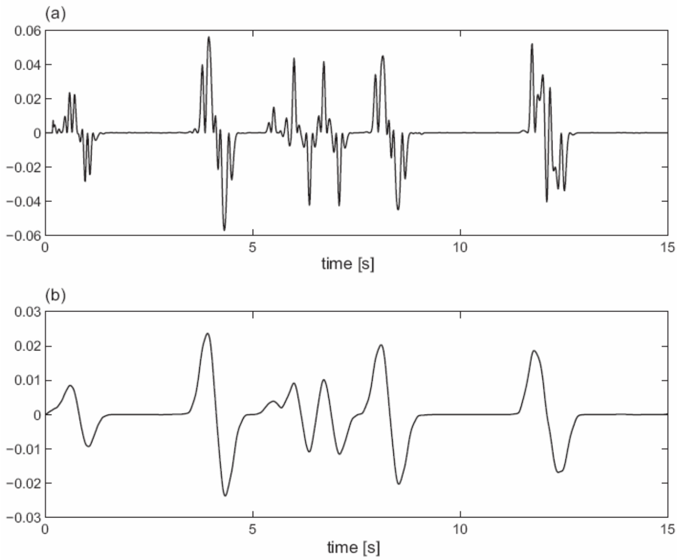


Fig. 9. The first derivative of the detection function before smoothing (a) and after smoothing (b)

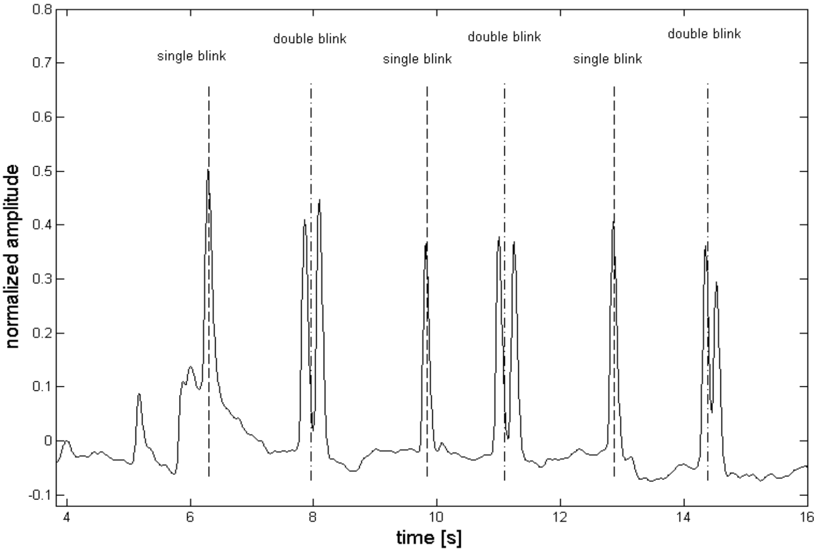


Fig. 10. Part of EOG signal with discriminated single (dashed line) and double (dot-dashed line) blinks

the Fig. 10. The EOG signal (recorded in vertical plane) presented in the Fig. 10 was recorded during the working with computer, and single and double blinks correspond to the mouse clicking in the window application.

4 Conclusions

In this paper a new method for blink detection is presented. The proposed method is based on the detection function. The main aim of the application of the detection function is suppression of noise and any other movements of eyes and enhanced blinks. The structure of detection function consists of the bandpass filter, stage of nonlinear operation and last stage is smoothing with the low-pass moving average filter. The square function or the absolute value function can be used as a nonlinear operation. But using one of these function requires choosing of an amplitude threshold level for proper detection and localization blinks in time domain. The duty of these stages is obtaining a single peak when a blink occurs. The cascade of proposed filters efficiently suppress any kind of eye's movements and amplify any type of blinks. The last stage of the creation of the detection function, when MA filter is applied, is critical. In this stage we should obtain peaks corresponding to blinks which are separated from other components of EOG signal. This process depends on the length of MA filter.

All experiments were run in the MATLAB environment.

The presented method can be applied for a human-machine interface, for example to control windows-based computer application as the virtual mouse for one or double clicking. Such approach can be very useful for people with limited upper body mobility or for testing human sight sense.

Another example of using this method is an application for accurate measurement of eye's blink parameters like blink frequency, amplitude or eyelid opening level and duration.

References

1. Akhbardeh, A., Farrokhi, M., Fakharian, A.: Voluntary and involuntary eye blinks detection using neuro-fuzzy systems and EOG signals for human-computer interface aids. In: Proc. 5th Intern. Workshop on Research and Education in Mechatronics, Kielce-Cedzyna, Poland (2004), http://www.pages.drexel.edu/~aa485/Akh-REM2004.pdf (accessed May 11, 2009)
2. Bacivarov, I., Ionita, M., Corcoran, P.: Statistical models of appearance for eye tracking and eye-blink detection and measurement. IEEE Trans on Consumer Electronics 54, 1312–1320 (2008)
3. Barea, R., Boquete, L., Mazo, M., López, E.: Wheelchair guidance strategies using EOG. J. Intel. Rob. Sys. 34, 279–299 (2002)
4. Caffier, P., Erdmann, U., Ullsperger, P.: Experimental evaluation of eye-blink parameters as a drowsiness measure. Eur. J. Appl. Physiol. 89, 319–325 (2003)
5. Frankiewicz, Z., Łęski, J.: Adaptive fiducial point detector for ECG stress testing systems. Int. J. Biomed. Comput. 28, 127–135 (1991)
6. Hammond, R.: Passive eye monitoring. Springer, Heidelberg (2008)
7. Chung, J.Y., Yoon, H.Y., Song, M.S., Park, H.W.: Event related fMRI studies of voluntary and inhibited eye blinking using a time marker for EOG. Neuroscience Letters 395, 196–200 (2006)

8. Lalonde, M., Byrns, D., Gagnon, L., Teasdale, N., Laurendeau, D.: Real-time eye blink detection with GPU-based SIFT tracking. In: Proc. 4th Canadian Conference on Computer and Robot Vision, Montreal, Canada, pp. 481–487 (2007)
9. Łęski, J.: A new possibility of non-invasive electrocardiological diagnosis. Politechnika Śląska, Zeszyty Naukowe, Gliwice (1994)
10. O'Haver, T.: Peak Finding and Measurement, `http://www.wam.umd.edu/~toh/spectrum/PeakFindingandMeasurement.htm` (accessed February 17, 2009)
11. Pan, G., Sun, L., Wu, Z., Lao, S.: Eyeblink-based anti-spoofing in sace recognition from a generic webcamera. In: Proc. 11th IEEE International Conference on Computer Vision, Rio de Janeiro, Brazil (2007), doi:10.1109/ICCV.2007.4409068
12. Pander, T.: An application of detection function in 3rd order spectrum shift method for high-resolution alignment of ECG cycles. In: Proc. 15th Biennial International Eurasip Conference BIOSIGNAL 2000, Brno, Czech Republic, pp. 66–68 (2000)
13. Park, K., Lee, K.: Eye-controlled human/computer interface using the line-of-sight and the intentional blink. Comp. Industry Eng. 30, 463–473 (1996)
14. Skotte, J., Nøjgaard, J., Jørgensen, L., Christensen, K., Sjøgaard, G.: Eye blinking frequency during different computer tasks quantified by electrooculography. Eur. J. Appl. Physiol. 99, 113–119 (2007)
15. Venkataramanan, S., Prabhat, P., Choudhury, S., Nemade, H., Sahambi, J.: Biomedical instrumentation based on electrooculogram (EOG) signal processing and application to a hospital alarm system. In: Proc. Intern. Conference on Intelligent Sensing and Information Processing, Chennai, India, pp. 535–540 (2005)

Self-directed Training in an Evidence-Based Medicine Experiment

M.M. Marta[1], S.D. Bolboacă[2], and L. Jäntschi[3]

[1] Department of Foreign Laguages, The Iuliu Hațieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania
mmarta@umfcluj.ro

[2] Department of Medical Informatics and Biostatistics,
The Iuliu Hațieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania
sbolboaca@umfcluj.ro

[3] Department of Chemistry, Technical University of Cluj-Napoca, Cluj-Napoca, Romania
lori@academicdirect.org

Abstract. The paper presents an interactive training and evaluation system designed for the evidence-based medicine (EBM) training of Romanian undergraduate students. Fourteen tutorials and a series of supplementary materials were developed and integrated into a virtual self-directed training environment in order to provide the opportunity to learn about and to assess evidence-based medicine knowledge and skills. The interactive web-based approach was efficient and effective in the EBM education of undergraduate students, suggesting that it could be the appropriate method of teaching the evidence-based medicine.

1 Introduction

Evidence-based medicine education, part of the concept of evidence-based medicine (EBM), introduced by Guyatt et all [1], refers to training undergraduate students [2], residents [3] and practitioners [4] to use the best available, valid and reliable evidence for individual daily medical decisions. Evidence-based practice seems to be a useful instrument for improving health care quality and controlling the costs of health services [5, 6]. EBM education is the first step in putting the concept into practice [7, 9]. Compared with no intervention, short EBM educational strategies have proved able to transfer knowledge and improve critical appraisal skills [9-12].

Information and communication technology lead to changes in medical practice [13] that influence the quality, efficiency and costs of medical care [14]. Moreover, changes in the translation of medical knowledge into medical practice, influencing the evidence-based practice [15] and education in medicine [16] are identified.

A self-directed computer-based curriculum in the Romanian language (run at the Iuliu Hațieganu University of Medicine and Pharmacy) and its effectiveness, were evaluated.

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 193–204.
springerlink.com

2 Material and Method

2.1 Training and Evaluation System

The computer-based EBM curriculum was designed with three core goals: (1) to promote access to EBM knowledge and resources in the Romanian language for undergraduate medical students, (2) to increase students' awareness and use of relevant medical evidence, and (3) to teach the calculation and interpretation of fundamental EBM metrics (therapy intervention, diagnostic or screening studies, disease prognosis studies). The computer-based EBM design respects the principles of educational psychology [17].

The EBM training and evaluation system contains tutorials, several supplementary materials and an evaluation system. Fourteen tutorials (see Table 1) were developed; each of them contains objectives, prerequisites, training materials divided into chapters, clinical-based problems, references, and a self-evaluation test. The self-evaluation test is interactive and comprises five multiple-choice questions (two problem-based questions) with one up to four correct answer(s).

Six types of supplementary materials [18] were available through the EBM training system. In addition, software for a) assisting the creation and browsing of critical appraisal topics [19-21], b) assisting the creation and browsing of guideline models and clinical practice guidelines [22], and c) calculating the 95% confidence interval for proportions[1] [24-33] is available. This material is completed

Table 1. Tutorials: evidence-based medicine curriculum

Tutorial	Remarks
1st Introduction to Evidence-Based Medicine 2nd Asking Answerable Questions 3rd Medical Evidence 4th Finding Evidence 5th Applying Evidence in Day-to-Day Practice 6th Decisions Based on Evidence*	The order of these tutorials must be respected for understanding all evidence based medicine concepts
Study Assessment of Therapy (7th) Diagnostic Test (8th) Screening Test (8th) Prognosis (10th) Etiology (11th) Economic Analysis (12th) Decisional Trees* (13th)	Can be completed in any order with one exception (*): the 13th tutorial must be complete after the 6th tutorial
Assessment of Evidence-Based Clinical Practice Guidelines (14th)	Must be the last broach tutorial

[1] http://vl.academicdirect.org/applied_statistics/binomial_distribution/

with twenty diagnostic and treatment guidelines published by the Romanian College of Physicians[2]; seventeen materials on proved based medicine published by the Stethoscope Journal[3]), and an EBM glossary.

An online tutor-assisted evaluation environment is merged into the system in order to facilitate the evaluation of the acquired knowledge. A detailed description of the evaluation system is available in [23].

The tutorials are in the Romanian language while the resources and both in Romanian and English.

The self-directed training and evaluation computer program system, is merged into a Windows help application by using the HTML Help Workshop (version 4, Microsoft®, free to use).

2.2 System Evaluation Methodology: Evidence-Based Medicine

System Evaluation

Third to sixth year students at the Faculty of Medicine (who do practicums in hospitals and dispensaries) represented the target population. The 4th year students of the Faculty of Medicine, The Iuliu Hațieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania were the available population. One group (out of five) of students, randomly selected, was included in the research. The students participated in a traditional two-hour EBM course that covered the basics of practicing evidence-based medicine. The evidence-based medicine knowledge and critical appraisal of evidence of each student has tested at the end of the EBM traditional course. It was done, using eighteen true/false paper-based questionnaires that incorporated five problem-based questions. Students also completed the baseline characteristics form that included demographic data (gender, age) and information about computer and Internet access. They also ranked their perception of continuing medical education, quality of health care, and patient satisfaction on a 1 to 5 scale (1 = extremely unimportant, 2 = unimportant, 3 = unconcerned, 4 = important, 5 = extremely important). A question regarding their previous experience with EBM concepts was also included. All the participants had previously received training in research methodology, epidemiology, and statistics.

Since the topic of the study is not included in the core medical curriculum, students enrolled voluntarily in the intervention group after the EBM traditional course. Students were eligible for the intervention group, if they met the following criteria: attended the traditional EBM education course, and filled in a participation and consent form. Access to an individual computer with CD-ROM, with or without an Internet connection, was necessary to get a prerequisite for the enrollment in the intervention group. The students in the intervention group received additional training via an interactive computer-based curriculum available on CD-ROM and on the Internet. We found, that the computer-based curriculum depends on EBM knowledge; it presents clinical problems with or without solutions and it offers links to evidence-based medicine resources in Romanian and English as

[2] http://www.cmr.ro

[3] http://www.stetoscop.ro

well as links to medical publications. EBM knowledge and skills, searching techniques and the critical appraisal of evidence were measured in the intervention group with a computer- and tutor-assisted online multiple-choice questionnaire at the beginning (pre-test) and at the end (post-test) of the three-month self-directed training. The questionnaire had forty-five questions with five possible options and one up to four correct answer(s); fifteen of them were clinical problem-based questions. A post-study survey, carried out in the intervention group, provides the evaluation of the developed system. Tested were the following problems: ▪ advantages offered by using the system in EBM training; ▪ usefulness of the developed application in EBM education; ▪ evaluation of ease of using the system; ▪ application usefulness in medical practice.

Statistical Analysis

The participant's characteristics were summarized and comparisons between the two groups (intervention and control) were carried out according to variable types. The differences in proportions were tested using chi-square statistics and the difference test for proportions (Statistica 8.0). The differences between the scores and the number of correct answers were evaluated using the Mann-Withney U test (comparison between control and intervention groups). The data from pre- and post-test knowledge assessment, in the intervention group were analyzed using the Wilcoxon test. In each statistical experiment, a significance level of 5% was accepted. The 95%-confidence intervals for proportions were computed with a method based on the binomial distribution hypothesis [24].

3 Results

3.1 Study Participants

Ninety-six students were included in the study: fifty-six in the control group and forty in the intervention group. For summary of the demographic data of the participants, in the intervention and control groups, see Table 2.

No statistically significant differences were identified between the groups in terms of gender ($p = 0.743$), age (mean 21.78 in the intervention group and 21.91 in the control group, $p = 0.235$), computer access ($p = 0.713$), and Internet access ($p = 0.676$). However, significant differences between the proportion of students who ranked Internet access as being "*relatively difficult*" was identified in the intervention and control groups ($p = 0.0432$). As far as the moment when students first heard about EBM was concerned, statistically significant differences were obtained ($p = 0.003$) between groups, a higher percent of students from the intervention group had previously heard about EBM concepts compared with the control group.

As far as the importance given by students to continuing medical education, quality of health care and patient satisfaction were concerned, no significant differences were found between groups ($p > 0.05$), with two exceptions. The exceptions were identified for quality of health care on two scales "*important*" ($p = 0.0019$) and "*extremely important*" ($p = 0.0001$).

Table 2. Demographic characteristics of participants

Characteristic	Intervention Group (n = 40) Percent (%) [95% CI]	Control Group (n = 56) Percent (%) [95% CI]
Gender Female	67.50 [50.06–82.44]	64.29 [50.03–76.75]
Age		
▪ 21 years old	32.50 [17.56–49.94]	19.64 [10.75–32.11]
▪ 22 years old	57.50 [40.06–72.44]	69.64 [55.39–82.11]
▪ 23 years old	10 [2.56–22.44]	10.71 [3.60–21.40]
Previously heard about EBM	55 [37.56–69.94]	25 [14.32–37.47]
Computer access		
▪ Easy	42.50 [27.56–59.94]	44.64 [30.39–58.90]
▪ Relatively easy	32.50 [17.56–49.94]	19.64 [10.75–32.11]
▪ Difficult	20.00 [10.06–34.94]	30.36 [17.89–44.61]
▪ No access	2.50 [0.06–12.44]	1.79 [0.03–8.90]
▪ I don’t know	2.50 [0.06–12.44]	3.57 [0.03–12.47]
Internet access		
▪ Easy	27.50 [15.06–44.94]	33.93 [21.46–48.18]
▪ Relatively easy	45.00 [30.06–62.44]	25.00 [14.32–37.47]
▪ Difficult	22.50 [10.06–37.44]	35.71 [23.25–49.97]
▪ No access	5.00 [0.06–17.44]	1.79 [0.03–8.90]
▪ I don’t know	0.00 [n.a.]	3.57 [0.03–12.47]
Continuing medical education		
▪ Unconcerned	0.00 [n.a.]	5.36 [1.82–14.25]
▪ Important	60.00 [42.56–74.94]	50.55 [35.75–64.25]
▪ Extremely important	37.50 [22.56–54.94]	44.64 [30.39–58.89]
▪ Missing data	2.50 [0.06–12.44]	0.00 [n.a.]
Quality of health care		
▪ Unconcerned	5.00 [0.06–17.44]	16.07 [7.17–28.54]
▪ Important	35.00 [20.06–52.44]	67.86 [53.60–80.32]
▪ Extremely important	55.00 [37.56–69.94]	16.07 [7.17–28.54]
▪ Missing data	5.00 [0.06–17.44]	0.00 [n.a.]
Patient satisfaction		
▪ Unconcerned	5.00 [0.06–17.44]	3.57 [0.03–12.47]
▪ Important	50.00 [32.56–67.44]	53.57 [39.32–67.82]
▪ Extremely important	40.00 [25.06–57.44]	42.86 [30.39–57.11]
▪ Missing data	5.00 [0.06–17.44]	0.00 [n.a.]

95% CI = 95% confidence interval, n. a. = not applicable

3.2 Intervention Group: Pre- and Post-test Assessment of EBM Knowledge

The students in the intervention group completed a questionnaire with forty-five multiple-choice questions in order to have their knowledge of EBM assessed at the beginning and completion of the self-directed training period. The tests were tutor-assisted and the students received maximum 45 points. The number of correct and incorrect answers (for each student) expressed as absolute frequency (f_a) and confidence interval for relative frequency ($95\%CI_{fr}$), are presented in Table 3.

The number of correct answers to the pre-test proved to be significantly lower compared with the number of correct answers to the post-test (n = 40, Wilcoxon Z = 5.51, p < 0.001); the number of incorrect answers to the pre-test was significantly higher compared with the number of incorrect answers to the post-test (n = 40, Wilcoxon Z = 5.51, p < 0.001).

Table 3. Distribution of the correct and incorrect answers to pre- and post-test

	Pre-test		Post-test	
	Correct	Incorrect	Correct	Incorrect
StdID	fa [95%CIfr]	fa [95%CIfr]	fa [95%CIfr]	fa [95%CIfr]
std_01	7 [6.72-28.84]	38 [71.16-93.28]	41 [77.83-97.73]	4 [2.27-22.17]
std_02	2 [0.05-15.51]	43 [84.49-99.95]	38 [71.16-93.28]	7 [6.72-28.84]
std_03	5 [4.49-24.40]	40[75.60-95.51]	38 [71.16-93.28]	7 [6.72-28.84]
std_04	5 [4.49-24.40]	40 [75.60-95.51]	37 [68.94-93.28]	8 [6.72-31.06]
std_05	7 [6.72-28.84]	38 [71.16-93.28]	41 [77.83-97.73]	4 [2.27-22.17]
std_06	5 [4.49-24.40]	40 [75.60-95.51]	36 [64.49-91.06]	9 [8.94-35.51]
std_07	9 [8.94-35.51]	36 [64.49-91.06]	32 [55.60-84.40]	13 [15.60-44.40]
std_08	6 [4.49-26.62]	39 [73.38-95.51]	37 [68.94-93.28]	8 [6.72-31.06]
std_09	3 [2.27-17.73]	42 [82.27-97.73]	36 [64.49-91.06]	9 [8.94-35.51]
std_10	5 [4.49-24.40]	40 [75.60-95.51]	37 [68.94-93.28]	8 [6.72-31.06]
std_11	5 [4.49-24.40]	40 [75.60-95.51]	41 [77.83-97.73]	4 [2.27-22.17]
std_12	8 [6.72-31.06]	37 [68.94-93.28]	40 [75.60-95.51]	5 [4.49-24.40]
std_13	4 [2.27-22.17]	41 [77.83-97.73]	35 [62.27-88.84]	10 [11.16-37.73]
std_14	2 [0.05-15.51]	43 [84.49-99.95]	35 [62.27-88.84]	10 [11.16-37.73]
std_15	6 [4.49-26.62]	39 [73.38-95.51]	37 [68.94-93.28]	8 [6.72-31.06]
std_16	4 [2.27-22.17]	41 [77.83-97.73]	42 [82.27-97.73]	3 [2.27-17.73]
std_17	4 [2.27-22.17]	41 [77.83-97.73]	38 [71.16-93.28]	7 [6.72-28.84]
std_18	8 [6.72-31.06]	37 [68.94-93.28]	39 [73.38-95.51]	6 [4.49-26.62]
std_19	3 [2.27-17.73]	42 [82.27-97.73]	38 [71.16-93.28]	7 [6.72-28.84]
std_20	2 [0.05-15.51]	43 [84.49-99.95]	38 [71.16-93.28]	7 [6.72-28.84]
std_21	6 [4.49-26.62]	39 [73.38-95.51]	38 [71.16-93.28]	7 [6.72-28.84]
std_22	6 [4.49-26.62]	39 [73.38-95.51]	36 [64.49-91.06]	9 [8.94-35.51]
std_23	3 [2.27-17.73]	42 [82.27-97.73]	37 [68.94-93.28]	8 [6.72-31.06]
std_24	6 [4.49-26.62]	39 [73.38-95.51]	41 [77.83-97.73]	4 [2.27-22.17]

Table 3. (*continued*)

std_25	4 [2.27-22.17]	41 [77.83-97.73]	37 [68.94-93.28]	8 [6.72-31.06]
std_26	6 [4.49-26.62]	39 [73.38-95.51]	36 [64.49-91.06]	9 [8.94-35.51]
std_27	8 [6.72-31.06]	37 [68.94-93.28]	37 [68.94-93.28]	8 [6.72-31.06]
std_28	7 [6.72-28.84]	38 [71.16-93.28]	37 [68.94-93.28]	8 [6.72-31.06]
std_29	5 [4.49-24.40]	40 [75.60-95.51]	38 [71.16-93.28]	7 [6.72-28.84]
std_30	2 [0.05-15.51]	43 [84.49-99.95]	36 [64.49-91.06]	9 [8.94-35.51]
std_31	5 [4.49-24.40]	40 [75.60-95.51]	38 [71.16-93.28]	7 [6.72-28.84]
std_32	6 [4.49-26.62]	39 [73.38-95.51]	40 [75.60-95.51]	5 [4.49-24.40]
std_33	6 [4.49-26.62]	39 [73.38-95.51]	40 [75.60-95.51]	5 [4.49-24.40]
std_34	5 [4.49-24.40]	40 [75.60-95.51]	38 [71.16-93.28]	7 [6.72-28.84]
std_35	3 [2.27-17.73]	42 [82.27-97.73]	38 [71.16-93.28]	7 [6.72-28.84]
std_36	3 [2.27-17.73]	42 [82.27-97.73]	38 [71.16-93.28]	7 [6.72-28.84]
std_37	2 [0.05-15.51]	43 [84.49-99.95]	40 [75.60-95.51]	5 [4.49-24.40]
std_38	7 [6.72-28.84]	38 [71.16-93.28]	40 [75.60-95.51]	5 [4.49-24.40]
std_39	3 [2.27-17.73]	42 [82.27-97.73]	34 [60.05-86.62]	11 [13.38-39.95]
std_40	1 [0.05-11.06]	44 [88.94-99.95]	41 [77.83-97.73]	4 [2.27-22.17]

3.3 Intervention Group: Usability Analysis of the Training and Evaluation System

Thirty-six (90%) medical students in the intervention group considered that the system offered a friendly interactive training environment. The participants identified the following advantages: self-evaluation facilities associated to each tutorial (95%), possibility of choosing the appropriate time and place for EBM education (80%), guidance in accessing electronic medical journals (60%), searching and retrieving medical information (95%), access to EBM resources in the Romanian language (97.5%).

The usefulness of the EBM education and evaluation system was regarded as indifferent by one student (2.5%), useful by twenty-eight students (77.5%), and very useful by eleven students.

The ease of using the application was perceived as *relatively difficult* by seven students (17.5%), as *relatively easy* by eleven students (27.5%), as *easy* by eight students (20%) and as *very easy* by fourteen students (35%).

Twenty-four students (60%) considered that the training and evaluation system was a helpful instrument in EBM training while twenty students (50%) regarded it as a useful instrument in medical practice.

3.4 Comparison between Control and Intervention Groups

Knowledge Evaluation: Traditional Course

The number of correct answers to the 18-question survey varied from 7 to 16. With one exception, the numbers of correct answers were not significantly different in the control and intervention groups (p = 0.7948, $n_{intervention}$ = 40, $n_{control}$ = 56). A significantly higher percent (p = 0.0426) of students in the intervention group (nine students out of forty) gave nine correct answers out of eighteen compared with the control group (three students out of fifty-six). In order to identify significant differences at the end of the EBM traditional intervention, the Mann-Withney U test was used to compare the average of the correct answers given by the students in the intervention and control groups. The results revealed that, at the end of the EBM course, there were not significant differences in EBM knowledge between the intervention and control groups (p > 0.05).

Groups' Comparison
EBM knowledge in the intervention (assessment at the end of the self-directed training) and control groups, were compared by analyzing the averages of the proportion of correct answers.

The results showed that the average proportion of correct answers in the intervention group (0.84, n = 40) was significantly higher (p = 0.0174) compared with the average proportion of correct answers in the control group (0.62, n = 56). In order to identify the source of this difference, four cases were analyzed by splitting the intervention and control group classes according to the percent of correct answers (see Table 4).

Table 4. Results of comparison between interventional and control groups

Parameter	Class 1 (≥50%)	Class 2 (≥60%)	Class 3 (≥70%)	Class 4 (≥80%)
	int ≥ 23; con ≥ 9	int ≥ 27; con ≥ 11	int ≥ 32; con ≥ 13	int ≥ 36; con ≥ 14
$f_{a\text{-}int}$	40	40	40	36
$f_{r\text{-}int}$ [95% $CI_{fr\text{-}int}$]	1 [0.90–1.00]	1 [0.90–1.00]	1 [0.90–1.00]	0.9 [0.78–0.97]
$f_{a\text{-}con}$	51	32	11	6
$f_{r\text{-}con}$ [95% $CI_{fr\text{-}con}$]	0.9 [0.80–0.96]	0.6 [0.43–0.70]	0.2 [0.11–0.32]	0.1 [0.04–0.21]
p	0.0421	< 0.001	< 0.001	< 0.001

f_a = absolute frequency; f_r = relative frequency; 95% CI_{fr} = 95% confidence interval for relative frequency;
int = intervention group (n = 40); con = control group (n = 56);
int ≥ 23 = the number of correct answers in intervention group greater than or equal to 23

4 Discussion

The EBM training and evaluation system proved to be efficient and effective. Although a number of EBM educational resources were already available on the Internet [34, 35], the proposed EBM computer-based curriculum is unique. The system is unique because it follows a deliberate sequence of educational activities in order to promote the reflection and active interaction of the students with online available resources, by requiring them to apply EBM concepts for solving real medical problems. It is also a unique resource for EBM education in the Romanian language.

The two investigated groups included homogenous participants in terms of gender, age, computer and Internet access. The enrolment of the students in the intervention group, could be explained by their previous contact with EBM concepts (a significantly higher proportion of students in the intervention group compared with the control group). A significantly higher proportion of students in the control group also ranked Internet access as "*relatively difficult*", which could explain their lower interest in self-directed EBM training.

The analysis of the students' perception of continuing medical education, quality of health care and patient satisfaction revealed the following:

- All participants ranked the importance of continuing medical education and patient satisfaction almost identically; but a higher percent of students in the control group ranked the quality of health care as "*important*", while a higher percent of students in the intervention group regarded it as "*extremely important*".
- The importance of health care quality was ranked by most of the participants as "*extremely important*" (intervention group) or "*important*" (control group).

No significant differences in terms of number of correct answers to the assessment of EBM knowledge performed at the end of the traditional course sustained the validity of the results obtained in the intervention group. The analysis of the performance of the intervention group students revealed that EBM knowledge and skills improved (a significantly higher percent of correct answers were obtained to the post-test compared to the pre-test, see Table 3). This result suggested that basic EBM knowledge could open the pathway to practicing evidence-based medicine, the individual being responsible for continuing personal training. Most of the participants in the intervention group found that the EBM training and evaluation system was friendly and allowed self-paced, independent training; the students were able to tailor the learning experience as well as the time and place of EBM education to personal preferences. Guidance to accessing medical journals, searching and retrieving medical information, as well as access to EBM resources in the native language have also been identified as advantages of this training and evaluation system. The EBM training and evaluation system was regarded by more than half of the students in the intervention group as a helpful EBM training instrument and a useful tool in medical practice.

The analysis of the comparison between the control and intervention groups revealed that the 95% confidence intervals for the proportion of students that gave a specified number of correct answers in three out of four criteria (≥ 60%, ≥ 70%, and ≥ 80%) did not overlap. These observations showed that the students in the intervention group obtained higher EBM performances compared with the students in the control group. The most relevant information in Table 4 is the result obtained for criterion ≥ 80, because for passing an exam students had to prove that they acquired 80 percent of the information.

Although the outcome variables were not identical in the two groups, the differences were higher than observed. The differences derived from different evaluation methods used for knowledge assessment in the two studied groups. The method selected to evaluate the knowledge of the control group were constructed as true/false statements. In terms of probability, the chance of guessing the correct answer to a question of this type equals 0.5. The probability equals 0.0278 in an eighteen-question questionnaire. On average, the probability of guessing the correct answer to a multiple-choice question with five choices and one up to four correct answers is of 0.15. For a sample of forty-five questions, the probability becomes 0.0033, and is ten times smaller in comparison with the probability of guessing the correct answer to a test with eighteen true/false statements.

The research had some limitations. The first limitation refers to the allocation of students into groups. Not all medical students included in the study had access to an individual computer. Because a realistic solution to the problem of student access to computers in the university computer labs, the student's allocation in the intervention or control group was not random. The second limitation refers to the research outcomes. Neither critical appraisal skills (as skills not as knowledge about skills) nor the attitude regarding EBM were included in the study outcome because more time could not be allotted to the traditional approach and web-based training for EBM due to the students' extremely busy schedule. The series of students randomly chosen was doing a six-week module with three subjects and a daily average of three hours of lectures and/or seminars plus four hours of practical and/or clinical activities. According to the students' availability for EBM education, the aim of the research was limited to EBM knowledge and problem-based assessment. Future intervention in evidence-based medicine skills and attitude is required.

The research compared only two methods of teaching EBM and focused on undergraduate medical students. More research aimed at comparing EBM educational strategies in other healthcare professionals such as nurses, residents, and practitioners could identify the best solution for EBM training according to learner category. The following also require future investigation: critical appraisal skills for searching the best available evidence able to answer a specific medical question, the assessment of the validity, reliability and usefulness of the evidence in treating an individual patient for including it in the daily medical decision-making process and the long-term effects on patient outcome.

5 Conclusions

The interactive web-based approach was efficient and effective in the EBM education of undergraduate students, suggesting that it could be the appropriate method of teaching evidence-based medicine.

The self-directed educational approach offered students the possibility of choosing the time, place and modality of learning. It also provided the possibility of using specialized online resources, an interactive self-evaluation environment, and an interactive web-based multiple-choice questionnaire for knowledge assessment.

However, more research aimed at comparing the proposed web-based curriculum with other educational models, applied on residents and practitioners is required.

References

1. Guyatt, G.H.: Evidence-based medicine. ACP J. Club. 114, A–16 (1991)
2. Riegelman, R.K., Garr, D.R.: Evidence-based public health education as preparation for medical school. Acad. Med. 83(4), 321–326 (2008)

3. Dahm, P., Preminger, G.M., Scales Jr., C.D., Fesperman, S.F., Yeung, L.L., Cohen, M.S.: Evidence-based medicine training in residency: A survey of urology programme directors. BJU International 103(3), 290–293 (2009)
4. Castillo, D.L., Abraham, N.S.: Knowledge management: how to keep up with the literature. Clin. Gastroenterol Hepatol 6(12), 1294–1300 (2008)
5. O'Kane, M., Corrigan, J., Foote, S.M., Tunis, S.R., Isham, G.J., Nichols, L.M., Fisher, E.S., Ebeler, J.C., Block, J.A., Bradley, B.E., Cassel, C.K., Ness, D.L., Tooker, J.: Crossroads in quality. Health Aff. 27(3), 749–758 (2008)
6. Wise, C.G., Bahl, V., Mitchell, R., West, B.T., Carli, T.: Population-based medical and disease management: an evaluation of cost and quality. Dis. Manag. 9, 45–55 (2006)
7. Alper, B.S., Vinson, D.C.: Experiential curriculum improves medical students' ability to answer clinical questions using the Internet. Fam. Med. 37, 565–569 (2005)
8. O'Neall, M.A., Brownson, R.C.: Teaching evidence-based public health to public health practitioners. Ann. Epidemiol. 15, 540–544 (2005)
9. Rhodes, M., Ashcroft, R., Atun, R.A., Freeman, G.K., Jamrozik, K.: Teaching evidence-based medicine to undergraduate medical students: a course integrating ethics, audit, management and clinical epidemiology. Med. Teach. 28, 313–317 (2006)
10. Yew, K.S., Reid, A.: Teaching evidence-based medicine skills: an exploratory study of residency graduates' practice habits. Fam. Med. 40, 24–31 (2008)
11. Nicholson, L.J., Warde, C.M., Boker, J.R.: Faculty training in evidence-based medicine: improving evidence acquisition and critical appraisal. J. Cont. Educ. Health Prof. 27, 28–33 (2007)
12. Kilian, B.J., Binder, L.S., Marsden, J.: The emergency physician and knowledge transfer: continuing medical education, continuing professional development and self-improvement. Acad. Emerg. Med. 14, 1003–1007 (2007)
13. Biswas, R., Maniam, J., Lee, E.W.H., Gopal, P., Umakanth, S., Dahiya, S., Ahmed, S.: User-driven health care: answering multidimensional information needs in individual patients utilizing post-EBM approaches: an operational model. J. Eval. Clin. Pract. 14(5), 750–760 (2008)
14. Chaudhry, B., Wang, J., Wu, S., Maglione, M., Mojica, W., Roth, E., Morton, S.C., Shekelle, P.G.: Systematic review: impact of health information technology on quality, efficiency and costs of medical care. Ann. Intern. Med. 144(10), 742–752 (2006)
15. Gartlehner, G.: Evidence-based medicine breaking the borders: working model for the European Union to facilitate evidence-based healthcare. Wien Med. Wochenschr. 154, 127–132 (2004)
16. Davis, J., Chryssafidou, E., Zamora, J., Davies, D., Khan, K., Coomarasamy, A.: Computer-based teaching is as good as face to face lecture-based teaching of evidence-based medicine: a randomized controlled trial. BMC Med. Educ. 7, art no 23 (2007)
17. Carlos, J.S., Levis, J.R.: Psychological perspectives on contemporary educational issues. Information age Publishing, Inc., NC (2007)
18. Bolboacă, S., Jäntschi, L.: Virtual environment for continuing medical education. Electronic Journal of Biomedicine 2, 19–28 (2007)
19. Bolboacă, S., Jäntschi, L., Drugan, T., Achimaş-Cadariu, A.: Creating therapy studies critical appraised topics. CATRom original software for Romanian physicians. App. Med. Inform. 15, 26–33 (2004)
20. Bolboacă, S., Jäntschi, L., Achimaş-Cadariu, A.: Creating diagnostic critical appraised topics. CATRom original software for Romanian physicians. App. Med. Inform. 14, 27–34 (2004)

21. Bolboacă, S., Jäntschi, L., Achimaş Cadariu, A.: Creating etiology/prognostic critical appraised topics. CATRom original software for Romanian physicians. Appl. Med. Inform. 13, 11–16 (2003)
22. Bolboacă, S.D., Achimaş Cadariu, A., Jäntschi, L.: Evidence-based guidelines assisted creation through interactive online environment. Appl. Med. Inform. 17, 3–11 (2005)
23. Bolboacă, S., Jäntschi, L.: Computer-assisted training and evaluation system in evidence-based medicine. In: Proc. 11th Intern. Symp. for Health Information Management Research, Halifax, Nova Scotia, CA, pp. 220–226 (2006)
24. Drugan, T., Bolboacă, S., Jäntschi, L., Achimaş Cadariu, A.: Binomial distribution sample confidence intervals estimation 1. Sampling and medical key parameters calculation. Leonardo El J. Pract. Techn. 3, 47–74 (2003)
25. Bolboacă, S., Achimaş Cadariu, A.: Binomial distribution sample confidence intervals estimation 2. Proportion-like medical key parameters. Leonardo El J. Pract. Technol. 3, 75–110 (2003)
26. Bolboacă, S., Achimaş Cadariu, A.: Binomial distribution sample confidence intervals estimation 3, post- and pre-test odds. Leonardo J. Sci. 3, 24–46 (2003)
27. Bolboacă, S., Achimaş Cadariu, A.: Binomial distribution sample confidence intervals estimation 4, post-test probability. Leonardo J. Sci. 3, 47–70 (2003)
28. Bolboacă, S., Achimaş Cadariu, A.: Binomial distribution sample confidence intervals estimation 5, odds ratio. Leonardo J. Sci. 4, 26–43 (2004)
29. Bolboacă, S., Achimaş Cadariu, A.: Binomial distribution sample confidence intervals estimation 6, excess risk. Leonardo El J. Pract. Techn. 4, 1–20 (2004)
30. Bolboacă, S., Achimaş Cadariu, A.: Binomial distribution sample confidence intervals estimation 7, absolute risk reduction and ARR-like expressions. Leonardo El J. Pract. Technol. 5, 1–25 (2004)
31. Bolboacă, S., Achimaş Cadariu, A.: Binomial distribution sample confidence intervals estimation 8, number needed to treat/harm. Leonardo J. Sci. 5, 1–17 (2004)
32. Bolboacă, S., Jäntschi, L.: Binomial distribution sample confidence intervals estimation for positive and negative likelihood ratio medical key parameters. In: AIMA annual symposium on biomedical and health informatics, Washington D.C. (2005)
33. Bolboacă, S.: Binomial distribution sample confidence intervals estimation 10, relative risk reduction and RRR-like expressions. Leonardo El J. Pract. Technol. 6, 60–75 (2005)
34. Cook, D.A., Dupras, D.M.: Teaching on the web: automated online instruction and assessment of residents in an acute care clinic. Med.Teach. 26(7), 599–603 (2004)
35. Schilling, K., Wiecha, J., Polineni, D., Khalil, S.: An interactive web-based curriculum on evidence-based medicine: design and effectiveness. Fam. Med. 38(2), 126–132 (2006)

Part III
Psychological and Linguistic Aspects of H-CSI

Emotion Recognition from Facial Expression Using Neural Networks

G.U. Kharat[1] and S.V. Dudul[2]

[1] Electronics & Telecommunication Department, Anuradha Engineering College,
Chikhli, Maharashtra, India
gukharat@rediffmail.com
[2] Applied Electronics Department, SGB Amravati University, Amaravati,
Maharashtra, India
dudulsv@rediffmail.com

Abstract. This research aims at developing "Humanoid Robots" that can carry out intellectual conversation with human beings. The first step of our research is to recognize human emotions by a computer using neural network. In this paper all six universally recognized principal emotions namely angry, disgust, fear, happy, sad and surprise along with neutral one are recognized. Various neural networks such as Support Vector Machine (SVM), Multilayer Perceptron (MLP), Principal Component Analysis (PCA), and Generalized Feed Forward Neural Network (GFFNN) are employed and their performance is compared. 100% recognition accuracy is achieved on training data set (seen examples) and test data set (unseen examples).

1 Introduction

It is highly expected that computers and robots will be used more for betterment of our daily life. Information Technology organizations expect a harmonious interaction or heart to heart communication between computers and / or robots and human beings. For its realization it seems to be necessary that computers and robots will be implemented with artificial mind that enables them to communicate with human beings through exchanging not only logical information but also emotional one. The first step to realize mind implemented robot is to recognize human emotions. Meharabian [1] indicated that the verbal part (i.e. spoken words) of a message contributes only for a 7% of the effect of the message, the vocal part (i.e. voice information) contributes for 38% while facial expressions of the speaker contributes for 55% of the effect of the spoken message. Hence in order to develop "Active Human Interface" that realizes heart to heart communication between intelligent machine and human beings we are implementing machine recognition of human emotions from facial expressions.

Affective computing addresses issues relating to emotion in computing and has been pioneered by the work of Picard at MIT [2]. Picard describes how "Affective

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 207–219.
springerlink.com

interaction can have maximum impact when emotion recognition is available to both man and machine" and goes on to say if one party can't recognize or understand emotion then interaction is impaired [3]. The problem of recognizing facial expressions had attracted the attention of computer- vision community [4-7]. Bassili [8] suggested that motion in the image of the face would allow emotions to be identified even with minimal information about the spatial arrangement of features.

Essa [13, 14] proposed FACS+ model extending Facial Action Coding System (FACS) model to allow combine spatial and temporal modeling of facial expressions. Optical flow computations for recognizing and analyzing facial expressions are used by [5, 7, 11-18]. Anthropometric facial points are used for feature extraction to recognize emotions [7, 19].

This paper provides an innovative approach of using images and obtain their Discrete Cosine Transform (DCT) to extract features along with physical parameters, say energy, entropy, variance, standard deviation, contrast, homogeneity and correlation to obtain intelligent feature vector for the recognition of facial expression. SVM, MLP, PCA and GFFNN are used for recognition of emotions.

2 Facial Expression Database

Facial expression database in six universally recognized basic emotions and neutral one is collected from Japanese female database. Ten expressers posed 3 to 4 examples of each of the six emotions along with neutral one for a total of 219 images of facial expressions. This data was prepared when expresser looked into the semi reflective plastic sheet towards camera. Hairs were tied away to expose all expressive zones of the face. Tungsten lights were positioned to create even illumination on the face. The box enclosed the region between camera and plastic sheet to reduce back reflections. The images were printed in monochrome and digitized using flatbed scanner. Sample images from Japanese Female database is shown in Fig. 1. Total 210 images are used for the experiment.

3 Feature Extraction

Feature extraction is an important step in the recognition of human emotions using neural network. Discrete Cosine Transform (DCT) is used to extract the features. The number of features in DCT is varied and it is found that optimal results are obtained when 64 features are used. Program in MATLAB is developed to obtain DCT of each image.

Statistical parameters of an image give important clue about the image. The programs in MATLAB are also developed to obtain statistical parameters of the images namely, Standard Deviation, Entropy, Co-relation, Energy, Homogeneity, Contrast and Variance. Optimal feature vectors are obtained containing 64 values of DCT and 7 statistical parameters. Thus each optimal feature vector contains 71 features.

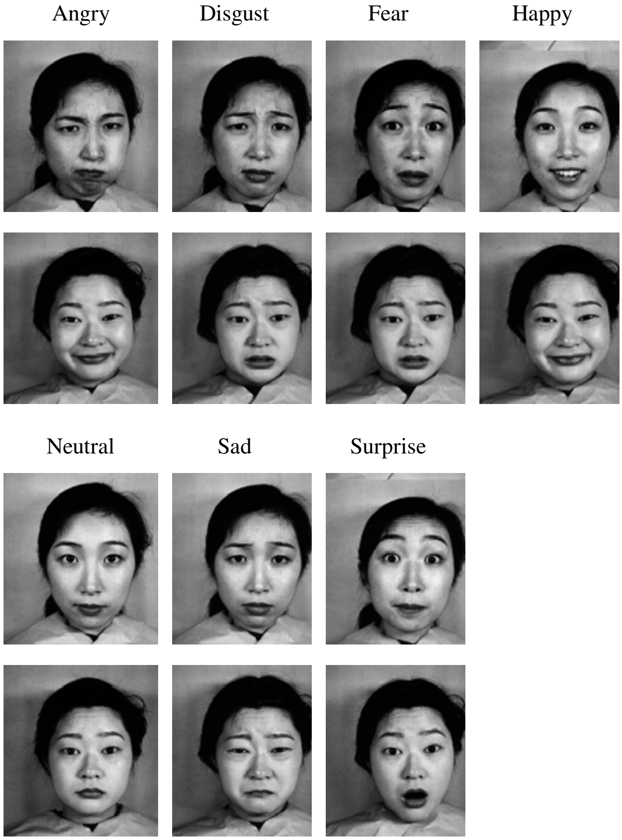


Fig. 1. Sample images from Japanese Female Database

4 Neural Networks for Emotion Recognition

The generalized procedure for emotion recognition from facial expressions using different neural networks is shown in Figure 2. Support Vector Machine (SVM), Multilayer Perceptron (MLP), Principal Component Analysis Neural Network (PCA NN), and Generalized Feed Forward Neural Network (GFFNN) are used one by one for emotion recognition using facial expressions. Number of input Processing Elements (PE) must be equal to that of input data of facial information. Since we have used 64 DCT & 07 statistical parameters of an image, 71 input Processing Elements are used in input layer. Seven Processing Elements are used in output layer for six emotions and neutral one.

4.1 SVM for Emotion Recognition

The randomized data is fed to the SVM network. The single hidden layer with 71 PEs delivers optimal result. The network is trained three times by varying the

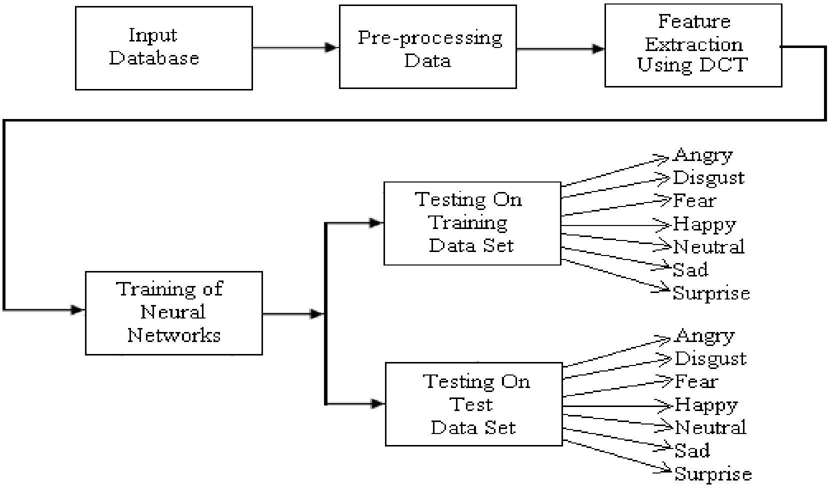


Fig. 2. Tentative scheme of procedure for emotion recognition

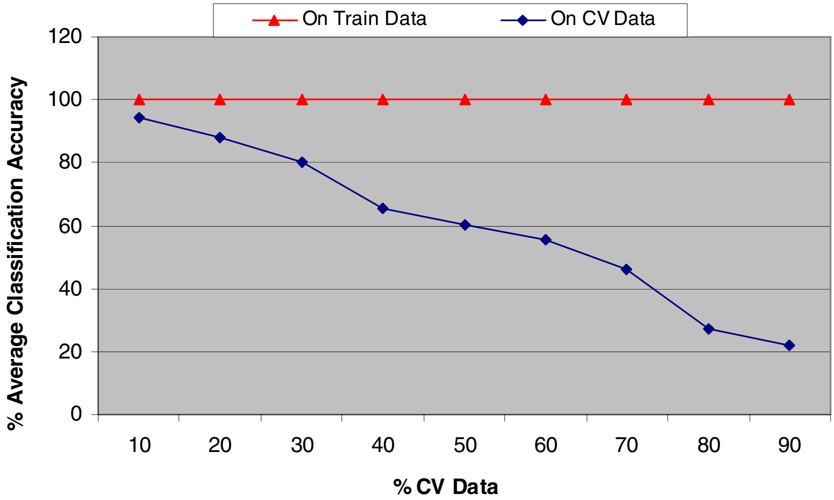


Fig. 3. Variation of % average classification accuracy with % of CV data

number of samples for training and cross validation (CV) data. The percentage average classification accuracy is calculated and is demonstrated in Figure 3. The optimal results are obtained when 90% data is used for training the network and 10% data for cross validation.

With 10% CV and 90% training data, the step size for training the SVM is varied from 0.01 to 1.0 and each time network is trained and tested on training and CV dataset. The graph of % average classification accuracy is plotted against step size in Figure 4. It is evident that average classification accuracy is 100% on train data and 94.28% on CV data.

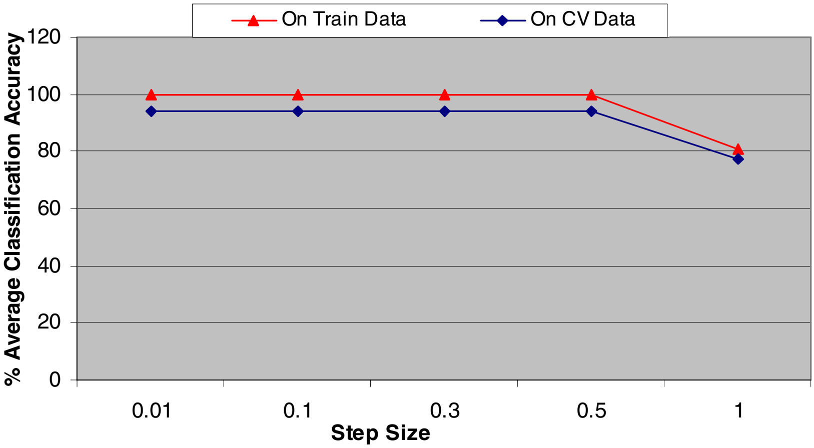


Fig. 4. Variation of % average classification accuracy with Step Size

Optimally designed SVM is

Learning control	= Supervised
Weight update	= Batch
Step size	= 0.5
Number of epochs	= 1000
Number of runs	= 03

Training is terminated at 100 epochs if there is no further improvement in MSE.

Time elapsed per epoch per exemplar	= 0.564mSec
Number of free parameters (P) of SVM	= 5616
Number of samples in training dataset (N)	= 189
N/P) ratio	= 0.0337

Finally, designed SVM is tested on training and CV dataset and results are depicted in table 1 and 2.

Table 1. Performance parameters for training data set using SVM

Performance	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
MSE	0.004573	0.004941	0.005143	0.005426	0.005126	0.005981	0.005363
Min Abs Error	0.001387	0.000248	0.001571	0.003599	0.000107	0.000127	0.002193
% Correct	**100**	**100**	**100**	**100**	**100**	**100**	**100**

The overall emotion recognition accuracy is = **100** %

Table 2. Performance parameters for CV data set using SVM

Performance	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
MSE	0.029818	0.045804	0.046596	0.065398	0.059857	0.085149	0.097773
Min Abs Error	0.001583	0.000860	0.013855	0.000672	0.000543	0.003794	0.065074
% Correct	**100**	**100**	**100**	**100**	**100**	**66.67**	**100**

The overall emotion recognition accuracy is = **95.24** %

4.2 MLP for Emotion Recognition

Japanese female database is randomized for generalization and true learning of neural networks. Randomized database is input to the MLP NN and is trained with different hidden layers. Maintaining one eye on simplicity of the designed network and other on maximum emotion recognition result, it is observed that single hidden layer entails better performance. The number of Processing Elements (PEs) used in the hidden layer is varied. The network is trained several times and minimum average MSE (0.0360) is observed on cross validation (CV) dataset when 10 PEs are used in hidden layer as depicted in Figure 5.

Data is partitioned into two parts: training data and cross validation (CV) data. The percentage of data used for training and cross validation is varied. The network is trained three times with different random initialization of weights so as to ensure true learning and avoid any biasing & affinity towards choice of specific initial connection weights. Minimum average MSE on training dataset and CV dataset along with average classification accuracy is calculated & it is observed that maximum recognition of emotion is obtained when 90% data is used for training and 10% data for cross validation (testing).

Using 90% data for training and 10 % data for CV, various transfer functions for the processing elements are used to train MLP NN. Average classification accuracy is measured and plotted in Figure 6. It is observed that tanh is the most

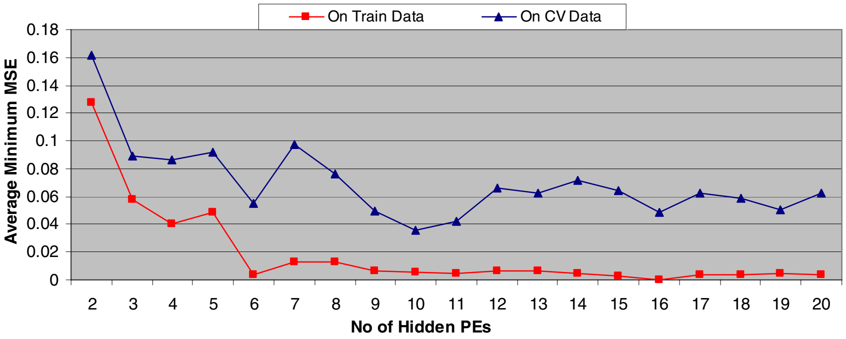


Fig. 5. Variation of average minimum MSE with number of processing elements in the hidden layer

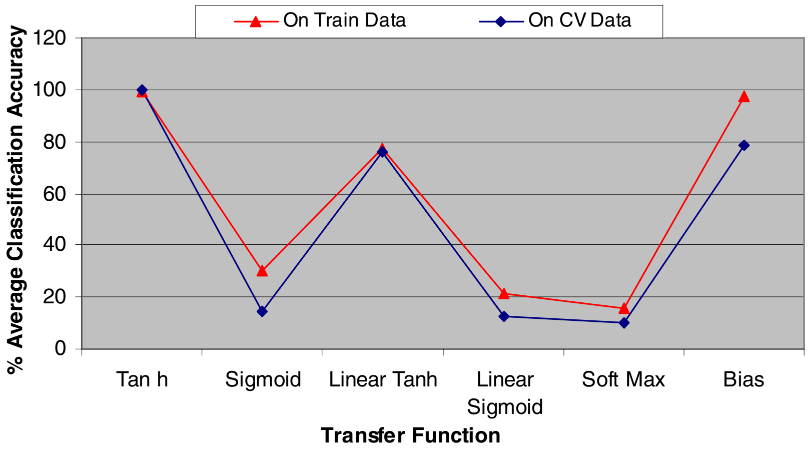


Fig. 6. Variation of % average classification accuracy with Transfer function

suitable transfer function as average classification accuracy is 99.47% on train data and 100% on CV data.

Now with 10 PEs in hidden layer, 10 % CV data, 90% training data and tanh transfer function, Neural Network is trained using various learning rules such as, momentum, Conjugate Gradient (CG), Quick Propagation (QP) and Delta Bar Delta (DBD). Minimum average MSE and average classification accuracy are computed. It was found that Momentum is the most appropriate learning rule for MLP NN.

The time required to process the data and complexity of the neural network are important performance parameters of any neural network in addition to emotion recognition accuracy. Time elapsed per epoch per exemplar is found out to be 0.0496 milliseconds. The ratio of number of instances in the training dataset to the number of free parameters of the network (N/P) is calculated and is found out to be 0.2371. From above experimentation selected parameters for MLP NN are:

MLP NN (71-10-7), Number of epochs	= 5000
Learning control	= Supervised
Weight update	= Batch
Step size	= 0.5
Number of epochs	= 3000
Number of runs	= 03

Training is terminated at 100 epochs without MSE improvement.

Time elapsed per epoch per exemplar	= 0.0496 mSec
Number of free parameters (P) of MLP	= 797
Number of samples in training dataset (N)	= 189
(N/P) ratio	= 0.2371

Thus, with rigorous experimentation and variation of parameters of the neural network, the optimal MLP is designed and tested on training and CV dataset and results are portrayed in table 3 and 4.

Table 3. Performance parameters for training data set using MLP

Performance	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
MSE	0.002260	0.002385	0.002210	0.002192	0.002843	0.002440	0.008098
Min Abs Error	0.000741	0.000314	0.000259	0.000227	0.000337	0.000748	0.001107
% Correct	**100**	**100**	**100**	**100**	**100**	**100**	**96.30**

The overall emotion recognition accuracy is = **99.47**%

Table 4. Performance parameters for CV data set using MLP

Performance	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
MSE	0.002895	0.045778	0.009529	0.006265	0.029642	0.038633	0.028476
Min Abs Error	0.013469	0.002336	0.008965	0.000310	0.001365	0.001955	0.003866
% Correct	**100**	**100**	**100**	**100**	**100**	**100**	**100**

The overall emotion recognition accuracy is = **100**%

4.3 PCA for Emotion Recognition

The randomized data is presented to the PCA neural network and is trained for different hidden layers. It is evident that PCA NN with single hidden layer delivers better performance. The number of Processing Elements (PEs) in the hidden layer is varied. The network is trained for 20 principal components (PC) and the average MSE is lowest on CV data when 19 PEs are used in the hidden layer as indicated in Figure 7.

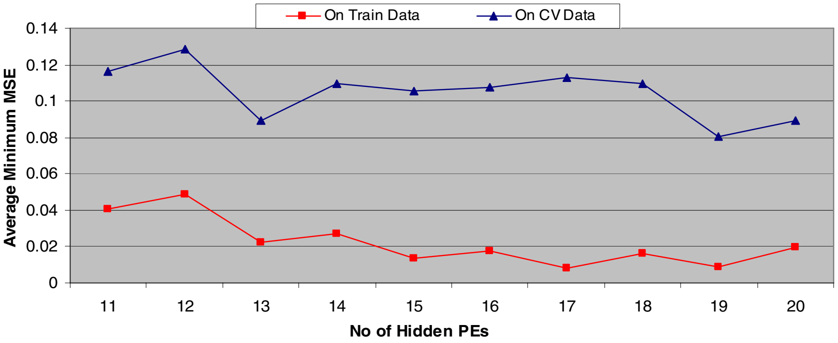


Fig. 7. Variation of average minimum MSE with number of processing elements in the hidden layer

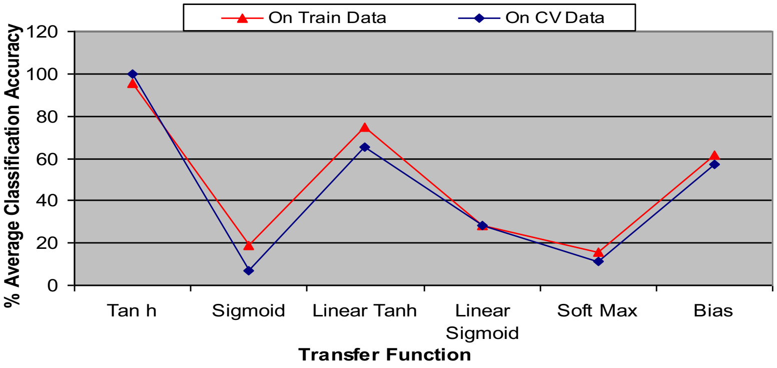


Fig. 8. Variation of % average classification accuracy with transfer function

Rigorous experimentation is performed by varying number of instances for training and cross validation data using 19 PEs in hidden layer. The network is trained three times. Average Minimum MSE on training and CV data set along with average classification accuracy is calculated and it is inferred that the best results are obtained when 15% instances are used for cross validation (CV) data and 85% for training.

Various transfer functions are used for training the network and average classification accuracy for different transfer function is plotted in Figure 8. It is inferred that tanh is the most suitable transfer function as average classification accuracy is 100% on train data and 95.47% on CV data.

Using tanh transfer function the PCA neural network is trained using learning rules namely, Sangers full, Ojas full, Momentum, Conjugate-Gradient (CG), Quick Propagation (QP), and Delta Bar Delta (DBD). Minimum MSE on training and CV data set is measured. Finally network is tested on training and cross validation dataset. It is observed that Sangers full and momentum are the most appropriate learning rules for our neural network.

From above experimentation selected parameters for PCA neural network are given below.

PCA NN (71-19-7), Number of epochs	= 5100
Unsupervised learning control epochs	= 100
Supervised learning control epochs	= 5000
Number of runs	= 03
Step size	= 0.5
Number of principal components	= 20
Instances for cross validation	= 15%
Instances for training	= 85%

Training is terminated at 500 epochs without MSE improvement.

Time elapsed per epoch per exemplar	= 0.0129 mSec.
Number of free parameters (P) of PCA NN	= 1508
Number of instances in training dataset (N)	= 178
(N/P) ratio	= 0.118.

Finally, designed PCA NN is tested on training and CV dataset and results are depicted in table 5 and 6.

Table 5. Performance parameters for training data set using PCA NN

Performance	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
MSE	0.010648	0.009319	0.009922	0.007877	0.019872	0.020167	0.017200
Min Abs Error	0.000374	0.001759	0.000226	0.000034	0.001261	0.000114	0.000720
% Correct	**96**	**95.65**	**92.59**	**100**	**100**	**94.74**	**89.29**

The overall emotion recognition accuracy is = **95.47**%

Table 6. Performance parameters for CV data set using PCA NN

Performance	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
MSE	0.001070	0.001596	0.001001	0.002180	0.002376	0.002049	0.001582
Min Abs Error	0.000091	0.000495	0.000202	0.000656	0.000499	0.000090	0.000262
% Correct	**100**	**100**	**100**	**100**	**100**	**100**	**100**

The overall emotion recognition accuracy is = **100**%

4.4 GFFNN for Emotion Recognition

The randomized data is fed to the GFF NN network and is trained for different hidden layers. It is evident that GFF NN with single hidden layer entails better performance. The number of Processing Elements (PEs) in the hidden layer is varied. The network is trained and average minimum MSE is obtained on CV data when 19 PEs are used in the hidden layer as indicated in Figure 9.

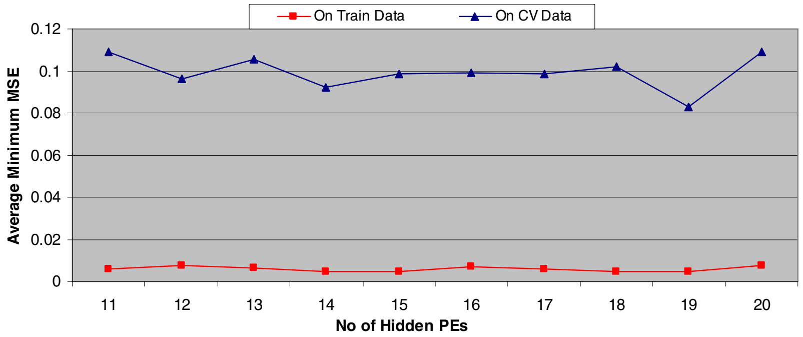


Fig. 9. Variation of average minimum MSE with number of processing elements in the hidden layer

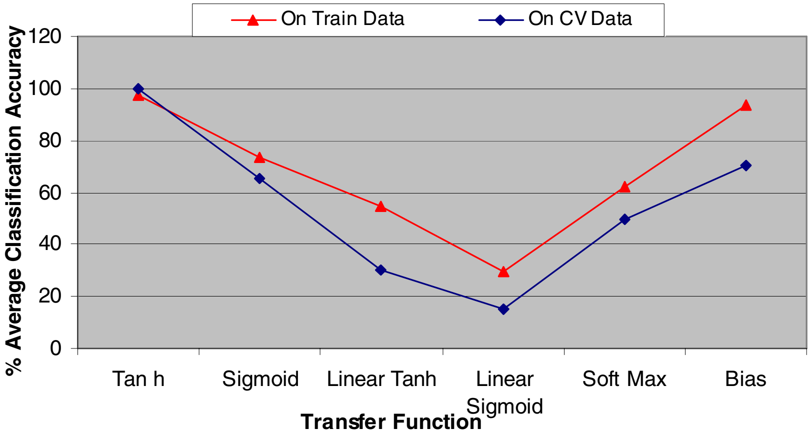


Fig. 10. Variation of % average classification accuracy with Transfer function

Rigorous experimentation is done by using 19 PEs in hidden layer and varying number of samples for training and cross validation data. The network is trained three times with different random initialization of connection weights. Average Minimum MSE on training and CV data set along with average classification accuracy is calculated and it is observed that the best results are obtained when 10% samples are used for cross validation (CV) and 90% for training.

Now various transfer functions are used for training the network and average classification accuracy for different transfer function is plotted in Figure 10. It is inferred that tanh is the most suitable transfer function because the average classification accuracy is 99.57% on train data and 100% on cross validation data.

With tanh transfer function the GFF NN neural network is trained using learning rules namely, Momentum, Conjugate-Gradient (CG), Quick Propagation (QP) and Delta Bar Delta (DBD). Average minimum MSE on training and CV data set is measured and it is observed that momentum is the most appropriate learning rule for our neural network.

From above experimentation selected parameters for Generalized Feed Forward Neural Network are given below.

GFF NN (71-19-7), Number of epochs	= 3000
Number of runs	= 03
Step size	= 0.5
Samples for cross validation	= 10%
Samples for training	= 90%

Training is terminated at 500 epochs if there is no further improvement in MSE

Time elapsed per epoch per exemplar	= 0.0496 mSec.
Number of free parameters (P) of GFF NN	= 1508
Number of samples in training dataset (N)	= 189
(N/P) ratio	= 0.1253

Finally, designed GFF NN is tested on training and CV dataset and results are portrayed in table 7 and 8.

Table 7. Performance parameters for training data set using GFF NN

Performance	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
MSE	0.003894	0.016805	0.008205	0.002288	0.008074	0.016625	0.013031
Min Abs Error	0.000176	0.000302	0.000471	0.000644	0.000028	0.000312	0.002085
% Correct	**100**	**92.31**	**100**	**100**	**100**	**93.10**	**96.55**

The overall emotion recognition accuracy is = 97.57%

Table 8. Performance parameters for CV data set using GFF NN

Performance	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
MSE	0.004986	0.002693	0.002151	0.003051	0.003722	0.002524	0.003095
Min Abs Error	0.002386	0.010128	0.000011	0.007991	0.015433	0.003593	0.020783
% Correct	**100**	**100**	**100**	**100**	**100**	**100**	**100**

The overall emotion recognition accuracy is = 100%

5 Conclusions

In this paper, the performance of the four Neural Networks MLP, SVM, PCA and GFFNN for recognition of emotions from facial expressions is evaluated to develop a “mind implemented computers and robots” and examined the validation of various Neural Networks in recognition of emotions and results are tabulated in table 9.

Table 9. Results of Emotion Recognition from Facial Expressions

NN	**t (μs)**	**N/P**	**MSE**		**% Accuracy**	
			Train	**Test**	**Train**	**Test**
MLP	49.6	.2371	.0229	.0014	99.47	100.00
SVM	564.0	.0337	.0051	.0614	100.00	95.24
PCA	12.9	.1180	.0135	.0016	95.47	100.00
GFF	49.6	.1253	.0098	.0031	97.57	100.00

Where t = time elapsed per epoch per exemplar per run for training
N/P = ratio of number of exemplars to the free parameters

It is evident from the table 9 that for MLP NN, the percentage of recognition of emotions is highest (100%),N/P ratio is highest indicating that the Neural Network is most simple to design and hardware realization, MSE is lowest and processing time for training the Neural Network is reasonably low. Hence MLP NN is recommended in recognition of emotions from facial expressions.

References

1. Meharabian, A.: Communication without words. Psychology Today 2(4), 53–56 (1968)
2. Picard, R.W.: Affective computing. MIT Press, Cambridge (1997)
3. Picard, R.W.: Toward agents that recognize emotion. In: Proc. IMAGINA Conference, Monaco, pp. 153–155 (1998)
4. Li, H., Roivainen, P., Forcheimer, R.: 3D motion estimation in model-based facial image coding. IEEE Trans. Pattern Analysis and Machine Intelligence 15(6), 545–555 (1993)
5. Mase, K.: Recognition of facial expression from optical flow. IEICE Trans. E 74, 3474–3483 (1991)
6. Terzopoulos, D., Waters, K.: Analysis and synthesis of facial image sequences using physical and anatomical models. IEEE Trans. Pattern Analysis and Machine Intelligence 15(6), 569–579 (1993)
7. Yaccob, Y., Devis, L.: Recognizing human facial expression from long image sequences using optical flow. IEEE Trans. Pattern Analysis and Machine Intelligence 18(6), 636–642 (1996)
8. Bassili, J.N.: Emotion recognition: the role of facial movement and the relative importance of upper and lower areas of the face. J. Personality and Social Psych. 37, 2049–2059 (1979)
9. Essa, I.A., Pentland, A.P.: Coding, analysis, interpretation and recognition of facial expressions. IEEE Trans. on Pattern Analysis and Machine Intelligence 13(7), 715–729 (1997)
10. Essa, I.R.: Analysis, interpretation and synthesis of facial expressions by IR. PhD dissertation, Massachusetts Institute of Technology, Cambridge, MA (1995)
11. Gianluca, D., Bartlett, M.S., Hager, J.C., Ekman, P., Sejnowski, T.J.: Classifying facial actions. IEEE Trans. Pattern Analysis and Machine Intelligence 21(10), 974–989 (1999)
12. Essa, I.R., Pentland, A.P.: Coding, analysis, interpretation and recognition of facial expressions. IEEE Trans. Pattern Analysis and Machine Intelligence 19(7), 757–763 (1997)
13. Rosenblum, M., Yacoob, Y., Davis, L.S.: Human expression recognition from motion using a radial basis function network architecture. IEEE Trans. Neural Networks 7(5), 1121–1138 (1996)
14. Mase, K., Pentland, A.: Recognition of facial expression from optical flow. IEICE Trans. E (74), 408–410 (1991)
15. Lanitis, A., Taylor, C.J., Cootes, T.F.: Automatic interpretation and coding of face images using flexible models. IEEE Trans. Pattern Analysis and Machine Intelligence 19(7), 743–756 (1997)
16. Matsugu, M., Mori, K., Mitari, Y., Kaneda, Y.: Subject independent facial expression recognition with robust face detection using a convolution neural network. Neural Network 16(5-6), 555–559 (2003)
17. Gargesha, M., Kuchi, P.: Facial expression recognition using artificial neural networks (2002), `http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.4.8462` (accessed May 11, 2009)
18. Yeasin, M., Bullot, B., Sharma, R.: Recognition of facial expressions and measurement of levels of interest from video. IEEE Trans. on Multimedia 8(3), 1312–1324 (2006)
19. Black, M., Yacoob, Y.: Recognizing facial expressions in image sequences using local parameterized models of image motion. Intern. J. Computer Vision 25(1), 23–48 (1997)

Emotion Eliciting and Decision Making by Psychodynamic Appraisal Mechanism

J. Liu[1,2] and H. Ando[1,2]

[1] Cognitive Information Science Laboratories, ATR
[2] Universal Media Research Center,
National Institute of Information and Communications Technology (NICT), Kyoto, Japan
juanliu@nict.go.jp, h-ando@nict.go.jp

Abstract. In this paper we describe an architecture named Psychodynamic Cognitive Construction (PCC) for artificial systems to develop affective attitudes from the history of human-machine interaction. Meanwhile the evolving emotion states are used as an appraisal for decision making. Psychodynamic appraisal mechanism is proposed based on our understanding that emotion is a physically grounded, dynamically constructed, and subjectively experienced process. Different from other mechanisms, in PCC the emotion aroused by perception is not just a predefined mapping to be learned by the agent, but changes along with the cognitive development of the agent during interaction. The agent's knowledge and history bias its feeling. The emotional incentive makes the system trying to live for its own well-being and keep improving its constructs of the external world. The fluctuation of tensions, which is induced by innate needs or acquired anticipations, drives associative learning and expressive behaviors. A growing network is devised to memorize and retrieve its past experiences. A semi-embodied system, named qViki, is built as an implementation of PCC architecture. The preliminary experiment shows how the system's attitude towards certain stimulus is elicited and regulated by interaction. We discuss the role that emotion may play in affective agents as a motivation system to facilitate autonomous social learning.

1 Introduction

The pivotal role of emotion in reasoning and decision-making has been manifested by the study of neurological patients and widely accepted [1]. It quickly called the attention of AI community, challenging the traditional rational cognitive models that neglect or demote emotion. Affective computing [2], taking emotion as the core issue, aims to "give a computer the ability to recognize and express emotions, developing its ability to respond intelligently to human emotion, and enabling it to regulate and utilize its emotions" [3]. For sociable robots [4], to express emotions in a natural and intuitive way, such as emotive facial expression or gestures, provides important cues for maintaining and regulating human-robot interaction. Although affective capabilities can make robots appear more "life-like" and "believable", how

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 221–232.
springerlink.com

much can emotion contribute to the cognitive development of artificial systems? The machine psychodynamic paradigm [5, 6] seems to fill the gap by assuming the system keeps seeking pleasure gained from the rapid discharge of psychic tensions. Here, we extend the assumption to the notion that continuous cognitive development may be motivated by emotional needs, which are aroused by both innate organism's needs and acquired experiences. This paper discusses emotion elicitation and affective decision making from the perspective of developmental robotics [7], exploring an approach toward the long-term goal of open-ended life-long learning.

The rest of the paper is organized as follows: section 2 presents the key principles that guide the design of our emotion eliciting system, as well as the system architecture. Section 3 gives more details of the components. Section 4 introduces our affective agent qViki and preliminary experimental results. We conclude the paper with our future direction of research effort in Section 5.

2 Psychodynamic Appraisal Mechanism

2.1 Emotion Synthesis

Emotion synthesis methods in existing work [2, 8, 9] are mainly based on predefined drives or releasers. Emotions are mapped to sensori-motor conditions and represented as states in an affect space. It is convenient to integrate behavior, perception, drive and emotion in a unified way and observe the expression changes over time. However, people have more complex emotional feelings, which are not just aroused by innate drives and physical preferences. In the affect space, for one moment the system can only have one emotional state, but in daily life, some events may give rise to several simultaneous emotions, occurring as a succession or superposition of different emotions, masking of one emotion by another, etc. [10]. Moreover, emotion elicited by certain sensori-motor context may not be stable because of habituation, adaptation or more profound cognitive changes.

Another approach is proposed as psychodynamic model [6], in which pleasure is the result of sudden plummeting of a tension. As a consequence, sometimes an agent should intentionally put itself in an inconvenient or dangerous situation to get pleasure. This model sheds light on the dynamic property of emotion.

The model of emotion as naturally existing and objectively measurable is grounded in the laboratory behavioral sciences in which subjective experience is suspect, but this tradition does not address salient aspects of emotion as experienced in interaction [11]. Our emotional states keep evolving in our life time. To apply emotion as a self-motivation and appraisal mechanism for task-nonspecific, cumulative learning, we propose a mechanism that underscores emotion as physically grounded, interactively constructed and subjectively experienced.

2.2 Psychodynamic Appraisal Mechanism

In our model, emotion may be elicited from two levels: primary level with innate settings, e.g. aversive and appetitive stimuli, instincts or hardwired needs; and secondary level with acquired memories and knowledge, which are flexible. Both of them can be represented by states, while the changes and differences of states also generate affective feelings.

The primary level includes the emotional reactions wired in at birth. Certain features of perceived stimuli in the world or in the agent's body make the agent respond with an emotion and/or action. These innate responses are the basic mechanism, from which more associations between situations and emotions are built.

The secondary emotion is not reactive. The system uses its learning machine to make prediction. We call it "anticipation". According to Kelly's Personal Constructs Theory [12], human beings experience the reality from different perspectives and have various constructions. "A person's processes are psychologically channelized by the ways in which he anticipates events" [12]. People have their own constructions of the reality, use their knowledge and previous experience with similar events to generate anticipations of how ongoing events will unfold, test those expectations, and then improve their understandings of reality by adaptation process [13]. So the emotion aroused by anticipation is personal and experience-dependent, with its root in the primary emotion.

The process of anticipation is intrinsically affective. The biological body may provide pleasure and pain. To have anticipation is to have anxiety (anticipation of pain) or hope (anticipation of pleasure), and the result could be distress, surprise, delight, anger or disappointment, etc. Sometimes those feelings are mixed. The secondary emotion is the affective side of adaptation and learning when we encounter and resolve the inconsistency of our construction and the reality.

Therefore in our understanding, emotion is a process, rather than a body state. Emotion is elicited and differentiated when the state transits. Tensions themselves cause pain, but the fluctuation of tensions is the source of pleasure. An agent's anticipation influences the feeling it experiences, which depends on the cognitive constructs that represent its knowledge and past interaction with the reality. For the same artificial agent, it may feel differently encountering the same situation in different time if its construct has been modified. As a result the agent's behavior may change correspondingly. The psychodynamic appraisal is an ever changing evaluation mechanism. The agent strives to gain more pleasure during its life time, but the source of pleasure varies from time to time.

2.3 System Architecture

Fig. 1 gives a system architecture implementing psychodynamic appraisal mechanism, called Psychodynamic Cognitive Construction (PCC).

The system starts with primitive actions, fundamental perceptual abilities with innate awareness of the pleasantness of stimuli, and some basic tensions (loneliness, boredom, etc. released by "Innate Need Module" (INM)). This is the primary level of emotion synthesis. We say emotion is physically grounded because most of primary emotions can be defined by the body. For an artificial agent, certain features of sensory data are hardwired to generate positive or negative emotional responses, such as comfortable or painful contact, preferred sound, hungry feeling and so on. Those instincts that cannot be directly represented by sensory data are defined in INM as tensions.

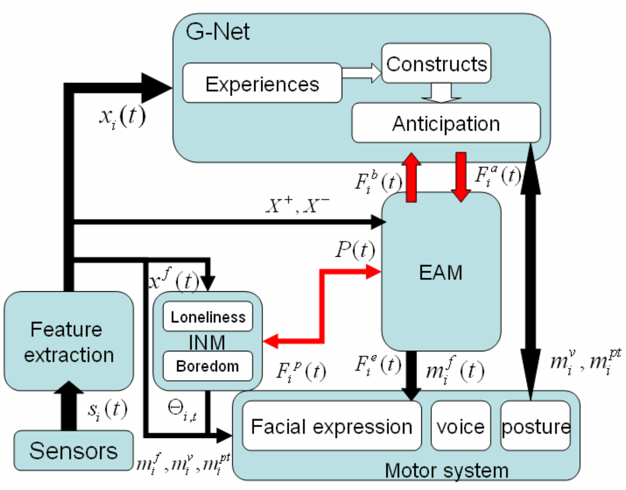


Fig. 1. An overview of psychodynamic cognitive construction agent's architecture. INM – Innate Need Module, EAM – Emotion Appraisal Module

To make the secondary level of emotion possible, "Emotion Appraisal Module" (EAM) and "Growing Network" (G-Net) module are incorporated in the architecture. G-Net builds cognitive constructions and produces anticipations using a memory of the experiences encountered by the agent. The past experience relevant to the concerns of the agent is stored in cells and their connections.

EAM appraises an event according to the bodily feeling (intrinsic pleasantness of sensory inputs), goal/need relevance (INM), its novelty and predictability (G-Net). A part of the outputs of EAM is provided to G-Net for learning emotionally important events and biasing behaviors. Another part is forwarded to motor system to express its affective states.

Initially G-Net has no function besides receiving stimuli from the sensors and other modules. The block acquires the association of its action, perception and anticipation when the agent interacts with the environment. If an event is emotionally influential enough, new cells and connections will be generated to memorize the event. However, if G-Net has been able to predict what would happen in a certain situation (where at least one cell's potential exceeds a threshold), it will not generate new cell, but adjust existing connections on account of the event. Events happening as expected will not have the same strong feeling as before. This influences the evaluation result of EAM. If the pleasure appraised by EAM can not satisfy the innate need of the agent, the agent will be driven to try some new actions.

The emotion process gets started and becomes differentiated when the agent starts to evaluate the significance of an unfolding event using its learned knowledge. The overall state of G-Net gives rise to "gut feelings" of desire or aversion to its current situation, while EAM provides conscious emotional feeling based on anticipation and reality. The internal representation is highly subjective and related to the agent's experiences. The system acts in the way that is promising to gain positive emotion and avoid negative one.

2.4 Emotion as a Motivation Mechanism

In PCC architecture, INM, EAM and G-Net form two processes between the input and output modules. One is the motion driven by INM and then the sensorimotor information and EAM evaluation are used for G-Net learning. The other is the anticipation and motion produced by G-Net that are fed to EAM for affective appraisal, and then influence the state of INM. The two flows are connected by EAM. Emotional states are used to drive acting and learning. Meanwhile they are elicited and differentiated during the learning process.

Without G-Net, the system can not anticipate the emotional impact of future rewards and punishment, i.e., it is not able to learn to adjust its action according to its experience. But the system still can receive stimuli and trigger emotional states through INM flow if its EAM is intact. However, without EAM, neither of the two flows works properly. The functions of G-Net and EAM are similar to those of the ventromedial prefrontal cortex and amygdala [14] in some aspects.

Emotion works as the appraisal of sensori-motor context. The acquired knowledge modulates the emotional status of the agent. Therefore there are two kinds of drives: biological needs and anticipations generated by the agent's construction of the world. These two drives are not separated. The construction is built during the agent's behaviors to fulfill needs and test anticipations. Anticipations may be motivated by an innate need. They two work together as our body and mind described in Damasio's Somatic marker hypothesis [1]. The two kinds of drives are both tensions in our psychodynamic appraisal mechanism. Different from some mechanisms driving the exploratory behaviors, such as intelligent adaptive curiosity [15], novelty [16], the EAM integrates the innate need and developed motivation in the framework of emotion.

In the next section, we will give a possible implementation of the mechanism.

3 Modules for Psychodynamic Cognitive Construction

3.1 Sensori-Motor Apparatus

A PCC agent has a number of sensors $s_i(t)$ gathering information from external world and its own body, which is preprocessed into $x_i(t)$ containing the features of interest, summarized as $X(t)$ (see Fig.1). Some of them intrinsically arouse feelings, i.e. $X^+(t)$ for positive features, and $X^-(t)$ for negative ones. There are also some unlabelled features $X^{null}(t)$, which will be discriminated and marked by the agent later. The agent's actions are controlled by a set of parameters $M(t)$. We use superscripts to show the types of action/motor parameters, such as $m_i^f(t)$ for facial expression, $m_i^v(t)$ for voice, $m_i^{pt}(t)$ for posture. The system has an internal clock that synchronizes the update of sensori-motor flow at every time step. At birth the system can present facial expressions showing its feeling to aversive or appetitive stimulus, randomly make phoneme sound or move its head if these actuators are motivated.

3.2 Innate Need Module

In this module, some instinct tensions are defined as a part of primary drives. A tension will accumulate in certain speed and discharge when the predefined condition is fulfilled. When a tension plummets, a pleasure signal is generated.
The level of *i*th tension $\Theta_{i,t}$ changes according to the following formula:

$$\Theta_{i,t+1} = \mathrm{G}(\Theta_{i,t} + {}^{1}\!/_{T_{S,i}} + {}^{A_{i,t}}\!/_{T_{A,i}} \text{ -}{}^{D_{i,t}}\!/_{T_{D,i}}, \varepsilon) \tag{1}$$

where:

- $G(a,b) = \begin{cases} 0 & \mathit{if}\; a \le b \\ a & \mathit{if}\; b < a < 1 \\ 1 & \mathit{otherwise} \end{cases}$; t – time steps; ε – lower limit (usually 0.01);
- $T_{S,i}$ – spontaneous increase/discharge time;
- $A_{i,t} = A_{i,0,t} + \cdots + A_{i,n,t}$ – accumulating signals, $T_{A,i}$ – tension increase time;
- $D_{i,t} = D_{i,0,t} + \cdots + D_{i,n,t}$ – discharging signals, $T_{D,i}$ – tension discharge time.

Pleasure $P(t)$ is related to the dynamics of tension discharge. The greater and the faster the tension discharged, the greater the initial intensity of the pleasure. When a tension level is decreased because of discharging signals, other than spontaneous reason, pleasure level increases. Otherwise it decays.

In PCC, we defined two kinds of primary tensions. One works as the basis of its social behavior, named “loneliness”. Another is for cumulative learning and curious behavior, named “boredom”. Since their discharging ports are connected to different sources, they function in two distinct ways in the system. Both of them accumulate slowly. The loneliness tension can be discharged if a human face image $x^f(t) \in X^0(t)$ is detected in the fovea. The discharging port of boredom tension is connected to the appraised pleasure output of EAM. The output of boredom tension is used to enable a random driver of voice and posture actuators. The pleasure generated by these two tensions is forwarded to EAM. We try to keep the primary level concise and not to add tensions that are not really fundamental.

3.3 Emotion Appraisal Module

EAM works as emotion elicitor for the affective agent. It takes dichotomous labeled sensory signals $X(t)$, pleasure output of INM $P(t)$ and anticipation produced by G-net $F_i^a(t)$ as input, appraising current situation $F_i^p(t)$ and generating emotional experience $F_i^e(t)$ of the agent. The internal affective feelings are expressed through facial expression controlled by a set of action parameters $m_i^f(t)$.

Appraisal is carried out dichotomously, i.e. evaluation is represented by two kinds of variables: positive (pleasure) $F_+^p(t)$ and negative (pain) $F_-^p(t)$ emotions.

$$F_+^p(t) = \textstyle\sum_{x_i \in X^+} x_i(t) + P(t) - F_+^a(t) = F_+^b(t) - F_+^a(t) \tag{2}$$

$$F_-^p(t) = \sum_{x_i \in X^-} x_i(t) - F_-^a(t) = F_-^b(t) - F_-^a(t) \tag{3}$$

Then it is clear that the emotion appraisal of a situation is not only decided by the body states and innate needs, but also the anticipation of the situation. The agent's expectation of positive/negative feeling will alleviate the effect of intrinsic feeling, since the agent has prepared for such a situation. Therefore, emotionally the consequence will not be so impact. However, if the anticipation is opposite to the fact, the feeling will be much stronger. If $F_+^p(t) < 0$, that means the agent is unsatisfied about the result. If $F_-^p(t) < 0$, the agent may feel relieved or lucky. $F_i^p(t)$ shows the subjective feeling of the agent. It is forwarded to INM's boredom tension. Insufficient fluctuation of this subject appraisal will drive the agent to try novel actions in order to gain new stimuli or interaction.

Emotional experiences $F_i^e(t)$ are calculated according to the value and relationship among $F_i^p(t)$, $F_i^b(t)$ and $F_i^a(t)$. In current experiments, six kinds of feelings are differentiated from the dichotomous appraisal: hopeful, elated, surprised, anxious, disappointed, and painful. The expressed emotion is a blend of feelings. In one time step, if $F_i^p(t)$ and $F_i^e(t)$ are not modified, they will decay similar to $P(t)$. Here the anticipation $F_i^a(t)$ is a critical factor for both emotion synthesis and appraisal. It is generated by G-net that equips the agent with a memory of past significant experiences.

3.4 G-Net

By saying "significant experiences", we emphasize that not all of the situations encountered by the agent are memorized. Only those experiences with fairly high emotional effect $F_i^b(t) > \delta$ will be used to update the growing network.

At the initial stage, the network has a set of sensory cells S for perceived features and motor primitives, two emotion cells $P = \{p^+, p^-\}$ for receiving pleasure and pain signal. Event cell set $E = \{E^+, E^-\}$ and action cell set M are empty. Newly created event cells are those generating anticipation signals for EAM. Learned actions are stored in M and activated by E.

The network's working procedure in one time step is:

1. Update cells' state $y_i(t)$ in S, P.

$$y_i(t) = \begin{cases} u_i & \mathit{if}\ u_i > 0 \\ G\left(\beta_i y_i(t-1), 0.01\right) & \mathit{if}\ u_i = 0 \end{cases} \tag{4}$$

where u_i is the input of cell i, $\beta_i < 1$ is its decay coefficient.

2. For cells in E, if $\exists i$ whose potential value $\varphi_i(t-1) > \theta$, i.e. last step the network anticipated that certain event represented by cell i would happen, then update the credit coefficient σ_i,

$$\sigma_i = \begin{cases} [\sigma_i + \varphi_i(t-1)] \big/ (\sigma_i + 1) & i \in E^+ \wedge y_{p^+} > \delta \text{ or } i \in E^- \wedge y_{p^-} > \delta \\ \sigma_i / (\sigma_i + 1) & i \in E^+ \wedge y_{p^-} > \delta \text{ or } i \in E^- \wedge y_{p^+} > \delta \end{cases} \tag{5}$$

By Equation (5), the credit coefficient σ_i of cell i is increased if the actual situation (pleasure or pain signal received by emotion cells) is consistent with the anticipated situation (pleasing or painful events), i.e. both of them are negative or positive. Otherwise, the credit value is decreased.
Then calculate $\varphi_i(t)$

$$\varphi_i(t) = \begin{cases} 0 & \text{if } \exists k \in H_i, y_k > 0 \\ G\left(\sum_{j \in C_i} \alpha_{ij} y_j \big/ \sum_{j \in C_i} \alpha_{ij}, 0\right) & \textit{otherwise} \end{cases} \tag{6}$$

where C_i is the excitatory presynaptic cell set, H_i is the inhibitory presynaptic cell set, α_{ij} is the strength of connection from j to i.

3. If $\exists i, \varphi_i(t) > \theta$, calculate the anticipation value ϕ_i of event cell $i \in E^\bullet$, where $\bullet \in \{+,-\}$, $\phi_i(t) = \sigma_i \alpha_{p^\bullet i} \varphi_i(t)$. If $\varphi_i(t) \le \theta$, $\phi_i(t) = 0$. $F_\bullet^a(t) = \sum_{i \in E^\bullet} \phi_i(t)$.

   Output action value for actuator z is $m_z(t) = \sum_{i \in C_z} \alpha_{zi} \varphi_i(t)$, if $F_+^a(t) > F_-^a(t)$, else $m_z(t) = -\sum_{i \in C_z} \alpha_{zi} \varphi_i(t)$. That means if the agent has a negative anticipation, the motivated action will be opposite to that produced in positive case. The emotional state influences decision making and behaviors of the agent.
4. If $\forall i \in E, \varphi_i(t) \le \theta \cap y_{p^\bullet}(t) > \delta$, generate new event cell $l \in E^\bullet$ and new action cell n if any actuator is working. The initial configuration is $\sigma_l = 0.1$; $\alpha_{p^\bullet l} = y_{p^\bullet}(t)$; for $k \in S$; if $y_k(t) > 0$, $\alpha_{lk} = y_k(t)$, $k \in C_l$; else $\alpha_{lk} = -1$, $k \in H_l$. For cell n, $\alpha_{nl} = y_h(t)$, where $h \in M$, $y_h(t-1) > \varepsilon$. There may be more than one action cell generated. By default, $\theta = 0.8$, $\delta = 0.5$, $\beta_i = 0.9$.

Introducing credit coefficient σ endows the agent with more delicate evolving process of emotion. The agent may feel most elated when it can predict something it is not sure about. It gains a lot of pleasure by testing its constructs. The growing rules that rank events as rewarding and quasi-rewarding situations according to temporal relations of events in our previous work [17] may also be applied here.

4 qViki System and Experiments

An artificial agent named qViki is built as a test-bed of our mechanism. qViki is a distributed system composed of a simulated part and physical parts (Fig. 2). The simulated part is mainly a face that is taken as a set of virtual actuators to express the emotion of the system and conduct communication with people. The physical part includes camera, microphone, speaker, tablet, touch sensor, and odor sensor.

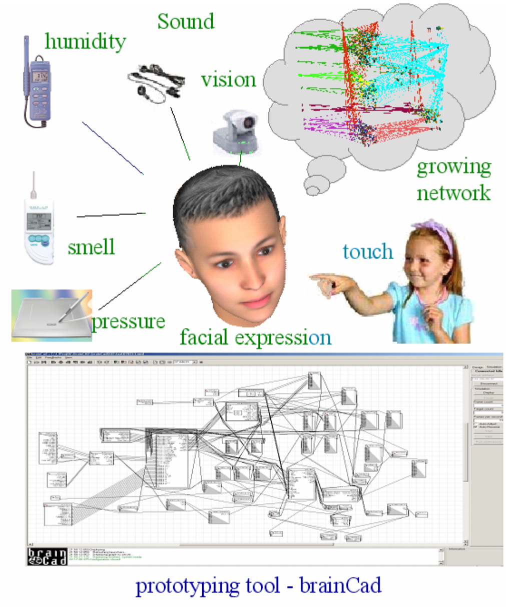
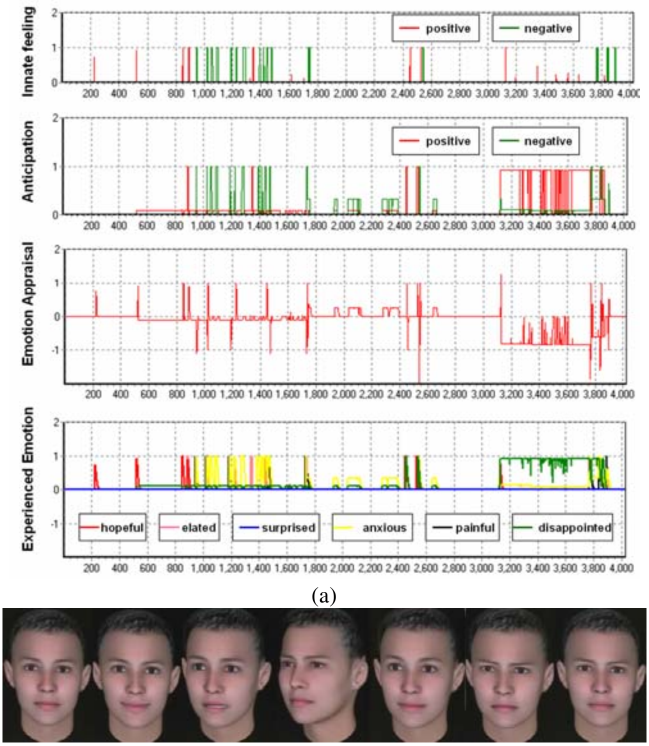


Fig. 2. Illustration of qViki system

Various actuators (e.g. remote controllable appliances, robotic arm, or mobile robot) can also be connected to sense the environment and manipulate objects. The whole system is build upon the software named brainCAD developed by our group for prototyping large-scale systems and running many modules in parallel on multiple workstations.

In our preliminary experiment, we observed how qViki's emotion is elicited and evolved through interaction, how it affectively appraises a situation by its experience and how the anticipation drives its action. A typical example is the varying attitude toward a human face. Fig. 3 shows a result of the experiment.

At the beginning qViki was indifferent to this stimulus. In the first stage of interaction, qViki learned positive anticipation to human face (second curve) because face was helpful to discharge the innate tension "loneliness", so she tended to track human face when seeing it. Since loneliness tension was innate and homeostatic, qViki should keep holding a positive anticipation to a human face if no other thing happens. But when the usual tender stroke became a hit or squeeze, qViki's surprising and painful facial expression showed that she suddenly realized her construct of human was deficient. After a sequence of unpleasant interactions, she started to generate negative anticipation and her action was changed. She avoided human face, instead of tracking it! This change of action was not predefined. However, the loneliness tension and later comfortable experiences changed her mind, again. But she looked more sophisticated than before, since the presence of a human aroused a lot of memories. G-net keeps recording and retrieving memories on-line. The synthesized emotion played an important role in decision-making and social learning in the artificial agent, like human emotion system, not only for expressive representation of internal states.



(a)

(b)

Fig. 3. A typical run of the human presence experiment: (a) Curves of innate feeling from sensors and INM, anticipation from G-net, emotion appraisal signal ($F_+^p(t) - F_-^p(t)$) and experienced emotion generated by EAM. (b) Facial expressions in corresponding stages, showing the evolution of qViki's attitude towards the same stimulus – human face via interaction

It is noticeable that the role of human-machine interaction is critical in the developmental scenario. Here a human being is a part of the external world that the agent intends to construct a mental model for. The PCC system seems suitable for social learning. qViki is like a naïve child, completely expressing her subjective feelings on her face. People may easily evaluate the validity of the emotion elicitation and differentiation model by direct communication with the face. They can also adjust their way of interaction to forge a desired appraisal mechanism in qViki. The emotional behaviors of qViki are not predefined to be carried out under corresponding conditions; instead, they are learned and keep changing.

5 Conclusion and Future Work

The brain systems that are jointly engaged in emotion and decision-making are generally involved in the management of social cognition and behavior. Similarly in robotic systems, emotion should also have some cognitive significance. In our system, the emotion expression is not the purpose of the implementation, but a way to facilitate cognitive development and social interaction. Emotion experiences are getting abundant along with the mental development. While the agent learns more about the world, her feeling changes. And it drives her to experience more. That is the main difference between psychodynamic appraisal mechanism and the work of [2, 8, 9, 18, 19]. Emotion is needed for mastering the knowledge behind proper social behavior, and also required for the deployment of proper behavior [1]. A key problem for scaling up the system is to find a set of essential laws of G-Net's growth that would result in efficient learning even in more complicated domains. The feature extraction part may also be involved in psychodynamic construction and influenced by the agent's subjective model of the reality as the top-down process in human perception flow.

Acknowledgments

The authors would like to thank Dr. Andrzej Buller for his crucial support and inspiration. We are also grateful to Michal Joachimczak and Beata Grzyb for their help in building the experimental system.

References

1. Damasio A (1994) Descartes' error: emotion, reason, and the human brain. Gosset, Putnam, New York
2. Picard R (1997) Affective Computing. MIT Press, Boston
3. Cañamero L (2005) Emotion understanding from the perspective of autonomous robots research. Neural Networks 18 :445–455
4. Breazeal C, Brooks A, Gray J, Hoffman G, Kidd C, Lee H, Lieberman J, Lockerd A, Chilongo D (2004) Tutelage and collaboration for humanoid robots. Int J of Humanoid Robotics 1 (2):315–348
5. Buller A (2002) Psychodynamic robot. In: Proc 2002 FIRA Robot World Congress, Seoul, pp 26–30
6. Buller A (2006) Machine psychodynamics: toward emergent thought. ATR Technical Report, TR-NIS-0005
7. Weng J, McClelland J, Pentland A, Sporns O, Stockman I, Sur M, Thelen E (2000) Autonomous mental development by robots and animals. Science 291 (5504):599–600
8. Breazeal C (2003) Emotion and sociable humanoid robots. Int J Human-Computer Studies 59 :119–155
9. Velásquez J (1998) When robots weep: emotional memories and decision-making. In: Proc AAAI-98 Conference, MadisonWI, pp 70–75
10. Martin JC, Niewiadomski R, Devillers L (2006) Multimodal complex emotions: gesture expressivity and blended facial expressions. Int J of Humanoid Robotics 3 (3):269–291
11. Boehner K, DePaula R, Dourish P, Sengers P (2007) How emotion is made and measured. Int J Human-Computer Studies 65 :275–291

12. Kelly G A (1991) The psychology of personal constructs. Vol. I, II. Norton, New York. 2nd edn, Routledge, London/New York
13. Boeree C G (1991) Causes and reasons: the mechanics of anticipation. http://www.ship.edu/~cgboeree/anticipation.html. Accessed 23 February 2009
14. Naqvi N, Shiv B, Bechara A (2006) The role of emotion in decision making: a cognitive neuroscience perspective. Current Directions in Psychological Science 15 (5):260–264
15. Oudeyer PY, Kaplan F, Hafner V (2007) Intrinsic motivation systems for autonomous mental development. IEEE Trans on Evolutionary Computation 11 (1):265–286
16. Huang X, Weng J (2002) Novelty and reinforcement learning in the value system of developmental robots. In: Prince C, Demiris Y, Marom Y, Kozima H, Balkenius C (eds) Proc 2nd International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems, Lund University Cognitive Studies 94 :47–55
17. Liu J, Buller A, Joachimczak M (2006) Self-motivated learning agent – skill-development in a growing network mediated by pleasure and tensions. Trans of the Institute of Systems, Control and Information Engineers (Trans of ISCIE), 19 (5):169–176
18. Broekens J, DeGroot D (2004) Scalable and flexible appraisal models for virtual agents. In: 5th Game-On International Conference: Computer Games, Artificial Intelligence, Design and Education (CGAIDE 2004), Reading, UK, pp 208–215
19. Haikonen PO (2007) Robot brains: circuits and systems for conscious machines. Wiley & Sons Inc, West Sussex, England

Positing a Growth-Centric Approach in Empathic Ambient Human-System Interaction

R. Legaspi[1], K. Fukui[1], K. Moriyama[1,2], S. Kurihara[1,2], and M. Numao[1,2]

[1] The Institute of Scientific and Industrial Research, Osaka University, Osaka, Japan
{roberto,fukui,koichi,kurihara,numao}@ai.sanken.osaka-u.ac.jp
[2] Department of Information Science and Technology, Osaka University, Osaka, Japan

Abstract. We define our empathy learning problem as determining the extent by which a system can perceive user affect and intention as well as ambient context, construct models of these perceptions and of its interaction behavior with the user, and incrementally improve on its own its models in order to effectively provide empathic responses that change the ambient context. In concept, system self-improvement can be viewed as changing its internal assumptions, programs or hardware. In a practical sense, we view this as rooting from a data-centric approach, i.e., the system learns its assumptions from recorded interaction data, and extending to growth-centric, i.e., such knowledge should be dynamically and continuously refined through subsequent interaction experiences as the system learns from new data. To demonstrate this, we return to the fundamental concept of affect modeling to show the data-centric nature of the problem and suggest how to move towards engaging the growth-centric. Lastly, given that an empathic system that is ambient intelligent has yet to be explored, and that most ambient intelligent systems are not affective let alone empathic, we submit for consideration our initial ideas on an empathic ambient intelligence in human-system interaction.

1 Introduction

Recent advances in human computing have resulted in significant enhancement of user experience in unprecedented fashion (e.g., [1, 2]). These involve projecting the human user into the foreground while pushing the computing technology to infuse itself unobtrusively into the background (human-centered design) and dealing with the difficult issues surrounding the front end, i.e., to understand human behaviors, specifically, affective and social signaling, through visible behavioral cues, and back end of human computing, i.e. anticipating human responses and then selecting, after deliberation, emulated human behaviors as system responses to meet user needs [3]. Dealing with the front and back end problems equates to engaging issues in user modeling and adaptive user interfaces, respectively.

Instead of more traditional user models that describe the cognitive processes, skills or preferences of the user [4], there has been a significant shift to model human emotional-social processes in intelligent systems that educate [5-9], comfort

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 233–244.
springerlink.com

or counsel [10-15], provide health care [16-18], aid children with special needs [19-21], game users [22, 23], or for music listening [24-28]. We posit, however, that a system obtains not only the perception of user affect, but include deliberations based on user intention and/or situational context (state of the surrounding environment) when providing perception-based user-centered responses.

Most of these existing affective social systems employ synthetic agents and are commonly confined in a sit-down set-up. It is to our knowledge that a system that has the features of ambient intelligence while being empathic has yet to be explored, and that the counter is equally compelling, that most ambient intelligent systems are not affective let alone empathic.

Moreover, existing interactive systems are implemented with either theory-centric or data-centric knowledge bases [29]. The former is a top-down approach that uses AI methods to apply domain expert knowledge while the second is bottom-up that uses huge amounts of recorded traces of human-system interactions to acquire generalizations of empathic behaviors or stereotypes using techniques in Machine Learning. Fig. 1 illustrates these two approaches to empathy modeling (with our proposed approach indicated in dotted lines).

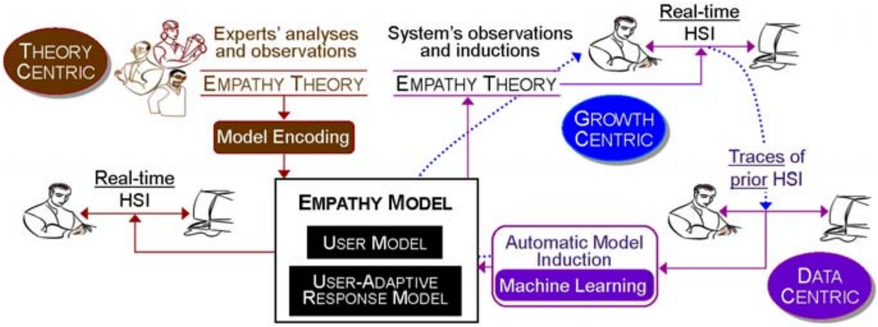


Fig. 1. Theory- data- and growth-centric approaches to empathy modeling

We reiterate here the tendency of theory- and data-centric approaches to fall short of optimal performance. The knowledge engineering required for the former is costly especially in novel applications where expertise is scarce and lackluster when expert analyses cannot be detailed directly into the machine. Furthermore, if the implementation is theory-specific, it can render the knowledge inflexible and not scalable. Lastly, most existing systems in their rule-based implementations are unable to adapt dynamically to never before encountered situations due to their preconditioned or scripted knowledge. These problems motivate constructing interactive systems that can generalize from human-system interaction data empirically grounded models of empathy. This approach, however, is impeded by the strict demand for a labeled dataset that needs to be huge to acquire a viable generalization. Problems arise due to an abundance of unlabeled data as credible annotators are costly to acquire, target subjects are unavailable or unwilling to participate, labeling mechanisms fail, or noises invalidate recorded data portions.

We posit a *growth-centric*, i.e. *self-improving*, approach to empathy learning in an ambient intelligent system. Research works have shown that the cognitive and perceptual abilities required for empathy mature in time and through experience (noteworthy citations in [29]). Emulating this in the realm of human-system interaction, an empathic machine learner may be allowed to acquire an initial knowledge that is partial, i.e., only a subset of the set of all possibly existing cases in the real world, permit it to be inaccurate and imperfect, but expect it to continuously improve its knowledge on its own each time it interacts with the user, hence, maturing or growing over time. This growth-centric approach is data-centric for a real-time incremental learning of empathy models (Fig. 1).

To conclude this section, we define our growth-centric empathy learning problem as determining the extent by which an empathy machine learner can intelligently perceive user and ambient context states and incrementally learn models of its perceptions and of its user interaction behavior patterns as means of improving its empathic responses. Although we have already elucidated in [29] our larger motivation for this growth-centric approach and our proposed learning framework for the empathic learner to self-improve its ambient responses based on the perception of affect, intention and behavior patterns, we deem it necessary to step back and expound on how compelling the problem is by drawing insights from our prior work, and from the result of an experiment showing the data-centric root of the problem suggest how to move towards engaging the growth-centric challenge.

2 Making the Case for a Growth-Centric Empathic System

We reckon that our empathy learner can achieve a growth-centric learning if it can demonstrate incremental learning. To illustrate the need for this capability, we return to the fundamental concept that underlies emotional-social computing, i.e., affect modeling, and report the results of our initial attempt to approximate continuous affect measures.

We collected user physiological readings using six wearable sensors[1] while one subject listened to affect-eliciting songs and sounds for about 27 minutes. The sensors measured blood volume pulse, abdominal respiratory movements, skin responses (conductance, temperature, and rate at which heat is dissipating from the body), and movements of the corrugator and left/right masseter. Evidences exist that affective states can be inferred from these features (e.g., [30-33]). From the collected readings, feature vectors were extracted, with each vector containing 51 attribute values. The software used to run these hardware provide facilities for the experimenter to extract the desired features (e.g., mean spectral values within an epoch of 20 ms). To label these vectors with affective states, we used another device[2] that uses an emotion spectrum analysis algorithm to induce from brainwave

[1] Acquired from Thought Technology (http://www.thoughttechnology.com/) and SenseWear BodyMedia (http://www.sensewear.com/solutions_bms.php)

[2] Acquired from Brain Functions Laboratory, Inc.(http://www.bfl.co.jp/english/top.html)

signals user affective states. We have discussed this algorithm and its outputs relative to our research in [28]. We have used this device to replace the semantic differential-based instrument we used in [24, 27] to more effectively measure the influence of music on user affect. The emotion readings provided by this device are novel for two reasons, namely, the affective states are measured in separate dimensions thereby capturing existing mixed emotions and the affect values are continuous as opposed to discrete and mutually exclusive (i.e., the presence of one precludes the others) emotion categories used by the majority of related research. Each affect dataset contains 3,058,658 labeled instances. From each dataset, training and test datasets were constructed for a leave-one-out two-fold cross validation of the model. Furthermore, two versions of each dataset were constructed, one using all 51 attributes and the other using a subset of features.

The primary objective of this experiment is to prune the initial feature set to contain only features that are relevant (their correlation with the continuous class labels is high), non-redundant (have low correlation with other features) [34] and effective (with accurate predictive function) when approximating user affect values. We contrast this to the majority of research that use frequently investigated features without attempting to explore other possibly affect-correlated features. Furthermore, the tendency for other research works is to employ feature extraction techniques that transform existing features into a lower dimensional space. The problem with feature extraction is that it can allow certain characteristics of the data to be lost and/or allow parts of noisy features to be inherited during transformation [35]. Moreover, we intend to employ the sensors corresponding to the subset of effective features as alternatives to the emotion spectrum analyzer whose readings we have been approximating, in effect, finding cheaper means of emotion detection. We leveraged the effectiveness of sequential forward floating selection [36] with a correlation-based filter [34] to select the subset of optimal features. For the regression task, we employed a support vector machine [37] with a linear kernel function. We refer the reader to these references for detailed discussions. We implemented the regression task in two instances, varying the amount of data to train the support vector machine (as well as the data to test the predictive models), i.e., 10% (train: 305,864, test: 2,752,794 instances) and 50% (train and test: 1,529,329 instances) of the total data.

It can be seen in Tables 1 and 2[3] that the feature subsets dynamically changed depending on the amount or content of the information read for three affect dimensions. The values in parentheses indicate that the number of needed sensors decreased from the original six. Later on, experts may need to look into these results and validate the learned relations. For example, there is evidence that skin conductivity is a general indicator of stress [32]. This appears in Table 1, which means that the model learner was able to discover existing real-world knowledge based on the data alone. The immediate question here, however, is whether these learned feature subsets are highly predictive of the approximated affective states.

[3] BVP-blood volume pulse, HRV-heart rate variability, EMG-electromyograph, VLF/LF/ HF-HRV standard frequency bands. For “%power”, the software algorithm takes an FFT or power spectrum channel as input and determines the moment-to-moment percentage of total power within the spectrum. We refer the reader to the manufacturers’ website for the details.

Table 1. Learned feature subsets from a 10%(train)×90%(test) dataset

Stressed (1 sensor)	*Sad* (1 sensor)	*Relaxed* (2 sensors)
Skin conductance mean	Heat flux mean	BVP VLF %power
	Skin temperature mean	BVP LF total power epoch mean
		BVP LF/HF epoch mean
		Respiration epoch mean
		HRV standard deviation
		BVP peak freq. mean (Hz)
		BVP LF %power mean
		BVP HF %power mean

Table 2. Learned feature subsets from a 50%(train)×50%(test) dataset

Stressed (1 sensor)	*Sad* (4 sensors)	*Relaxed* (2 sensors)
Skin temperature mean	BVP HF total power epoch mean	BVP HF total power epoch mean
	EMG epoch mean	Respiration epoch mean
	Respiration epoch mean	HR max-min mean
	HR max-min mean	Abdominal amplitude mean
	Abdominal amplitude mean	
	Heat flux mean	

Fig. 2 shows the accuracy of the predictive models within an estimation error bandwidth of 0.1, i.e., the predictive value is correct if it differs with the actual test value by $\leq$ bandwidth. Table 3 shows the average estimation error between all predicted and actual test values. This shows that for some cases the bandwidth can still be lowered towards the average and still obtain satisfying results. The chart shows the effectiveness of the selected features as these increased the accuracy of the models by an average of 31 percent. It also shows that in using all 51 attributes, the tendency is for the model to become more accurate with the increase in data. This does not seem to hold, however, with the feature subsets. We suspect that this tendency to be less accurate can be attributed to overfitting as the learning algorithm is trained too many times on the same data.

Though we cannot claim that the results here are conclusive since our methodology needs to progress (e.g., experiment with more subjects, perform 10-fold cross validation to account for all tendencies in the data, etc.), it is fair to suggest that the results show traces that to learn incrementally and effectively, a system must consider learning from few examples, intelligently select optimal features and account for various tendencies in the data as the amount of data increases. A dynamic attribute selection will prove useful in solving problems that might emanate from high dimensionality of data (i.e., large number of features) that usually impedes the machine learning process or from changing contexts due to changes in the problem domain or in the physical environment [29]. There should be mechanisms to include new useful features that can bring about new relations and remove noisy features, which is analogous to unlearning wrong ways of relating. Several works in emotion recognition have employed dynamic feature selection (e.g., [38-40]). There is a need, however, to investigate these in light of the various detection schemes of our self-improving empathy learner.

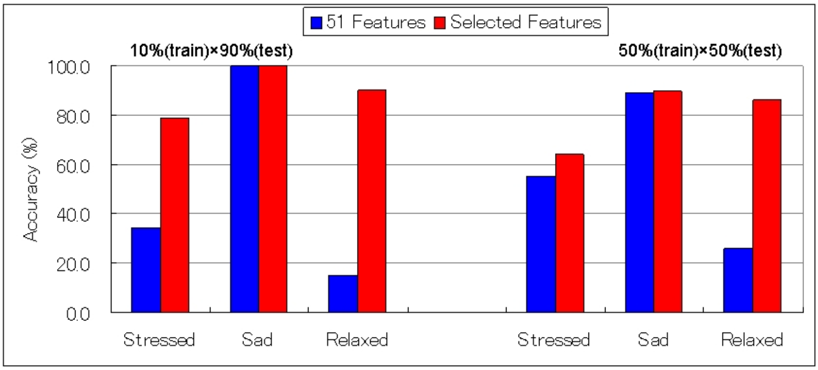


Fig. 2. Performance evaluation of the learned models

Table 3. Average estimation error when two-fold cross validation was performed

	10%(train)×90%(test)			50%(train)×50%(test)		
	Stressed	Sad	Relaxed	Stressed	Sad	Relaxed
51 features	0.152	0.021	0.189	0.222	0.097	0.118
Selected features	0.076	0.031	0.049	0.138	0.058	0.064

3 Making the Case for an Empathic Ambient Intelligent System

We build our case on the point that finding relations between user affective and behavioral patterns from whole body movements is still an open problem. Although there are evidences to suggest that motion properties (e.g., acceleration, velocity, etc.) of human body parts (e.g., upper body, limbs, torso) are manifestations, hence predictive cues, of emotions and changes in emotional intensity [41, 42], generalizations about the movement characteristics associated with affect remain insufficient due to issues in the methods by which these generalizations have been observed [42]. Most prior works on emotion analyses use data from induced or acted emotions [42, 43], hence deliberate and unnatural. Furthermore, most studies done to detect movement-based affect utilized body markers to facilitate easier segmentation of which body parts signal particular types of emotions. Second, although there are ubiquitous systems that detect normal and novel behaviors in everyday living [44], and use such knowledge to predict other needs of the user [45], most of such systems are not empathic on the basis of behavior patterns. Our challenge in this regard is for the system to deliberate on empathic responses

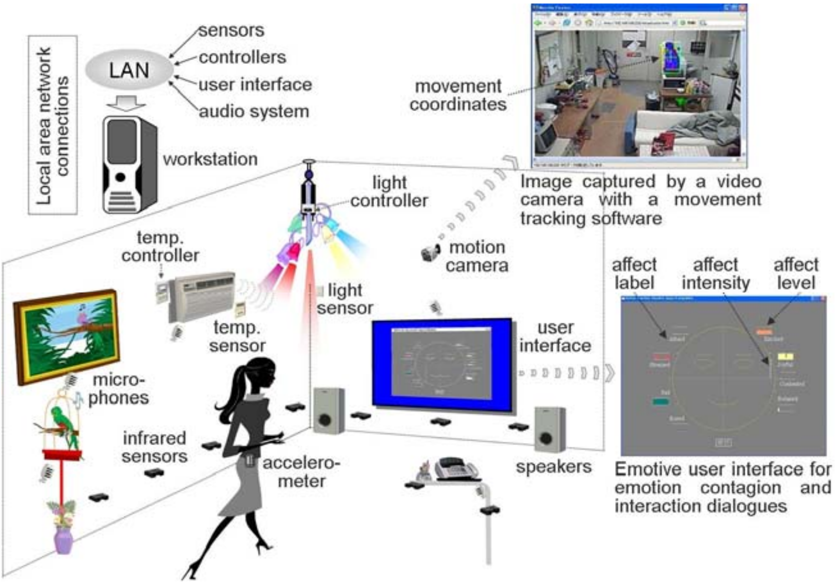


Fig. 3. Proposed empathic space for daily living

based on perceived real-life user features, i.e., affect and behavior that spontaneously surface in natural contexts such as everyday situations, without obtrusive means and without the need for body markers. It has been shown, for example, that habitual patterns [46] and high-level behaviors [47] can be determined even with the use of low-level sensors (e.g., infrared sensors, GPS).

Fig. 3 illustrates our proposed empathic ambient space with the devices that will comprise it. What is essential is for our empathic space to capture the dynamic flow of natural human-system interaction, i.e., it can provide immediate response to current user states given the context that the user is in which requires the user and context states being interpreted, and the system responses being deliberated and modified when necessary, in real-time; and permit mobility of the user in the space. We have detailed in [29] our learning framework that may (1) permit mapping affective states to their expressions, (2) determine causes of emerging affect which include the user goal and behavior patterns and the state of the environment, and infer empathic actions optimal for (1) and (2). At this conceptual stage, we have identified the system empathic responses as changing light color and intensity, adjusting room temperature, and modifying ambient audio expressions. The difficulty is to permit online learning of the empathic response model. Is there motivation for us to believe that online self-improvement is feasible?

The advantage of online learning is that training instances are continuously processed upon arrival, consequently updating the concept theory that covers all seen instances [48]. Online variants of machine learning algorithms are available for decision trees, naïve Bayes models, and nearest-neighbor classifiers [48], for adversarial setting, online agnostic learning, and learning a drifting target concept [49]. There are also online algorithms for active [50, 51]) and semi-supervised [52, 53]) learning and for real-time feature selection [54-56]). Semi-supervised [57, 58] and active learning [59, 60] are important and useful techniques when labeled data are hard to obtain while there is an abundance of unlabeled data. Their combination also proved useful [58, 61, 62]. Another ML technique that can cope with online experiential learning is reinforcement learning that relies on experience to discover a policy for acting optimally in given situations [63]. Reinforcement learning clearly suits unknown or changing contexts. Furthermore, a model of system empathic dynamics that generates simulated experiences [63] and/or similar cases [64] can be used for planning to supplement real-time direct learning from experience. A notable progress has been demonstrated recently by a reinforcement learning algorithm that requires only an approximate model and only a few real-life training samples to achieve a near-optimal performance in a deterministic environment [65]. Although these approaches have found many useful applications, these still need to be investigated and tried in the area of human-system interaction.

4 Conclusions

The main goal of this paper is to put forth for consideration a growth-centric learning process when implementing an ambient intelligent system that has the capacity to be empathic. Addressing the issues when developing this process calls for the system to have notions of user and ambient states recognition and provisions of empathic support based on that perception, as well as provisions of machine learning techniques that can support the system's self-improvement of its knowledge when interacting with the user in order to overcome the restrictions imposed by inaccurate approximated models and in real-time.

We have also shown that the challenge to achieve a growth-centric empathy learning stems from addressing the problems rooted in the data-centric, i.e., to find means for the system to efficiently learn when labeled data is scarce, and refine its feature set dynamically because of the presence of changing contexts.

Acknowledgment

This research is supported in part by the Global COE (Centers of Excellence) Program of the Ministry of Education, Culture, Sports, Science and Technology, Japan and by the Japan Society for the Promotion of Science.

References

1. Philips Research, `http://www.research.philips.com/technologies/projects/ambintel.html` (accessed May 15, 2008)
2. Wasserman, K.C., Eng, K., Verschure, P.F.M.J., Manzolli, J.: Live soundscape composition based on synthetic emotions. IEEE Multimedia 10(4), 82–90 (2003)

3. Pantic, M., Pentland, A., Nijholt, A., Huang, T.: Human computing and machine understanding of human behavior: a survey. In: Proc. 8th Intern. Conference Multimodal Interfaces, pp. 239–248 (2006)
4. Webb, G.I., Pazzani, M.J., Billsus, D.: Machine learning for user modeling. User Modeling and User-Adapted Interaction 11(1-2), 19–29 (2001)
5. Picard, R.W., Papert, S., Bender, W., Blumberg, B., Breazeal, C., Cavallo, D., Machover, T., Resnick, M., Roy, D., Strohecker, C.: Affective learning – amanifesto. BT Technology J. 22(4), 253–269 (2004)
6. Paiva, A., Dias, J., Sobral, D., Aylett, R., Woods, S., Hall, L., Zall, C.: Learning by feeling: evoking empathy with synthetic characters. Applied Artificial Intelligence 19(4), 235–266 (2005)
7. Wang, H., Yang, J., Chignell, M., Ishizuka, M.: Empathic multiple tutoring agents for multiple learner interface. In: Proc. IEEE/WIC/ACM Intern. Conference Web Intelligence and Intelligent Agent Technology, pp. 339–342 (2006)
8. Zakharov, K.: Affect recognition and support in intelligent tutoring systems. MS Thesis, Dept. of Computer Science and Software Engineering, Univ. of Canterbury, New Zealand (2007)
9. D'Mello, S., Jackson, T., Craig, S., Morgan, B., Chipman, P., White, H., Person, N., Kort, B., El Kaliouby, R., Picard, R.W., Graesser, A.: (2008), http://affect.media.mit.edu/publications.php (accessed May 15, 2009)
10. Klein, J., Moon, Y., Picard, R.W.: This computer responds to user frustration: theory, design and results. Interacting with Computers 14, 119–140 (2002)
11. Bickmore, T.W., Picard, R.W.: Towards caring machines. In: Proc. ACM SIGCHI Conference Human Factors in Computing Systems, pp. 1489–1492 (2004)
12. Paiva, A., Dias, J., Sobral, D., Woods, S., Aylett, R., Sobreperez, P., Zoll, C., Hall, L.: Caring for agents and agents that care: building empathic relations with synthetic agents. In: Proc. 3rd Intern. Joint Conference on Autonomous Agents and Multiagent Systems, pp. 194–201. ACM Press, New York (2004)
13. Brave, S., Nass, C., Hutchinson, K.: Computers that care: Investigating the effects of orientation and emotion exhibited by an embodied computer agent. Intern. J. Human-Computer Studies 62(2), 161–178 (2005)
14. Bickmore, T., Schulman, D.: Practical approaches to comforting users with relational agents. In: Proc. ACM SIGCHI Conference on Human Factors in Computing Systems, pp. 2291–2296 (2007)
15. McQuiggan, S.W., Lester, J.C.: Modeling and evaluating empathy in embodied companion agents. Intern. J. Human-Computer Studies 65(4), 348–360 (2007)
16. Lisetti, C., Nasoz, F., LeRouge, C., Ozyer, O., Alvarez, K.: Developing multimodal intelligent affective interfaces for tele-home health care. Intern. J. Human-Computer Studies 59, 245–255 (2003)
17. Bickmore, T., Gruber, A., Picard, R.: Establishing the computer-patient working alliance in automated health behavior change interventions. Patient Education and Counseling 59(1), 21–30 (2005)
18. Liu, K.K., Picard, R.W.: Embedded empathy in continuous, interactive health assessment (2005), doi:10.1017/S1351324908004956
19. El Kaliouby, R., Robinson, P.: The emotional hearing aid: An assistive tool for children with Asperger syndrome. In: ACII, pp. 582–589 (2005)
20. Teeters, A.C.: Use of a wearable camera system in conversation: Toward a companion tool for social-emotional learning in autism. MS Thesis, Media Arts and Sciences, School of Architecture and Planning, MIT, Cambridge, MA (2007)

21. Lee, C.H., Kim, K., Breazeal, C., Picard, R.W.: Shybot: friend-stranger interaction for children living with autism (2008), `http://affect.media.mit.edu/pdfs/08.lee-etal-shybot-chi.pdf` (accessed May 15, 2009)
22. Johnson, D., Wiles, J.: Effective affective user interface in design games. Ergonomics 46(13-14), 1332–1345 (2003)
23. Boukricha, H., Becker, C., Wachsmuth, I.: Simulating empathy for the virtual human Max. In: Proc. 2nd Intern. Workshop on Emotion and Computing, Osnabruek, Germany, pp. 23–28 (2007)
24. Numao, M., Takagi, S., Nakamura, K.: Constructive adaptive user interfaces – Composing music based on human feelings. In: Proc. National Conference on Artificial Intelligence, pp. 193–198. AAI Press (2002)
25. Kim, S., Andre, E.: Composing affective music with a generate and sense approach (2004), `http://mm-werkstatt.informatik.uni-augsburg.de/Elisabeth-Andre.html` (accessed May 15, 2009)
26. Dornbush, S., Fisher, K., McKay, K., Prikhodko, K., Segall, Z.: XPod – human activity and emotion aware mobile music player. In: Proc. 2nd Intern. Conference on Mobile Technology, Applications and System, Augsburg, Germany, pp. 1–6 (2005)
27. Legaspi, R., Hashimoto, Y., Moriyama, K., Kurihara, S., Numao, M.: Music compositional intelligence with an affective flavor. In: Proc. ACM Intern. Conference on Intelligent User Interfaces, pp. 216–224 (2007), doi:10.1145/1216295.1216335
28. Sugimoto, T., Legaspi, R., Ota, A., Moriyama, K., Numao, M.: Modeling affective-based music compositional intelligence with the aid of ANS analyses. Knowledge Based Systems 21(3), 200 (2008)
29. Legaspi, R., Kurihara, S., Fukui, K.-I., Moriyama, K., Numao, M.: Self-improving empathy learning. In: Proc. 5th IEEE Intern. Conference on Information Technology and Applications (ICITA 2008), pp. 88–93 (2008)
30. Nasoz, F., Alvarez, K., Lisetti, C.L., Finkelstein, N.: Emotion recognition from physiological signals using wireless sensors for presence technologies. Cognition, Technology and Work 6(1), 4–14 (2004)
31. Branco, P., Encarnacao, L.M.: Affective computing for behavior-based UI adaptation. In: Proc. Workshop Behavior-based User Interface Customization, Intern. Conf. Intelligent User Interfaces (2004)
32. Healey, J.A., Picard, R.W.: Detecting stress during real-world driving tasks using physiological sensors. IEEE Transactions on Intelligent Transportation Systems 6(2), 156–166 (2005)
33. Wilhelm, F.H., Pfaltz, M.C., Grossman, P.: Continuous electronic data capture of physiology, behavior and experience in real life: Towards ecological momentary assessment of emotion. Interacting with Computers 18(2), 171–186 (2006)
34. Hall, M.A.: Correlation-based feature selection for discrete and numeric class machine learning. In: Proc. 17th Int. Conf. Machine Learning, pp. 359–366 (2000)
35. Wagner, J., Kim, J., André, E.: From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification. In: IEEE Int. Conf. Multimedia and Expo., pp. 940–943 (2005)
36. Pudil, P., Novovičová, J., Kittler, J.: Floating search methods in feature selection. Pattern Recognition Letters 15(11), 1119–1125 (1994)
37. Joachims, T.: Making large-scale SVM learning practical. In: Advances in Kernel Methods: Support Vector Learning, pp. 169–184 (1999)

38. Petrushin, A.: Emotion recognition agents in real world. In: Proc. AAAI Symposium on Socially Intelligent Agents: Human in the Loop (2000)
39. Amershi, S., Conati, C.: Unsupervised and supervised machine learning in user modeling for intelligent learning environments. In: Proc. 12th Int. Conf. Intelligent User Interfaces, pp. 72–81 (2007)
40. Alvarez, A., Cearreta, I., Lopez, J.M., Arruti, A., Lazkano, E., Sierra, B., Garay, N.: Application of feature subset selection based on evolutionary algorithms for automatic emotion recognition in speech. In: Proc. Int. Conf. Non-Linear Speech Recognition, pp. 273–281 (2007)
41. Atkinson, A.P., Dittrich, W.H., Gemmell, A.J., Young, A.W.: Emotion perception from dynamic and static body expressions in point-light and full-light displays. Perception 33(6), 717–746 (2004)
42. Crane, E.A., Gross, M.: Motion capture and emotion: Affect detection in whole body movement. In: Paiva, A.C.R., Prada, R., Picard, R.W. (eds.) ACII 2007. LNCS, vol. 4738, pp. 95–101. Springer, Heidelberg (2007)
43. Devillers, L., Vidrascu, L., Lamel, L.: Challenges in real-life emotion annotation and machine learning based detection. Neural Networks 18(4), 407–422 (2005)
44. Rivera-Illingworth, F., Callaghan, V., Hagras, H.: Towards the detection of temporal behavioral patterns in intelligent environments. In: Proc. 2nd IET Int. Conf. Intelligent Environments, pp. 119–125 (2006)
45. Mori, T., Takada, A., Noguchi, H., Harada, T., Sato, T.: Behavior prediction based on daily-life record database in distributed sensing space. In: Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems, pp. 1703–1709 (2005)
46. Honda, S., Fukui, K.-I., Moriyama, K., Kurihara, S., Numao, M.: Extracting human behaviors with infrared sensor network. In: Proc. 4th Int. Conf. Networked Sensing Systems, pp. 122–125 (2007)
47. Patterson, D., Liao, L., Fox, D., Kautz, H.: Inferring high-level behavior from low-level sensors. In: Proc. 5th Int. Conf. Ubiquitous Computing (2003)
48. Oza, N.C.: Online bagging and boosting. In: Proc. IEEE Int. Conf. Systems, Man and Cybernetics, pp. 2340–2345 (2005)
49. Blum, A.: On-line algorithms in machine learning. In: Fiat, A. (ed.) Dagstuhl Seminar 1996. LNCS, vol. 1442, pp. 306–325. Springer, Heidelberg (1998)
50. Baram, Y., El-Yaniv, R., Luz, K.: Online choice of active learning algorithms. Jrnl. Machine Learning Research 5, 255–291 (2004)
51. Dasgupta, S., Kalai, A.T., Monteleoni, C.: Analysis of perceptron-based active learning. In: Proc. 18th Annual Conf. Learning Theory, pp. 249–263 (2005)
52. Li, Y., Guan, C.: A semi-supervised SVM learning algorithm for joint feature extraction and classification in brain computer interfaces. In: Proc. 28th IEEE EMBS Annual Int. Conf. (2006)
53. Zheng, H., Olaf, H.: Discrete regularization for perceptual image segmentation via semi-supervised learning and optimal control. In: Proc. IEEE Int. Conf. Multimedia and Expo., pp. 1982–1985 (2007)
54. Last, M., Kandel, A., Maimon, O., Eberbach, E.: Anytime algorithm for feature selection. In: Ziarko, W.P., Yao, Y. (eds.) RSCTC 2000. LNCS (LNAI), vol. 2005, pp. 532–539. Springer, Heidelberg (2001)
55. Auer, P., Cesa-Bianchi, N., Gentile, C.: Adaptive and self-confident on-line learning algorithms. Jrnl. Computer and System Sciences 64(1), 48–75 (2002)
56. Grabner, M., Grabner, H., Bischof, H.: Real-time tracking with on-line feature selection. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition (2006)

57. Chapelle, O., Scholkopf, B., Zien, A.: Semi-Supervised Learning. MIT Press, Cambridge (2006)
58. Zhu, X.: Semi-supervised learning literature survey. Computer Sciences TR 1530, Univ. of Wisconsin, Madison (2007)
59. Freund, Y., Seung, H.S., Shamir, E., Tishby, N.: Selective sampling using the query by committee algorithm. Machine Learning 28(2-3), 133–168 (1997)
60. Angluin, D.: Queries revisited. Theoretical Computer Science 313(2), 175–194 (2004)
61. Muslea, I., Minton, S., Knoblock, C.A.: Active + semi-supervised learning = Robust multiview learning. In: Proc. 19th Int. Conf. Machine Learning, pp. 435–442 (2002)
62. Zhu, X., Lafferty, J., Ghahramani, Z.: Combining active learning and semi-supervised learning using Gaussian fields and harmonic functions. In: Proc. 20th Int. Conf. Machine Learning, Workshop on the Continuum from Labeled to Unlabeled Data (2003)
63. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. MIT Press, Cambridge (1998)
64. Sharma, M., Holmes, M., Santamaria, J., Irani, A., Isbell, C., Ram, A.: Transfer learning in real-time strategy games using hybrid CBR/RL. In: Proc. 20th Int. Joint Conf. AI, pp. 1041–1046 (2007)
65. Abbeel, P., Quigley, M., Ng, A.Y.: Using inaccurate models in reinforcement learning. In: Proc. 23rd Int. Conf. Machine Learning, pp. 1–8 (2006)

Toward Daydreaming Machines

S.I. Ahson[1] and A. Buller[2]

[1] Patna University, Patna, Bihar, India
[2] Cama-Soft, Gdynia, Poland
a.buller@cama-soft.com

Abstract. This paper provides some insights related to building a working computational model of human-level mind. We propose to take a fresh look at some ideas propounded more than a century ago by William James and Sigmund Freud, which were recently reconsidered by Peter Naur and ATR Brain-Building Group, respectively. Naur proposes his Synapse-State Theory of Human Mind (SST), while the research at ATR resulted in the Machine Psychodynamic paradigm (MΨD). We argue that SST and MΨD propose complementary ideas about implementation of mental functionalities, including those related to the quest of consciousness. The 20th-century AI gave machine the ability to learn. The great challenge in the 21th-century AI is to make a robot actually want to learn. MΨD proposes a solution based on pleasure defined as a measurable quantity to be used as a general reinforcement. SST proposes a neuroscience-inspired architecture, where the key blocks are item-nodes, attention-node, and specious-present excitation. MΨD may supplement SST with a pleasure node and related Pleasure Principle.

1 Introduction

How the mind really works? Neither mainstream cognitive modeling nor mainstream robotics brings us closer to understanding the mystery. Even the ACT-R the flag computational model for cognitive psychology [3] deals with processing of sensory information toward *rational* behaviors out of their emotional context. Yet so many human actions is just irrational! As for speaking mascots or human-shaped "reception-desk staff" that can welcome guest and even answer their questions, their "intelligence" is nothing but a masquerade aimed to impress laymen. Although capturing the fundamental nature of generalized perception, intelligence, and action is still recognized as the challenge for AI [31], the ingenious insights of fathers of psychology, like William James or Sigmund Freud, are not being considered by the researchers who do not want to risk their careers [1]. Fortunately, recently one can hear some strong voices against this ill situation.

Peter Naur – recipient of 2005 Turing Award – writes: "William James's *Principles of Psychology* from 1890 is a supreme scientific contribution of mankind, on par with Newton's *Principia*" [28]. Eric Kandel – 2000 Nobel Prize winner – writes: "Psychoanalysis still represents the most coherent and intellectually satisfying view

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 245–256.
springerlink.com

of the mind" [23]. Inspired by the legacy of William James and Charles S. Sherrington (neurophysiologist, 1932 Nobel Prize), Peter Naur proposes the *Synapse-State Theory of Mental Life* (SST) [27]. The legacy of Sigmund Freud stimulated the ARS-PA (Artificial Recognition System-PsychoAnalysis [29], as well as the emergence of *Machine Psychodynamics* (MΨD) [10,11,13]. We argue that SST and MΨD do not contradict each other at any point and that their development may converge toward an artifact that can actually think human-like way.

What does it mean "to think human-like way?" Daniel C. Dennett [16] wrote: "There are robots that can move around and manipulate things adeptly as spiders. Could a more complicated robot feel pain, and worry about its future, the way a person can? (…). We know that the people the world over have much the same likes and dislikes, hopes and fears. We know that they enjoy recollecting favorite events in their lives. We know that they all have rich episodes of waking fantasy, in which they re-arrange and revise the details deliberately."

One cannot disagree with the above. Therefore, as before building a machine with human-like intelligence it is necessary to study the spontaneous flow of thought that occurs in daydreaming, for this phenomenon is not a by-product of "more useful" mental mechanisms [1].

2 Synaptic-State Theory (SST)

According to SST, the nervous system consists of three kinds of things: *synapses*, *neurons*, and *nodes*. Excitations are transmitted along the neurons, through the synapses, and are summed and distributed in the nodes. Naur's view of the system's structure is based on five layers: (1) *item-layer*, (2) *attention-layer*, (3) *specious-present-layer*, (4) *sense-layer*, and (5) *motor-layer*, as well as (6) a *special connection* from motor layer to the sense-layer.

Item-layer contains nodes that are the neurobiological embodiments of acquaintance objects. When two nodes to which a given synapse is connected are both excited, the synapse becomes more conductive. This means that the nodes become associated. If no re-excitation takes place, the synapse's conductivity will decay over years.

Attention-layer contains a set of synapses linking selected nodes with a single special node called attention-excitation. When a given item node N_i gets excited (through its connection to certain sense-layer nodes or other item nodes) beyond a certain level, the related attention synapse becomes conductive for about one second and in this period it conducts a strong excitation from the attention node. It can be said that at this moment the attention sits in the node N_i.

Specious-present-layer has similar structure as the attention layer. Its synapses normally have no conductivity. When a given item node gets excited above a certain level (especially by an impulse from an attention-excitation), a related specious-present synapse becomes conductive and provides the node with an excitation from the specious-present excitation. This extra excitation will fall off within about 20 seconds. It is assumed that always there are 5-10 specious-present synapses at the stage of falling off in their excitation. They form a queue of nodes that have been strongly excited within the latest minute or so.

The notion of specious present, taken from William James's *Principles of Psychology* [21] is the key to awareness of time-flow. Our attention jumps back and forth within a thought object, which happens within a time scale of about one second. Accordingly, whatever we think about at a certain moment contains in a weaker form what we thought about a few seconds before, in a yet weaker form what we thought yet earlier, and so on.

Sense-layer has a number of nodes. Each of them gets excited from one of the available senses (through appropriate transducers) and sends impulses to all item nodes. In the young individual the related synapses have only low conductivity. When they are excited from both sides simultaneously they gain in a conductivity which is higher in the direction towards the item-node, and lower in the opposite direction. When a particular item-node gets excited strongly enough to excite its attention-synapse, this is called perception of the corresponding object.

Motor-layer has a number of nodes each of which excite a dedicated muscle or a gland. Motor synapses connect each of the nodes with all item-nodes and are subject to training.

The *special connection* lets muscular activations influence sense cells. Owing to this we can feel the state of our body. According James's suggestion that has been caught in SST, all what we call *feelings* and *emotions* is a matter of the muscles having sense cells in them.

SST sorts out James's view of the mind and brings it quite close to a computational model, however, it can be noted that in his lecture Naur speaks nothing about mechanisms responsible for such phenomena as ambivalence, adventurousness, heroism, hunger of knowledge, or proneness to imitate. Yet this does not mean that SST rejects the mechanisms. On the contrary, this theory contains a lot of blank space for a new contribution related to the mentioned phenomena. We argue that the school of thinking that can contribute to SST a lot is Machine Psychodynamics.

3 Machine Psychodynamics (MΨD)

Machine Psychodynamics is an out-of-the-main-stream approach to building brains for robots, where the key concepts are *tension*, *pleasure*, and *ambivalence*. A tension relates to the degree to which a part of a robot's body deviates from its state of resting or to which a drive (such as fear, anxiety, excitation, boredom, or expected pain) deviates from its homeostatic equilibrium. The pleasure is a function of tension dynamics. When the tension abruptly drops, the pleasure volume rises and then slowly decays. A psychodynamic robot is defined as a creature that always seeks an opportunity to enhance its pleasure record [10].

Unlike a mainstream bio-mimetic robot that is expected to always try to keep each of its drives within a homeostatic regime [6], a psychodynamic robot may sometimes deliberately let a bodily or psychic tension increase - even to extreme values. Though this may be madness, there is method in it. The higher the tension is before plummeting, the stronger the enhancement of the pleasure record [9].

The term ambivalence refers to a perpetual struggle of ideas in the robot's working memory. So, unlike a conventional robot that in the case of contradictory ideas is expected to quickly decide which of them to implement, a psychodynamic robot may change its mind in unpredictable moments [12]. These properties are believed to enhance a robotic brain's potential to self-develop.

3.1 Pleasure Principle

For a psychodynamic robot, any tension-increase is an occasion to get a new portion of pleasure. However, in order to succeed, the robot must have learned what to do to cause a possibly rapid discharge of the tension. For each kind of tension there is a sensory pattern that triggers the process of discharge. Pleasure signals strengthen the circuits that contributed the most to the pleasure-resulting behavior. This principle forces the robot, all by itself, to develop smarter and smarter methods of changing its relation to the environment or changing the environment itself, just in order to enjoy perceiving various tension-discharging patterns.

As an example of pleasure-principle application, let us consider an insect-like robot equipped with a pair of tentacles. Assume it has no innate mechanism for obstacle avoidance. Upon each encounter with an obstacle, the robot performs random movements, while its pleasure generators receive signals representing the degree to which each of the tentacles is deformed. When one and only one tentacle touches something and, by accident, stops touching, a related pleasure generator produces a signal which reinforces the circuits that most strongly contributed to the recent movement. In this way, the creature learns, unsupervised, to more and more smoothly avoid obstacles, driven only by pleasure defined in psychodynamic terms. In the same way, a psychodynamic creature equipped with a camera may learn to chase objects of interest. It was also experimentally confirmed that, based on pleasure principle, a robot equipped with a microphone and speaker, is able to develop together with its caregiver a mutually understandable proto-language [10, 25].

3.2 Freud in Machine

Let us consider an SST-based robot equipped with a set of item-nodes constituting a world model including the model of the robot itself. Having one day completed machinery for handling the model, we may make a breakthrough in the issue that today is maybe the hottest one in the field of robotic intelligence – learning from observation and instruction. Related projects have resulted in giving robots the ability to imitate selected human behaviors [4] or learn from verbal communication [32]. MΨD intends to supplement these achievements with mechanisms that will make a robot actually want to learn.

Let us imagine a robot whose memory hosts two world models – a *model of perceived reality* and a *model of desired reality*. The desired reality may develop driven by, among other things, several sorts of challenges. Let us consider the following story: The robot notices a person (or other robot) juggling balls. A question emerges: "Would I be able to do the difficult thing that the other individual can do?" The question induces a challenge that results in the mental image of oneself juggling too. The difference between the desired reality and the perceived reality may be a source of a strong tension. How to discharge the tension and have the resulting pleasure? Just by learning. Needless to say, the learning can be recognized as a *voluntary* learning. But what to do with such tension when the learning cannot succeed? The theory of machine psychodynamics considers Freudian defense mechanisms [18], i.e., the possibility of redirecting the desire toward another challenge and acting toward a substitute satisfaction. Another defense may consist in a distortion of perceived reality, e.g. a repression of inconvenient memories to "unconscious" zones of the mind, which may reduce the difference between perceived and desired reality. To have the collection of world models complete, let us also mention a *model of ideal reality* to be acquired during upbringing. Having such a model the mind may develop tensions related to moral dilemmas that may result from the difference between the ideal reality and desired reality [8].

3.3 Machine Adventurousness

As Marvin Minsky proposes, one 'secret of creativity' may be to develop the knack of accepting the unpleasantness that comes from awkward or painful performances [26]. Indeed, only an adventurous individual may deliberately select a challenging gain over an easy one. Psychodynamic robots are adventurous owing to the pleasure principle and the adventurousness pays. First, adventurousness may facilitate survival. Second, it may accelerate cognitive growth.

As for survival, let us consider an ecosystem psychodynamic filter (ΨD-Filter) in which a population of robots whose habitat is separated from the rest of the environment by an unsafe zone. Let us assume that the supply of vital resources started decaying in this habitat. Let us also assume that the robot's knowledge includes neither the dimensions of the unsafe zone nor what lies beyond the zone. Needless to say, in this situation, if the robot were not psychodynamic, its fate would be sealed. Fortunately, unlike conventional robots that have no mechanisms facilitating an emergence of the idea to engage in a "purposeless" risk, a psychodynamic robot from time to time ventures into the unsafe zone and deliberately exposes itself to dangers - just to increase the fear-related tension and to get pleasure from discharging it. If the robot does not carry things too far or is simply lucky, it has a good chance of discovering an area rich in vital resources on the other side of the unsafe zone (Fig. 1). An instance of ΨD-Filter was tested experimentally using a simulation model [11].

The pleasure principle may make a psychodynamic robot penetrate "purposelessly" various areas of its own memories. Unlike a non-psychodynamic robot that confines the usage of its long-term memories to finding only data that are helpful to solve a problem that is already being faced, its psychodynamic cousin may daydream in the literal meaning. Daydreaming allows it to experience, to a certain extent, an increase of bodily and psychic tensions, as well as pleasures resulting from the discharging of the tensions. In order to magnify such substitute pleasures, the robot may embellish facts, design new adventures, or even imagine completely fantastic worlds. It can later try to implement the ideas it has dreamed out. Hence, This way a circuitry that facilitates creativity may emerge just as a result of the robot's strive for more pleasure.

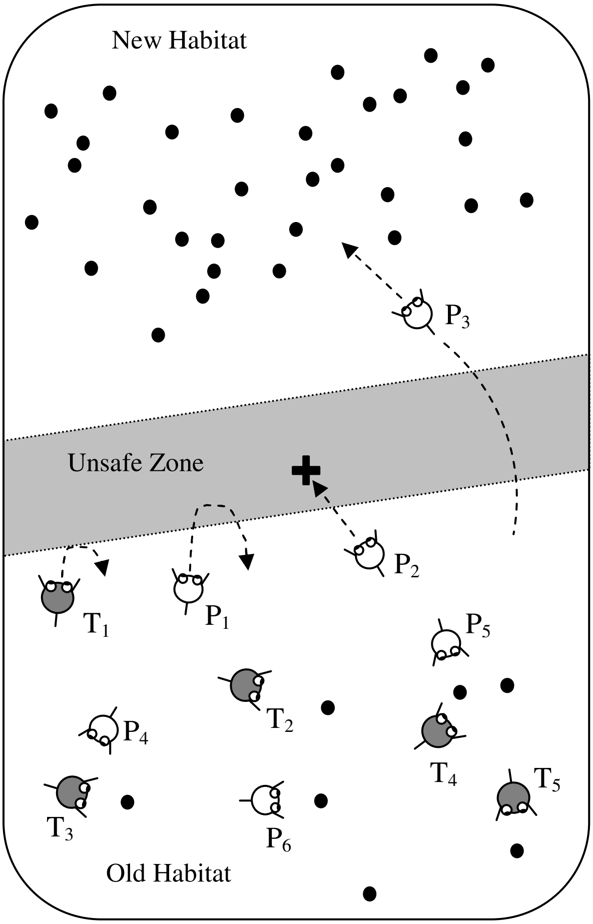


Fig. 1. ΨD-Filter. Black dots denote vital resources whose number decays. Grey robots always try to reduce tensions (hunger, fear). White robots seek for pleasure that comes from rapid reduction of a tension. Hence, grey robots avoid entering the unsafe zone (see T_1), while white ones sometimes venture into it to increase the fear-related tension and to get pleasure from discharging it (see P_1). Although a number of the adventurers may get killed in the unsafe zone (see P_2), the bravest ones may, by chance, reach a better habitat (see P_3). Equipped with a circuitry for daydreaming, a psychodynamic robot may experience tension-increasing/reducing adventures in its imagination.

3.4 Constructive Ambivalence

Is the distant object a snake, or only a snake-shaped branch? To accept the challenge, or to give up? To select a longer, but safer route, or a shorter but riskier one? To imitate the other individual's behavior, or rather refrain form the imitation? Unlike a conventional robot, which in face of such dilemmas tries to quickly work out an explicit decision (and thus employs an algorithm of multi-criteria optimization or a voting mechanism), a psychodynamic robot may endure ambivalence. In a psychodynamic mind contradictory ideas coexist and each of them tries to achieve domination in the working memory, for such a domination gives an access to motor drives and, in general, an influence on the course of things. However, as for the competing ideas, fortune is fickle. A winning idea may, after a while, lose to a rival one, and after an unpredictable time win again. An intrinsic dynamics of the process may cause irregular switches of beliefs or some inconsistencies in the robot's behavior [11].

Ambivalence may force a robot to develop new methods of judgment and to test their efficiency versus those developed earlier. Also, ambivalence gives a chance to sometimes implement a stupid idea, which, as long as the resulting behavior is not too devastating, may give the robot useful knowledge about its own physical or mental capacity.

4 Quest of Machine Consciousness

Machine Consciousness is a difficult subject. The notion is commonly related to intentionality and free will. Maybe most of people seem to reject the idea that robots can be conscious. This is based on the belief that "humans are the most significant beings in the universe" [17]. Some people do not like the idea of "machine ever being so much like us" [30]. In early 1900s ascribing mental processes to animals was thought to be against the principle of parsimony. Similar arguments are being used against robots attaining human-like qualities by doubters and fearers [30].

D. Griffin in [19] proposed a science of mental processes in the animals and gave it a name "cognitive ethology". His proposal, though controversial, has been advocated by William Farthing [17] because it is consistent with the "liberal attitude toward consciousness". A similar attitude can be extended to robots; however, the number of views and opinions related to the subject gives a little chance for development of a commonly accepted theory of consciousness before the end of 21st century. The most visible difference of approaches concerns the requirements the participants of the debate use to recognize an entity as conscious one. It can be noted that very hard requirements, as well as surprisingly modest requirements are being proposed.

4.1 Hard-Requirement Camp

Philip Johnson-Laird suggested [22] a hierarchically organized parallel-processing model of the mind in which the conscious mind interacts with the lower level non-conscious sub-systems. The levels are defined in terms of the availability of contents to reflective consciousness. Primary consciousness is the direct experience of sensory percepts and emotional feelings, spontaneously arising memories, thoughts, images, dreams and daydreams. Reflective consciousness concerns thought about one's own conscious experience and makes it possible for us to judge our own knowledge and to interpret our feelings. It helps to revise and improve our thoughts, to evaluate our actions. It helps to plan our future actions. Reflective consciousness is necessary for an elaborated self-awareness involves the realization that you are unique individual, separate from others, with your own personal history and your own future. In reflective consciousness you relate your current experience to your self concept, which is your concept of your personal nature, including your personal desires, values, goals, interest, intellectual and physical abilities, and personality traits. Reflective consciousness includes the process of introspection. Both primary and reflective consciousness is varying in content and quality. Primary consciousness evolves in children as they mature. Reflective consciousness develops out of primary consciousness.

Yang and Bringsjord [33] wrote: "Cognitive modelers need to step outside the notion that mere computation will suffice. They must face up to the fact, first, that the human mind encompasses not just the ordinary, humble computation that Newell and all his followers can't see beyond, but also *hyper*computation: information processing at a level *above* Turing machines, a level that can be formalized with help from chaotic neural nets, trial-and-error machines, Zeus machines, and the like."

Even if they were right, yet there are so many mental phenomena that remain to be covered below the Turing limit that there is no need to even consider an abandoning of SST or MΨD in favor of hyper-computation. Moreover, even if hyper-computing someday riches the stage of practical implementation, it will probably be possible to add a hyper-computing layer to a psychodynamic below-Turing-limit architecture [11]. The structure suggested by Johnson-Laird seems not to contradict the Naur's SST; however, an implementation of it remains an utopia. There are no tips how to represent in technical terms such concepts as "to judge our own knowledge" or "to interpret our feelings." Fortunately, there is the modest-requirement camp, which gives the enthusiasts of "strong AI" some hope.

4.2 Modest-Requirement Camp

Braitenberg in [5] proposed that free will can be attributed to *Vehicle 12* – a robot equipped with a module that, based on the number of brain elements activated at a given moment calculates the number of brain elements to be activated in the next moment. The related function was a U-curve, inverted and somehow distorted, so the generated numbers were virtually unpredictable for both a human observer and the robot itself.

Brooks in [7] attributed an intentionality to *Genghis* – a legged insect-like robot whose brain is a collection of interconnected AFSMs (augmented finite-state machines), where no AFSM has an intelligence higher than a soda machine. When Genghis's array of sensors caught sight of nothing, it waited. When perceiving a moving infrared source, the robot treated it as prey and chased it scrambling over anything in its path. Brooks argued that it was the robot's own will, since there was no place inside the control systems of Genghis to represent any intent to follow something.

Regardless of whether one agrees with the thesis that Vehicle 12 and Genghis are intentional creatures, we can at least admit that Braitenberg and Brooks proposed two criteria based on which we may consider a potential for intentionality.

The *Braitenberg criterion* is the lack of causality in a creature's behavior, while the *Brooks criterion* is the lack of a particular goal-related place in a creature's control system. In order not to provoke justified objections from philosophers, the notion of *proto-intentionality* was introduced [11] as the name of a feature one may attribute to a robot based on the Braitenberg criterion, Brooks criterion, and other criteria of this kind.

Creatures working in the framework of Machine psychodynamics (MΨD) paradigm meet the above criteria by definition. Moreover, MΨD provides another criterion based on the ΨD filter discussed in IIIC. MΨD intends for the pleasure signal not only to reinforce the learning of particular reactive behaviors but, aiming at the self-development of the human-level intelligence, intends to use pleasure as a motivator of the sophisticated planning and executing of plans. A fully psychodynamic robot must be able to deliberately expose itself to inconveniences and dangers – just to increase related tensions and then let them discharge, which might result in pleasure-signal generation. Hence, the *psychodynamic criterion* of intentionality is the ability to achieve a state defined as pleasurable by deliberately plunging oneself into a state defined as unpleasant.

According to Koch [24], one of the signs that a creature may be endowed with consciousness is a behavior revealing hesitation about what to do. Although the suggestion applies to living creatures, we could apply it also to artifacts, provided that hesitation demonstrated by an artifact is not a pre-programmed masquerade. This criterion, called the *Koch criterion* [11] fits well to psychodynamic machines since they demonstrate an emergent ambivalence, which seems to elevate their minds to a level that is not available to insects. Indeed, there is perhaps no doubt that dogs sometimes hesitate whether to attack or to escape, or whether to obey a calling or to ignore the caller.

Dawkins [15] proposed that consciousness may arise when the brain's simulation of the world becomes so complete that it must include the model of oneself. Psychodynamic architecture uses several kinds of imagined reality where, in each case, a robot itself is an actor (IIIB). Hence, MΨD-based fully developed artifacts will meet *Dawkins criterion* by definition.

Haikonen [20] proposes that the ultimate consciousness test should show that the machine has the flow of inner speech with grounded meanings and it is aware of it.. The existence of the flow can be determined from the architecture of the machine. As for the awareness of having the inner speech, we could rely on the machine's own report. As it can be noted, Haikonen [20] provides the idea of a higher-level daydreaming in which the tested scenarios are not only sequences of images, but also verbalized scripts that help the machine to make order in its long-term memories and more efficiently plan pleasure-oriented actions.

5 Concluding Remarks

How the mind really works? It often works irrationally. But, as we argue in this paper, a bit of irrationality may supports survival and cognitive development.

Hence the conclusion that the key to understanding how the mind works and building a truly thinking machine is probably hidden in the works of fathers of psychology forgotten by the mainstream cognitive robotics. The recent research by Peter Naur, by the ARS-PA project team in Vienna, and by the ATR Brain-Building Group in Kyoto aimed to rewrite William James' and Sigmund Freud's insights in technical terms to be eventually implemented using 21th-century technology. Machine psychodynamics (MΨD), as a new subfield on the borders of AI and cognitive sciences, emerged.

MΨD does not replace the homeostatic mechanisms employed by the mainstream AI robotics. The psychodynamic approach intends to supplement the mainstream solutions with mechanisms for pleasure-seeking and "deliberate irrationality". These mechanisms are expected to bring a breakthrough in the quest for a machine's self-development toward a human-level intelligence, especially when an apparatus for daydreaming and inner speech is implemented. This belief is justified by a number of experimental results, especially the experiment with emerging proto-language mentioned in IIIA. It is also believed that a set of methods for seeking pleasure that a robot develops may lean toward a state that is compatible with the caregiver's system of values. Hence, despite the pleasure-principle-based "selfishness", a psychodynamic robot may become useful to its human master.

Unfortunately, a robot designed to deliberately expose itself to inconveniences and dangers may be hardly welcomed by today's corporate investors. The same undoubtedly applies to a robot that displays visible signs of indecisiveness. Nonetheless, we argue that such troublesome properties may be an unavoidable price for a robot's cognitive self-development up to a level beyond that which can be achieved via handcrafting or simulated evolution.

References

1. Ahson, S.I., Buller, A.: Machine daydreaming: Self-rearrangement of long-term memories. In: Akerkar, R. (ed.) Artificial intelligence – future trends, New Delhi, pp. 179–192. Allied Publ. Pvt. Limited (2002)
2. Ahson, S.I., Buller, A.: Toward machines that can daydream. In: Proc. IEEE conference on human system interaction, Cracow, Poland, pp. 609–614 (2008)
3. Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., Qin, Y.: An integrated theory of the mind. Psychological Review 111(4), 1036–1060 (2004)
4. Bentivegna, D., Atkeson, C.G., Ude, A., Cheng, G.: Learning to act from observation and practice. International Journal of Humanoid Robots 1(4), 585–611 (2004)
5. Braitenberg, V.: Vehicles: experiments in synthetic psychology. MIT Press, Cambridge (1984/1986)
6. Breazeal, C.: Cognitive modeling for bio-mimetic robots. In: Bar-Cohen, Y., Breazeal, C. (eds.) Biologically inspired intelligent robotics, pp. 253–283. SPIE Press, Bellingham (2003)
7. Brooks, R.A.: Flesh and machines: how robots will change us, pp. 48–50. Pantheon Books, New York (2002)

8. Buller, A.: Psychodynamic robot. In: Proc. 2002 FIRA robot world congress, Seoul, pp. 26–30 (2002)
9. Buller, A.: From q-cell to artificial brain. Artificial Life and Robotics 8(1), 89–94 (2004)
10. Buller, A.: Building brains for robots: a psychodynamic approach. In: Pal, S.K., Bandyopadhyay, S., Biswas, S. (eds.) PReMI 2005. LNCS, vol. 3776, pp. 70–79. Springer, Heidelberg (2005)
11. Buller, A.: Machine psychodynamics: toward emergent thought. Technical Report TR-NIS-005, ATR Network Informatics Labs, Kyoto (2006)
12. Buller, A.: Mechanisms underlying ambivalence: a psychodynamic model. Estudios de Psicologia 27(1), 49–66 (2006)
13. Buller, A.: Four laws of machine psychodynamics. In: Dietrich, D., et al. (eds.) Simulating the mind. A technical neuropsychoanalytical approach, pp. 320–331. Springer, Heidelberg (2009)
14. Buller, A., Joachimczak, M., Liu, J., Shimohara, K.: ATR Artificial brain project - Progress report. Artificial Life and Robotics 9(4), 197–201 (2005)
15. Dawkins, R.: The selfish gene. Oxford University Press, Oxford (1976/1999)
16. Dennet, D.C.: Kind of minds: towards an understanding of consciousness (1996), `http://en.wikipedia.org/wiki/DanielDennett` (accessed April 4, 2009)
17. Farthing, G.W.: The psychology of consciousness. Prentice Hall, Englewood Cliffs (1992)
18. Freud, S.: An outline of psycho-analysis. W. W. Norton & Company, New York (1940/1989)
19. Griffin, D.: Animal thinking. Harward University Press, Cambridge (1984)
20. Haikonen, P.: Robot brains: circuits and systems for conscious machines. J. Wiley & Sons, Chichester (2007)
21. James, W.: The principles of psychology. Dover Publications, Inc., New York (1890/1950)
22. Johnson-Laird, P.N.: Mental models. Cambridge University Press, Cambridge (1983)
23. Kandel, E.R.: Biology and the future of psychoanalysis - a new intellectual framework for psychiatry revisited. In: Kandel, E.C. (ed.) Psychiatry, psychoanalysis and the new biology of mind, pp. 63–106. American Psychiatric Publishing, Inc., Washington (2005)
24. Koch, C.: The quest of consciousness: a neurobiological approach. Roberts & Co. Publishers, Englewood (2004)
25. Liu, J., Buller, A., Joachimczak, M.: Self-motivated learning agent – Skill-development in a growing network mediated by pleasure and tensions. Transactions of the Institute of Systems, Control and Information Engineers 19(5), 169–176 (2006)
26. Minsky, M.: The emotion machine. Simon & Schuster, New York (2006)
27. Naur, P.: A synaptic-state theory of human mind (2004), `http://www.naur.com/synapse-state.pdf` (accessed April 4, 2009)
28. Naur, P.: Computing versus human thinking. Communications of the ACM 50(1), 85–94 (2007)
29. Palensky, P., Lorenz, B., Clarici, A.: Cognitive and affective automation: Machines using the psychoanalytic model of human mind. In: Proc. 1st intern. engineering & neuro-psychoanalysis forum (ENF 2007), Vienna, pp. 49–73 (2007)

30. Sloman, A.: Architectural requirements for human-like agent both natural and artificial (What sort of machine can love). School of Computer Science, University of Birmingham (1999)
31. Stone, M., Hirsh, H.: Artificial intelligence: the next twenty five years. AI Magazine 26(4), 85–97 (2005)
32. Weng, J.: Developmental robotics: theory and experiments. International Journal of Humanoid Robots 1(2), 199–236 (2004)
33. Yang, Y., Bringsjord, S.: Newell's program, like Hilbert's, is dead; let's move on. Behavioral and Brain Sciences 26(5), 627 (2003)

Eliminate People's Expressive Preference in the Mood of Fuzzy Linguistics

Z. Qin and Q.C. Shang

School of Information Management & Engineering,
Shanghai University of Finance and Economics, Shanghai, China
sqc116@yahoo.com.cn

Abstract. Fuzzy linguistics is popular in the fields like human system interaction and intelligent systems today, so to eliminate the expressive preference in fuzzy mood has become a crucial issue for computer to understand human's real meaning well. After defining and discussing the concept of mood operator and two typical kinds of expressive preference in fuzzy mood – attitude preference and degree preference, we propose a method to eliminate this preference, which is to do a preparatory test in advance and get rid of the preference from data by mathematical translation. Both the process and mathematical translation method will be illustrated, and an experiment will be done to verify the effect of the method as well. The results of the experiment reflect that it's effective to some extent. At last, the method will be concluded as well as the study direction in future.

1 Introduction

Human language is plentiful and diversiform. For an object, people can use various expressions to describe it. So for the "stiff" computer, it's very difficult to identify and understand human language well and truly. Hereinto, the fuzziness in human language is one of the most important problems. In human's manner of expression, not all things can be described in a precise and quantitative way. In a lot of cases, people express things in a way between qualitative and quantitative expression, by which they indeed represent some difference in degree, but can't say the difference very accurately in a numerical way even by themselves. That's exactly an obvious feature of fuzzy linguistics in the ordinary course of events.

To describe the problems in real world, fuzzy linguistics is undoubtedly necessary. Actually, it provides a means by which people can handily represent their opinions in a natural and intuitive manner rather than a precise manner that they are not adept at ordinarily [3]. Therefore, to translate human's fuzzy natural language into a quantitative and relative precise form which can be understood well by the "stiff" and programmed computer is exactly what we need to do. In this field, one method for fuzzy linguistic measurement and translation is proposed by Zadeh.

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 257–268.
springerlink.com

He has given an approach to translate the fuzzy terms into numerical computations, which uses a fuzzy linguistic variable to represent a variety of values in natural language [4, 5]. Based on his idea of using fuzzy subsets for “computing with words” [6], a lot of representations have been proposed [8, 9]. Another of the most popular method adopted in practice is ranking, which gives an ordinal scale which indicates the fuzzy terms’ relative position but not their distance [10-12]. Without the magnitude of difference, the degree of fuzzy terms is still unclear, and the results of survey cannot be analyzed by traditional statistical methods [13-15]. Besides, there are also many other means to depict fuzzy linguistic variables, like triangular, trapezoidal and other suitable geometric representations [16, 17].

Each of the above methods has its respective pitfalls in some respects, such as inconsistency with human intuition and indiscrimination of interpretation [18]. But there is a common pitfall of all these methods, which is not taking people’s expressive preference in language into consideration. In fact, every person has its own expressive preference or habit, and different people have different preference in their linguistic expression. Due to these expressive preferences, sometimes what people say is not exactly what they mean or think. What we really would like to do is to catch what they mean rather than what they say, which means we need to find out their real meaning through their apparent language.

Nowadays fuzzy linguistics is so popular that to eliminate people’s expressive preference in fuzzy linguistics has become a basic and crucial issue in lots of fields, such as human system interaction, intelligent decision-making and knowledge-based systems [19, 20]. In this paper, we’d like to focus on the fuzzy mood terms and put forward a method to eliminate the expressive preference in the mood of fuzzy linguistics. After the section of introduction, the concept of mood operator and two typical kinds of expressive preference in the mood of fuzzy linguistics will be defined and discussed in depth with some actual examples. Then a general method to eliminate this preference will be proposed, and both its process and the mathematical translation method in it will be illustrated. To verify the effect of the method we proposed, we will also do an experiment by it, and the results of the experiment reflect it’s effective to some extent. Finally, the method will be concluded and its disadvantages will be given as study direction in future.

2 Expressive Preference in Fuzzy Mood

When people express in human’s natural language, there’re a lot of preferences and habits in their expressions, such as being used to using a genus of words, making sentences in a changeless grammatical form and representing things in a certain mood. These expressive preferences often make people’s linguistic expression deviate from their real meaning and form an individual style of their own which presents a certain rule. Whatever, the expressive preference has become an indispensable part of human language. And hereinto, the expressive preference in fuzzy mood is a very important aspect.

2.1 Mood Operator

In human's natural language, there is a species of words especially used to represent some degree of mood, such as "very", "comparative", "a little" and so on. In general, these words cannot be used solely as it will be meaningless, but only can be used with the words owning true meanings. By adding these words in front of the true-meaning words, the mood degree of their previous meaning will be changed along with them. So in fuzzy linguistics, we call this species of words as "mood operators", which are used to adjust the mood degree of fuzzy words [21].

Actually, mood operators are only useful for fuzzy words [21], so people's expressive preference in the mood of fuzzy linguistics can be exactly considered as the expressive preference in mood operators, which is eventually represented on the fuzzy membership. Here we mainly divide the expressive preference in fuzzy mood into two categories – attitude preference and degree preference.

2.2 Attitude Preference

As we all know, every person has their own attitude towards things. Some of them are positive and optimistic, and they usually pay more attention to the positive aspects of all things; while the others are negative and pessimistic, and they usually pay more attention to the negative aspects of all things. Similarly, this kind of attitude problem also exists in human's linguistic expression, which makes people be inclined to adopt different expressive preference. Some people are optimistic in linguistic expression, and they are usually inclined to magnify the positive representation and compress the negative representation; while some others are pessimistic in linguistic expression, and they are inclined to compress the positive representation and magnify the negative ones. Here we define this preference of attitude in human linguistics as "attitude preference", and consider it may influence the effect of human's linguistic expression.

Different from people's attitude towards things, the attitude preference here purely points at their habit or fancy in linguistic expression. In many cases, what people consider is one thing and what they represent is another, which usually proceeds from their different will of expression. People always would like others to know a certain part and certain degree of their opinions rather than all, so they would reveal something and hide some others for certain. This attitude preference results in the distance between people's expressions and opinions, which makes their linguistic expression don't always reflect what they really think and take on asymmetry for positive and negative things (Fig. 1).

In fuzzy linguistics, people's attitude preference is in action by mood operators to a great extent. People whose attitude preference is positive would generally choose strong mood operators to represent the positive meanings and weak ones to represent the negative meanings; to contraries, people whose attitude preference is negative would usually use weak mood operators to represent the positive meanings and strong ones to represent the negative meanings. E.g. for a film, the person with positive attitude preference may say "it's very interesting" while another whose attitude preference is negative may say "it's a bit interesting", though their feelings about the film may be similar. And for a traffic accident, the former may say “I’m a little unlucky” while the latter say “I’m very unlucky”, though their feelings about it is not very different. So as we can see, the attitude preference can really distort people’s meaning by mood operators in fuzzy linguistics.

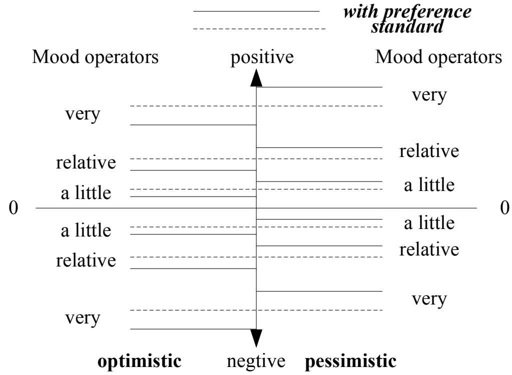


Fig. 1. Asymmetrical distribution of mood operators in attitude preference

2.3 Degree Preference

Besides the attitude preference, there is also another expressive preference in fuzzy linguistics, which may influence the effect of human’s linguistic expression as well. To describe an object, some people are exaggerated in linguistic expression and inclined to adopt an exaggerated expressive manner to represent it; contrarily, some others are conservative in linguistic expression and used to represent it in a cautious expressive manner. Owning to the difference in characters and culture, people may always adopt different expressive manners in linguistics which are consistent with their style of doing things. Here we define this exaggerated or conservative preference in linguistic expression as “degree preference”, and consider it may distort people’s real meaning in fuzzy linguistics as well.

The degree preference’s reaction on the mood of fuzzy linguistics is also actualized by mood operators. For all things, the exaggerated people generally would like to choose strong mood operators for description or expression while the conservative people would like to choose the weak mood operators. E.g. after degusting a delicious dish, while a conservative person just says “it’s delicious”, an exaggerated person may say “it’s extremely delicious”, though in fact it’s not so much more delicious than any other dishes. So this is just the habit or fancy of this person in linguistics, which may make there be a certain distance between his or her expressions and opinions. Different from the attitude preference, degree preference generally magnifies or compresses the expressive degree of things in the same proportion no matter they’re positive or negative, so it makes people’s linguistic expression towards positive and negative things symmetrical (Fig. 2).

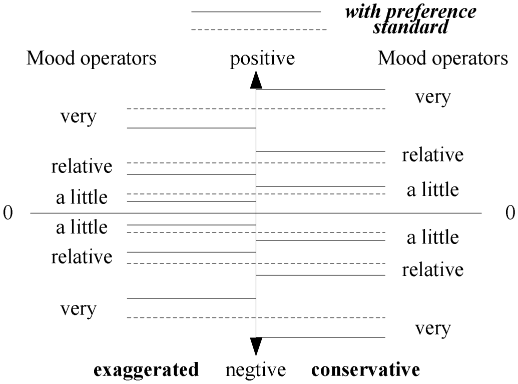


Fig. 2. Symmetrical distribution of mood operators in degree preference

3 Method

To make the computer understand human being's real meanings well through their apparent linguistic expression, it's necessary to eliminate the influence of expressive preference in fuzzy linguistics by dealing with the original data with some method, no matter it's the attitude preference or degree preference. And the method we proposed to eliminate people's expressive preference in the following is based on a simple idea, which is to do a preparatory test to get the respondents' expressive preference in advance, and then getting rid of their preference from the survey data by some mathematical translation. By this way, the processed data after translation should be the information without people's attitude preference or degree preference, and represent the respondents' real meanings in opinion.

3.1 Process

To eliminate the expressive preference in fuzzy mood, the steps of the whole process this method being actualized should be as follows:

1. For the theme that would like to be studied, make sure of the proper respondents being investigated.
2. Take the preconcerted respondents' background and the prospective answers of thematic investigation into account, design one or several questions as the prior questionnaire to test their expressive preference. These questions should be easy to answer and have nothing to do with both the theme and respondents' background, which will be answered only according to the respondents' opinions and intuitions. They should be able to represent the scale of mood operators in the thematic investigation. And what is most important is that there should be a standard index to measure the difference in respondents' fuzzy

mood of linguistics by them. So the answers should be able to reflect the expressive preference of the respondents.

3. Do the testing survey, and get their answers as the original data for analysis along with that of the thematic investigation. Actually, as having nothing to do with the questions of thematic investigation, the testing questions can be answered ahead of, along with or after the thematic questions though it's called as preparatory investigation, and it won't influence the testing survey's result. Sometimes doing the testing survey after thematic investigation is better, because the mood operators that the respondents use can be analyzed first and the testing questions can be designed suitably.
4. Analyze the original data being got in the testing survey. Calculate the distance of different grades in the respondents' answers by the measurement index. Set a series of normal distance, and compare the calculated distance with the normal ones. Adjust the fuzzy numerical value of mood operators in the thematic investigation according to a mathematical method and rule, when the departure of two distances exceeds the limit set in advance. And the value got after adjusting are the data we would like to get, which have been eliminated the influence of respondents' expressive preference.

3.2 Mathematical Translation

To eliminate the expressive preference in fuzzy mood well, the mathematical method used for translating the original data into processed information without expressive preference's influence is most important except the questionnaire's design. The mathematical translation method should be able to get rid of the influence of expressive preference in people's answer to the thematic investigation as much as possible without changing their real meanings. The method we proposed here is to adjust the fuzzy value of mood operators when someone's cognitive distance of mood operators are different much from the normal standards.

Suppose X to be a fuzzy concept or judgment, and A represents the mood word used for adjusting its degree, like "very". The membership of A belonging to X which means how much degree of "very X " belonging to X can be expressed as $\mu_X(A)$, where $\mu \in [0,1]$. So the fuzzy membership of mood operator can be expressed as:

$$\mu_X(A) = H_\lambda(A(X)) = H_\lambda(A)(X) \tag{1}$$

where λ takes different value for different mood operators and $\lambda \in R^+$. $H_\lambda(A)$ represents the fuzzy value's expression of mood operator A.

Before translating the original data into numerical values by a mathematical algorithm, set a benchmark A_0 for the different standard mood operators

$A_1, A_2, ..., A_n$ in a scale of mood operator set S first. Then measure the distance d_{i0} between all the mood operators used by respondents in answers and A_0 with the standard measurement index we set, which can be expressed as:

$$d_{i0} = D(A_i - A_0) \tag{2}$$

Measure that in the numerical algorithm we used as well, and compare the respondents' d_{i0} with the standard d'_{i0} according to the limit ε we set in advance. Adjust the value of λ according to some rule when the departure exceeds the limit, which can be expressed as:

$$\mu_X(A) = \begin{cases} H_{\lambda}(A)(X), | d_{i0} - d'_{i0} | \le \varepsilon \\ H_{\lambda'}(A)(X), | d_{i0} - d'_{i0} | > \varepsilon \end{cases} \tag{3}$$

where $\lambda' \in R^+, \varepsilon \in Z^+, n, i \in N$. The rule for adjusting is not stated, but it should make λ change in the same direction with the departure, and consider about the real distance between different mood operators in respondents' answers.

4 Experiments

To illustrate and verify the method proposed above, a simple experiment has been done. Based on not influencing the result of experiment, taking the limits and convenience of existing condition into account, we have designed a simple questionnaire according to the testing purpose, and then selected 30 students at random in the library of SUFE (Shanghai University of Finance and Economics) to answer the questionnaires. The process of the experiment is in the following.

1. Considering about the respondents' background in common, we chose "consumption capacity of market" and "demand for college graduates" as the themes of investigation, and thought their cognitions wouldn't differ too much. In the questionnaire, the respondents would be asked to select the adjectives as answers from the given sets (strong, weak) and (big, small), and add mood operators in front of the adjectives to adjust the degree they'd like to express as well.
2. Taking the independence and simpleness into account, we chose "the film you like" and "the dish you dislike" as the questions to test the respondents' expressive preference, and used the price that they would like to pay as the measurement index. Set the difference of degree as three grades, and the testing questions were: "how much would you like to pay for a film you like very much, comparatively, or a little?" and "how much would you like to pay for avoiding eating a dish you dislike very much, comparatively, or a little?"

3. Do the investigation by the selected students, and ask them to answer the testing questions along with the thematic ones in the questionnaire. Then withdraw all the questionnaires, and get their answers as the original data for analysis. For the 30 selected students, totally 27 questionnaires of this experiment were withdrawn, and hereinto 23 of them are valid.
4. Analyze the original data of the thematic investigation we got. Find out all the mood operators used in the answers, and build up a mood operator set $S =$(very, comparative, relative, a little, not too). For the reverse word "not too" in the set, considering about the meaning and degree it actually expresses, convert it into "a little" in the answers, and the adjectives into the corresponding antonyms at the same time. Take the mood operator "relative" as equal degree as "comparative", and thereby there are only three kinds of mood operators in the answers, each of which respectively represents a grade that is corresponding with the scale we set in the testing survey. So set up the scale with four grades totally including the case that there's no mood operator in front of the adjective at all (Table 1).

Table 1. Scale of mood operators

ID	Grade
1	a little, not too
2	comparative, relative
3	(none)
4	very

5. Analyze the original data of the testing survey we got. Get the median price of each grade for the first testing question, which is 50, 30, and 10. Set the grade of "comparative" as the benchmark for comparison, and calculate the relative distance among them by the expression as follows:

$$d'_{i0} = D(A_i - A_0) = \frac{|A_i - A_0|}{A_0} \quad (i \in N) \tag{4}$$

Deal with the data of the second testing question in the same way, and the results are shown in Table 2.

Table 2. Distance of mood operators

	M(A_1)	M(A_2)	M(A_4)	d'_{12}	d'_{42}
Question 1	10	30	50	0.67	0.67
Question 2	2	5	10	0.6	1

M = Median

6. Use a classical expressive form of fuzzy value as the mathematical algorithm to translate original data into numerical value, which is expressed as [21, 22]:

$$H_{\lambda}(A) = e^{\lambda} \quad (e \in (0,1)) \tag{5}$$

Take the calculated results of the two testing questions as the cognitive distance of different grades for positive and negative case respectively, according to which set the parameter in the expression above as $e = 0.7$, and the value of λ and H for every grades are shown in Table 3.

Table 3. Value of parameters in the expression

grade ID	1	2	3	4
Positive				
λ	5	1.5	1	0.1
H	0.17	0.59	0.7	0.96
Negative				
λ	5	2	1	0.1
H	0.17	0.49	0.7	0.96

7. Calculate the distances of different grades in every respondent's answers to the testing questions according to Formula (4), which are represented as d_{i0}. Take the relative distance of the mood operators' median d'_{i0} which is calculated in Step 5 as the normal departure, and compare d_{i0} with it. The calculation expression used for comparison is as follows:

$$l_{i0} = D(d_{i0} - d'_{i0}) = \frac{|d_{i0} - d'_{i0}|}{d'_{i0}} \quad (i \in N) \tag{6}$$

8. Take 50% as the limit, and adjust the numerical value of mood operators for the thematic investigation got in Step 6 according to the rules as follows: (Fig. 3)

$$H_{\lambda}(A) = e^{\frac{\lambda}{2}}, \quad \text{when } l_{i0} > 50\% \tag{7}$$

Finally, the processed data are got, in which the respondents' expressive preference in fuzzy mood has been eliminated.

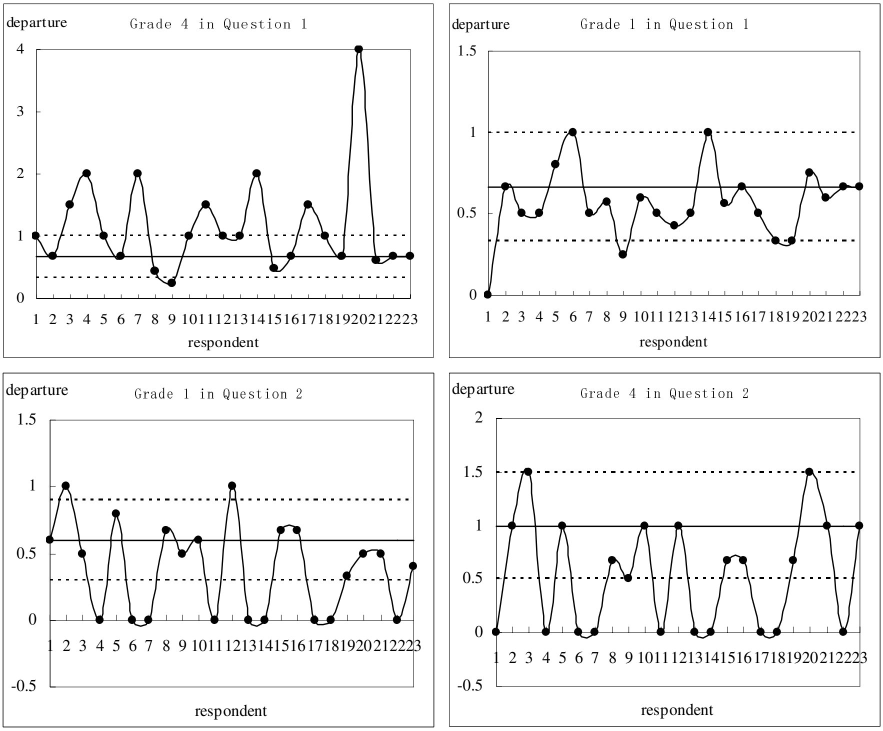


Fig. 3. Asymmetrical distribution of mood operators in attitude preference

5 Results

To test the effect of the method we proposed to eliminate people's expressive preference in fuzzy mood, we calculate the standard deviation of mood operators' numerical value for the two questions in the thematic questionnaire that both before and after the translation. The results are shown in Table 4.

Table 4. Standard deviation before & after elimination

	before	after
Question 1	0.20	0.14
Question 2	0.25	0.21

As we can see, the standard deviation of mood operators' numerical value has been reduced after the translation by the method we proposed. It can be considered as the expressive preference which may make people's expression deviate from their opinions has been eliminated, so the expression reflecting people's opinion become more coincident (Fig. 4).

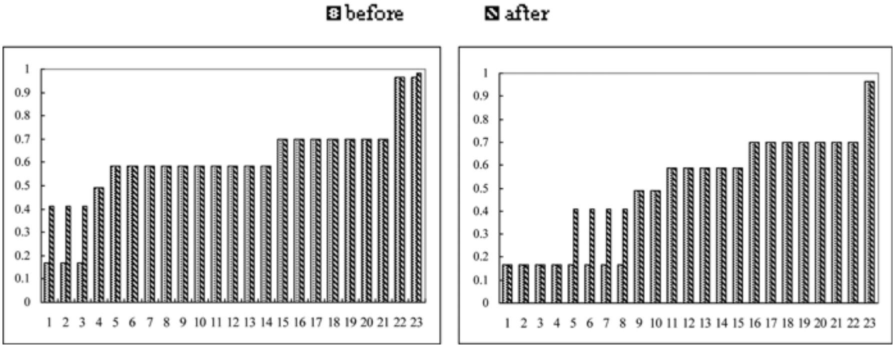


Fig. 4. Numerical value of mood operators before & after elimination in Question 1 and 2

6 Conclusions

As to understand fuzzy linguistics well has become a crucial issue in human system interaction, intelligent systems and so on, identifying and eliminating the expressive preference in human's fuzzy mood is one of the most important issue in this field. In this paper, we propose a method aiming at eliminating this preference. On the basis of defining the concept of mood operator and two typical kinds of expressive preference in the mood of fuzzy linguistics, we put forward a method of doing a test investigation to get the respondents' expressive preference in advance and getting rid of it from the survey data by mathematical translation. Both its process and mathematical translation are illustrated, and an experiment has been done to verify it, whose results illustrate its effect to some extent.

However, the greatest problem in this method is to design the testing questions and translation rules, which should eliminate the expressive preference well without change people's real meanings. It's really difficult and should be the study direction of this field in future.

References

1. Cassone, D., Ben-arieh, D.: Successive proportional additive numeration using fuzzy linguistic labels. Fuzzy Optimization and Decision Making 4(3), 155–174 (2005)
2. Zadeh, L.A.: The concept of linguistic variable and its application to approximate reasoning. Parts 1 and 2, Information Sciences 8(3), 199–249, 8(4) 301–357 (1975)
3. Zadeh, L.A.: Concept of a linguistic variable and its application to approximate reasoning. Part 3. Information Sciences 9(1), 43–80 (1976)
4. Yager, R.R.: Fuzzy logic in the formulation of decision functions from linguistic specifications. Kybernetes 25(4), 119–130 (1996)
5. Zeshui, X.: An interactive procedure for linguistic multiple attribute decision making with incomplete weight information. Fuzzy Optimization and Decision Making 6(1), 17–27 (2007)

6. Petrovic, S., Fayad, C., Petrovic, D., Burke, E., Kendall, G.: Fuzzy job shop scheduling with lot-sizing. Annals of Operations Research 159(1), 275–292 (2008)
7. Wang, X., Kere, E.E.: Reasonable properties for the ordering of fuzzy quantities. Fuzzy Sets and Systems 118(3), 375–405 (2001)
8. Liou, T.S., Chen, C.W.: Subjective appraisal of service quality using fuzzy linguistic assessment. Intern. J. Quality & Reliability Management 23(8), 928–943 (2006)
9. Arfi, B.: Linguistic fuzzy-logic game theory. J. Conflict Resolution 50(1), 28–57 (2006)
10. Malhotra, N.K.: Marketing research: An applied orientation, 3rd edn. Prentice-Hall, Upper Saddle River (1999)
11. Mason, R.D., Lind, D.A., Marchal, W.G.: Statistical techniques in business and economics, 10th edn. McGraw-Hill, New York (1999)
12. Eslami, E., Khosravi, H., Sadeghi, F.: Very and more or less in non-commutative fuzzy logic. Soft Computing 12(3), 275–279 (2008)
13. Kappoor, V., Tak, S.S.: Fuzzy application to the analytic hierarchy process for robot selection. Fuzzy Optimization and Decision Making 4(3), 209–234 (2005)
14. Chen, T.C.: Extensions of the TOPSIS for group decision-making under fuzzy environment. Fuzzy Sets and Systems 114(1), 1–9 (2000)
15. Zhang, H., Li, H., Tam, C.M.: Fuzzy discrete-event simulation for modeling uncertain activity duration. Engineering, Construction and Architectural Management 11(6), 426–437 (2004)
16. Lu, J., Zhang, G., Ruan, D.: Intelligent multi-criteria fuzzy group decision-making for situation assessments. Soft Computing 12(3), 289–299 (2008)
17. Mantas, C.J.: A generic fuzzy aggregation operator: rules extraction from and insertion into artificial neural networks. Soft Computing 12(5), 493–514 (2008)
18. Baoqing, H.: Fuzzy Theory Foundation, pp. 340–354. Wuhan University Press (2004)
19. Yang, W., Gao, Y.Y.: Principle and application of fuzzy mathematics, pp. 373–379. South China Polytechnic Press (2003)

VoiceXML Platform for Minority Languages

M. Brkić, M. Matetić, and B. Kovačić

Department of Informatics, University of Rijeka, Rijeka, Croatia
{mbrkic,maja.matetic,bkovacic}@inf.uniri.hr

Abstract. Not only that the use of different services is faster and easier in a voice-based user interface, but it also has benefits for people with physical impairments. This paper investigates the possibilities that minority languages have in developing spoken dialogue applications. We present our future platform and discuss difficulties that we as Croatian language speakers face in the planning and development phases. Furthermore, we set forth our Dialogue Manager Strategy, discuss properties which affect it and create weather-forecast dialogues. We have opted for VoiceXML as a dialogue modeling language since it is a language for creating voice user interfaces which simplifies application development, especially for minority languages.

1 Introduction

The number of telephone and mobile phone users increases daily and it is substantially greater than the number of Internet users. The number of telephone user services which enable DTMF (Dual Tone Multi Frequency) and voice input grows accordingly. Furthermore, beside visual browsers, there are also voice browsers which enable us to surf the Internet over the telephone from any place at any time [6]. Almost immediate information access is granted to people around the world. Research in this field aims at achieving natural language communication with the system. It is quite plausible that such a goal cannot be achieved in the foreseeable future. Therefore, it is better to talk about human-like dialogues with the system.

VXML (Voice Extensible Markup Language) is a programming language for creating VUIs (Voice User Interface) similar to XML (Extensible Markup Language) or HTML (Hyper-text Markup Language) [13]. It enables web programmers to create applications that can be accessed via any telephone or mobile phone. The development of voice applications is simplified because it is based on familiar web techniques, tools and infrastructure.

Time efficiency of voice interaction is indisputable. When it comes to telephone user services, voice interaction reduces costs and waiting time. Another advantage of the utmost importance is that these applications have benefits for people with physical impairments [7]. These systems are applicable in different spheres like support desks, airline and train arrival and departure information, booking services, weather and traffic conditions, etc [19]. However, it needs to be pointed out that voice interaction is transient and more susceptible to errors [19].

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 269–280.
springerlink.com

Spoken dialogue systems usually have modular design as shown in Fig. 1. Modularity simplifies integration of new features [10]. The system is connected to the user through ASR (Automatic Speech Recognition) and TTS (Text-to-Speech) modules [19]. ASR translates user utterances into text. In order to reduce complexity and improve quality, ASR usually has a defined set of acceptable words. ASR is connected to the linguistic analysis module which serves for interpreting syntactic expressions [3]. TTS plays the answer to the user either by using prerecorded audio files or synthesized speech [13]. The answer is generated by the answer generation module which receives appropriate information from the Dialogue Manager [3]. The Dialogue Manager is therefore a central component which extracts information from a database connected to the Internet and directly or indirectly communicates with the remaining modules in order to ensure system functionality [7].

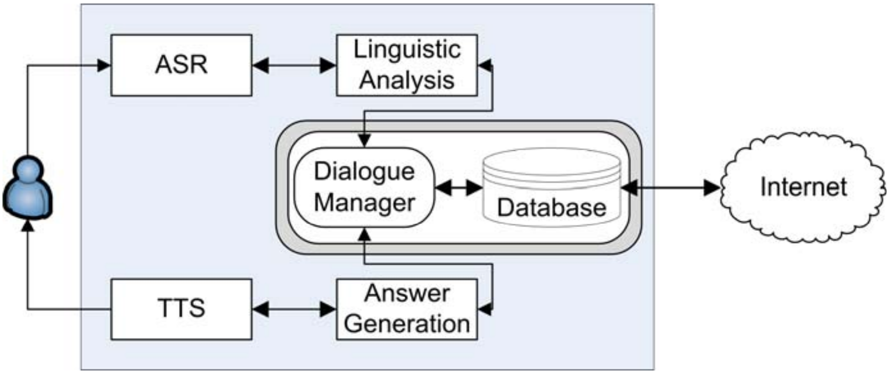


Fig. 1. Spoken dialogue system architecture

We have opted for VXML as a dialogue modeling language because of its simplicity [1]. In this paper we introduce VXML as our language of choice. Next, we give a detailed account of our future platform. Then we discuss properties that influence the Dialogue Manager Strategy. Based on the Dialogue Manager Strategy we set forth, we create simple VXML weather forecast dialogues, illustrating turn taking behavior and confirmation policy. Special attention is paid to the difficulties that minority languages like Croatian face in developing spoken dialogue applications and solutions to these problems are proposed. Lastly, we describe our CROVREP system and set directions for future work.

2 W3C Speech Interface Framework

W3C (World Wide Web Consortium) Speech Interface Framework is a collection of interrelated languages for developing speech applications. It consists of VXML, SRGS (Speech Recognition Grammar Specification), SSML (Speech Synthesis Markup Language), PLS (Pronunciation Lexicon Specification), SISR (Speech Recognition for Grammar Specification), CCXML (Call Control), and SCXML (State Chart XML) [19]. These languages can be used independently.

The goal of W3C is to lead the Web to its full potential and that is possible only by using compatible hardware and software. The responsibility for developing languages within the W3C Speech Interface Framework lies with the W3C Voice Browser Working Group. From this point on, we will turn our attention to VXML [19].

2.1 VXML

Typical components of a spoken dialogue system are listed in the introductory part. The Dialogue Manager is a central component and it can be written in VXML. The Dialogue Manager requires input from the user, interprets the input and determines the dialogue flow based on the script written in VXML. VXML can be used in DTMF applications as well [18].

The Dialogue Manager can be seen as a set of components. It consists of Input Interface, Grammar's Handling Unit, Output Interface, Document Manager, Logging Interface, VXML Interpreter, XML Parser, and ECMAScript Unit [11]. The Input Interface unit catches input-related events like *noinput* or *nomatch* events. It also sets and resets timers. The input is forwarded to the Grammar's Handling Unit, which implements Grammar Activation Algorithm and therefore activates and deactivates different grammars. It is in charge of grammar scope and also grammar generation if *option* or *menu* elements are used. More than one grammar can be active simultaneously, which makes multiple slot filling feasible [11]. However, as pointed out in [12], vocal grammars are inevitably limited. Recording user utterances and sending them to an external NLU (Natural Language Understanding Unit) might bypass this limitation. The Output Interface Unit is responsible for selecting a prompt and inserting appropriate variable values. The underlying algorithm is known as Prompt Selection Algorithm. The Document Manager is in charge of file downloads. The Logging Interface Unit makes error logs, warning logs, diagnostics logs and detailed interaction logs [11].

W3C VXML 2.1 recommendation was published in 2004 [17]. It relies on the principle of modularity and asynchrony which eases the integration of external media and applications.

2.2 VXML Platform Architecture

VXML is downloaded from HTTP servers in the same way as HTML. They differ only in layout mode. Since HTML is a visual language, the emphasis is on visual layout, while VXML puts emphasis on voice layout [19].

First, we will describe a typical cycle in our system. The overall architecture is shown in Fig. 2. A user dials a number and actually calls an Internet service provider that has a VXML Interpreter. When the connection is established, the user listens to the welcoming message and then makes a voice request. The request is received and sent to a particular web server. The web server returns a VXML page created after processing the request. The VXML Interpreter interprets the VXML page to provide the user with the desired information [5]. The web server is connected to a database which is automatically updated and from which a response to the user's query is generated.

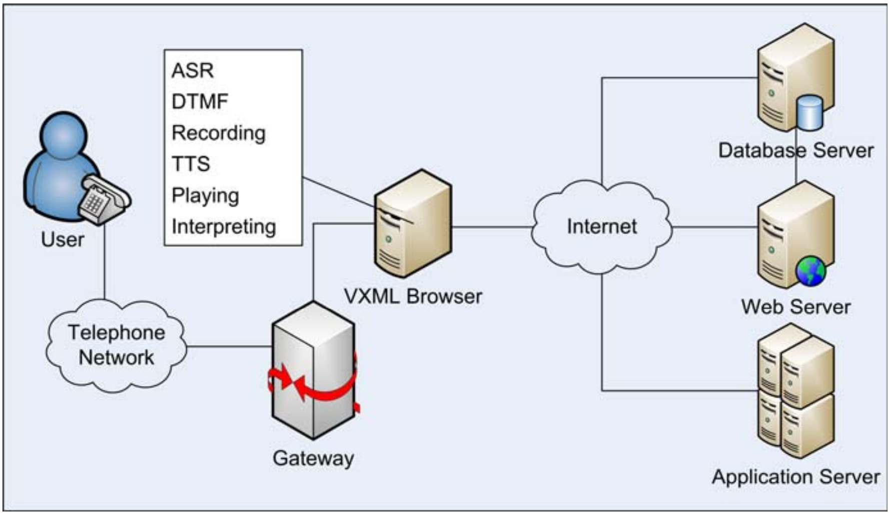


Fig. 2. VXML platform architecture

Voice browser resides on the voice server. It fetches VXML documents from the web server and sends these documents to the Interpreter and actually integrates modules responsible for ASR, DTMF interpretation, input recording, TTS, and playing prerecorded audio [18].

Gateway is equipped with the telephone card and it connects one or multiple telephone lines with the Interpreter [12]. It enables two-way connection between the telephone network and the world of Internet protocols and acts as a client to the application server [5].

Although the voice browser and the gateway provide two different functionalities, they can often be found in the same casing.

VXML Interpreter is responsible for interpreting commands. It controls speech and telephony resources like ASR, TTS, play and record functions, and a telephone network interface [5]. It is, therefore, responsible for answering calls, sending voice prompts to the user, accepting DTMF and voice input, recognizing or recording the words, sending requests to a web site and finally receiving information from the Internet and giving it to the user. The Interpreter can access a remote database or run application in the same way as HTML browser [18].

Interpreting is not a sequential procedure. Its flow is determined by different VXML algorithms, some of which have already been described. We will continue with the description of the remaining algorithms. Document Loop determines the flow of dialogue and keeps track of variables and document settings [11]. Form Interpretation Algorithm selects a dialogue element to be interpreted next and interprets it. It also catches input or events initialized after interpreting. Event Handling Algorithm catches and controls events that are thrown during interaction like *noinput* or *disconnect*. Resource Fetching Algorithm controls content fetching from different URIs (Unifies Resource Identifiers) [9].

There are a couple of open source Interpreters and a great number of commercial Interpreters. Open source VXML Interpreters do not usually have all of the VXML elements implemented. On the other side, commercial VXML Interpreters are often part of a complex IVR (Interactive Voice Response) system and therefore unsuitable for adaptation. This is of crucial importance when dealing with minority languages like Croatian. There is also a third possibility and that is building an Interpreter from scratch. Since there are no Interpreters with Croatian language support, second choice has to be excluded from further considerations.

3 Dialogue Modeling

Application is a collection of interrelated VXML documents which share the same root document. Root document variables are global scope variables and root document grammars are active throughout the application [7]. VXML document or a set of documents that make up an application is a finite conversation automaton. That means that the user is always in one of the possible conversation states or dialogues [16]. A document, therefore, contains one or more dialogues [7]. Each dialogue determines the next dialogue the user will engage in. These transitions are specified by URIs. If there is no document specified, the default document is the one currently active. If there is no dialogue specified then the execution starts with the first dialogue in a document. A conversation comes to an end if a dialogue does not specify its descendant [16]. A session starts when the connection is established and lasts until the user, the document or the system signals the end [7].

Prior to setting the strategy, we will discuss properties which affect the Dialogue Manager. Main Dialogue Manager goals are just-in-time information delivery, user-friendly interface and user guidance. The Dialogue Manager Strategy gives shape to the dialogue flow, confirmation policy, data quantity per turn, and available help [15]. Turn-taking and confirmation policy are human-human conversation properties which affect the Dialogue Manager Strategy.

Turn-taking rule has a number of important applications for dialogue modeling. We distinguish between three turn-taking situations. In the first situation a speaker selects the next person to speak by addressing him or her. This is reflected in two-part structures called adjacency pairs. Representatives of these structures are question-answer structures, greeting followed by greeting, request followed by grant, proposal followed by acceptance or rejection, etc. The second situation would be when the speaker does not specifically select another speaker, and then anybody can take turn. In the third situation nobody wants to take their turn, so the first speaker can take another turn [4]. These three situations can be mapped onto two system behaviors. In system-initiative or single-initiative systems, the system completely controls the conversation with the user by a series of questions. Mixed-initiative systems allow for initiative shifts between the user and the system [4]. The latter make the dialogue sound more natural, hence they are preferred.

Understanding is facilitated by confirmation, which can be explicit or implicit. Explicit confirmation includes additional questions to confirm system understanding at each point in a conversation. It lengthens the conversation and sounds awkward. Implicit confirmation sounds more natural and that is why it should be employed whenever possible. An alternative to a confirmation is a rejection [4].

3.1 Weather Forecast Dialogues

A detailed account of the Dialogue Manager Strategy in our weather forecast system is to be presented next. *What is the weather like in Rijeka today? Is it snowing in Zagreb? Is it windy in Senj?* These are only some of the questions that the user might ask. As an answer, our system would give a complete weather forecast or just one part of it.

The confidence obtained as a result of a recognized grammar in a *field* element would determine whether the system should ask for explicit confirmation (*Would you like to hear the weather forecast for Rijeka for today?*). While there are empty fields left, the system would continue questioning the user and filling the remaining fields. Each question would implicitly confirm previously filled slot (*You would like to hear the weather forecast for Rijeka. And for what period of time, please?*). In cases of input misrecognition, the system would follow the strategy of rapid re-prompting *Could you repeat that, please*? When the input is rejected for the second time, the system would follow the strategy of progressive prompting or escalating detail. The third rejection would lead to an explicit question *What is the name of the city?* Furthermore, the system would automatically switch to the entry at the telephone keypad in cases of multiple input misrecognitions. Moreover, DTMF support would be useful in cases of noisy environment or in situations where it is not proper to speak.

A form is one of the two existent dialogues in VXML. It is interactive section of a document and it has a set of item variables which are initialized during interaction. A form-level grammar, if present, can be used to fill several fields from one utterance. Another existent dialogue is a menu but we will not discuss it in this paper. Although we have opted for a form-filling mechanism, VXML also defines a mechanism for handling events not covered by the form mechanism. In case the user does not respond, needs help or the input is not intelligible, the platform throws an event, which is then caught by catch elements or their syntactic shorthand. Each input item may have associated shadow variables like the above mentioned *confidence* variable which belongs to the *field* element. They return additional execution results [16].

Interesting weather forecast voice browsing based on a VXML platform is presented in [6].

3.2 VXML Syntax Basics

VXML documents can be created with any text editor. They consist of plain text and tags. Tags are keywords or expressions enclosed by angle brackets (<and>). A tag can have attributes inside the angle brackets which consist of a name and a value (name=”value”). Tags which have counterparts, like <vxml> and </vxml>, are called containers, and contain either content or other tags. Tags that do not have counterparts are called empty tags or stand-alone tags, and they do not have anything besides attributes. An empty tag has a slash (/) just before the closing> character. Writing VXML documents demands strict syntax usage [8].

3.2.1 Basic VXML Elements

A simple VXML document code is presented below. This simple document generates a voice request for the user to choose a city, accepts the response from the user and sends it to a server-side script named *izbor.jsp*. The first line is known as a header line and it is followed by the body. The body contains at least one *form* element [16].

```
<vxml version="2.1">
  <property name="universals" value="all"/>
  <form>
    <field name="grad">
      <prompt> Odaberite grad za koji želite čuti vremensku prognozu – Osijek, Zagreb, Rijeka,
       Split ili Dubrovnik. </prompt>
      <grammar type="text/gsl"> [Osijek Zagreb Rijeka Split Dubrovnik] </grammar>
      <noinput> Molimo vas ponovite unos. <reprompt/> </noinput>
      <nomatch> Nismo vas razumjeli. Molimo vas ponovite unos. <reprompt/> </nomatch>
      <help> Odaberite grad <reprompt/> </help>
      <filled>
        <prompt> Vremenska prognoza za grad <value expr="grad"/>. </prompt>
      </filled>
    </field>
    <field name="sea" cond="grad=='Rijeka' || grad=='Split' || grad=='Dubrovnik'">
      <prompt> More ili kopno? </prompt>
      <grammar type="text/gsl"> [more kopno] </grammar>
      <filled>
        <prompt> Vremenska prognoza za grad <value expr="grad"/> <value expr="sea"/>.
        </prompt>
      </filled>
    </field>
    <block>
      <submit next="izbor.jsp"/>
    </block>
  </form>
</vxml>
```

The form used in the example above is a rigidly controlled form also known as directed form. It processes fields in a sequence. It contains only one field for storing the response.

A field can have a *type* attribute that can facilitate input processing. The *type* attribute of our interest is the *date*. It defines a field which accepts a calendar date. Its value is a string with eight digits – year, month and day, respectively. If the field does not have a *type* attribute, it must have a grammar. The element *grammar* defines elements that will be accepted by the Interpreter. The documents in this paper use the GSL grammar format used by the popular Nuance advanced speech recognition system. If the user's response can be found in the grammar, the Interpreter sets the field's *izbor* variable to the response. A *dtmf* tag can be used in the same manner as the *grammar* tag. It is worth to mention two additional tags not included in the example, *subdialogue* and *record* tags, which can be nested in the *form* tag. The first transfers control to another document, which can, in turn, return the control using the *return* tag, and the latter audio records the input [8].

The *prompt* tag tells VXML Interpreter to use TTS to read aloud the text enclosed in these brackets. If the *audio* tag is enclosed in prompt, the recorded audio files are to be played to the user. This tag may be a container or a stand-alone tag. If it is a container, a text can be enclosed in it. The text is pronounced only if the associated audio file cannot be found or loaded. Otherwise, the text is ignored. The *src* attribute specifies the URL of the audio file. A *prompt* tag can also have the *bargein* attribute which can be set to true or false to enable or disable barge-in. The *block* tag contains a section of executable content, which does not interact with the user. In the example given above, a *block* container has a *submit* tag which transfers control to another document or script and sends the value of the *izbor* variable to that script. However, it can serve the same purpose as the *prompt* tag [8].

Any text found between the delimiters <!—and --> is considered to be a comment and is, therefore, ignored by the Interpreter. A standalone tag *clear* is responsible for resetting the fields and resuming the execution at the top of the form. Universal commands are available anywhere in the document if there is a line of code including that property. The most interesting universal command is *help*, which has its counterpart and usually gives a detailed explanation of the application. In our example *help* is used to give a specific field related message if the user asks for help. The tag *filled* is used for defining specific actions once the field is filled in. The tag *value* is used to include the field value in the prompt by referencing the field name [8].

The *var* tag is used for declaring variables. Variables can hold numbers, text and several other data types. Every field has an associated variable based on its *name* attribute. The *var* tag is used to hold values computed by the script which do not have their own fields. Attributes that belong to the form items and that control their execution are *cond* and *expr* attributes. The Interpreter executes a form item if its guard condition is set to true. A guard condition can be any JavaScript expression [8].

3.2.2 Event Handlers

If the user fails to provide input, the *noinput* tag specifies what the document should do. The amount of time that the Interpreter waits for input can be adjusted by the *timeout* property. Properties are VXML predefined variables. There is another useful tag and that is the *reprompt* tag, which makes the Interpreter repeat the prompt from the current form field. There is also *nomatch* tag, which has functionality similar to those of the *noinput* tag and is used when the input does not match any active grammar. These two tags can be defined at the top level of the document. However, such usage makes them too general. Tags *noinput*, *nomatch*, and *prompt* can have *count* attributes with various values. When the value reaches a certain number, the prompt with that *count* value is played [8].

3.2.3 The Flow of Dialogue and Mixed Initiative Dialogues

In the example below, the user can ask for three services in any order which illustrates mixed initiative dialogue behavior. The top-level rule in the grammar is a rule called *Request* and it represents the entire input. The first letter has to be capitalized so it could be identified as a name, rather that part of the spoken text. The presence of a question mark allows the user to use the word following the question mark in any reasonable place. *Service* is the name for another rule which can recognize the choices. Two choices are optional, as marked by a question mark. The portions enclosed in curly brackets are commands that store the value *true* into special variables called slots. The first element processed by the Interpreter in the mixed-initiative form is the *initial* element, which contains the opening prompt. It can also be used for initializing variables. Notice that the tag *filled* has *mode* attribute set to *any*. Otherwise, the form would execute only if all the fields have been filled in.

```
<vxml version="2.1">
 <form>
  <grammar type="text/gsl">
   <![CDATA[Request (?[ (zanima me) (htjela bih čuti) (htio bih čuti) ] Service ?and ?Service
    ?and ?Service ?molim vas) Service ([temperatura    { <temperature true> }  vjetar  { <wind
    true> }  oborine   { <precipitation true> }])]]>
  </grammar>
  <initial>
   <prompt> Što vas zanima </prompt>
  </initial>
  <field name="temperature"    type="boolean"> </field>
  <field name="wind" type="boolean"> </field>
  <field name="precipitation" type="boolean"> </field>
  <filled mode="any">
   <if cond="temperature"><prompt>Odabrali ste tempreraturu.</prompt> </if>
   <if cond="wind"><prompt>Odabrali ste vjetar.</prompt> </if>
   <if cond="precipitation"><prompt>Odabrali ste oborine.</prompt> </if>
  </filled>
 </form>
</vxml>
```

4 CROVREP System

Slavic languages are a group of Indo-European languages spoken in most of Eastern Europe, much of the Balkans, part of Central Europe, and the northern part of Asia. Slavic languages have similar vocabulary, as well as similar basic grammar. All of these languages are free word order because of the rich morphological system. However, there is an unmarked subject-verb-object order. They have three noun/adjective genders, gender and number agreement for nouns, adjectives and sometimes verbs, developed grammatical case system (from six to seven cases), number and person agreement for subjects and verbs, and lack of articles *the* and *a* [2]. All these grammatical subtleties complicate the process of modeling Slavic languages and developing their language tools.

The system we are building is to be based on VXML. Since existent voice browsers do not have Croatian language support, we intend to engage in the integration process. We will use already developed ASR and TTS engines based on Hidden Markov Models and integrate them with the OpenVXI Interpreter [6]. However, ASR engine needs to be modified in order to report recognition confidence values. Meteorological and Hydrological Service[1] is to be used as the data source. The web application for fetching data has already been built. It runs on a LAMP (Linux, Apache, MySQL, and PHP) platform and adds weather forecast reports to MySQL database three times a day at certain hours. The script for obtaining weather forecast reports is run by *cron*, a Linux daemon for executing scheduled commands.

5 Conclusions

There are several ways of accessing web pages via telephone devices. WTE (Microsoft Web Telephony Engine), WAP (Wireless Application Protocol), and state-of-the-art VXML are among the most important ones [13]. In the foreseeable future, we can expect hundreds of applications with input modalities that include not only speech and keyboard, but also pointing as emphasized in [14]. The W3C Multi Modal Interaction Working Group (MMIWG) has already started working on standards which would allow this [12].

Open standards have proven to be the key to accepting new technologies primarily because of their flexibility and portability. In general, we can say that VXML is a powerful, though very simple, programming language for spoken dialogue applications development. More importantly, it is platform independent. However, different service providers often add their own extensions which, along with unspecified grammar format, might hinder portability. Still, VXML connects telephony and World Wide Web in a unique fashion, and that is why we should continue our work in that direction.

When it comes to the quality of service, Ajax technology is worth mentioning [7]. If VXML documents are atoms of speech applications, time efficiency is affected and other services are temporarily unavailable. Therefore, if only small parts of a dialogue need to be updated, Ajax technology should be employed. Ajax supports dynamic web page updates, either synchronous or asynchronous. Since VXML uses ECMAScript which, on the other side, does not have support for Ajax, the authors in [7] propose using X+V (XHTML+Voice Profile) instead. X+V is a web language with support for visual and voice interaction.

Since Croatian is a language of only about 6 million speakers, conducting researches is not a straightforward task. Its complicated syntax and morphology

[1] http://meteo.hr

pose even more obstacles. Croatian is a language which is threatened with extinction, and that is why we need to strive to develop Croatian language tools, which would enable effective communication with foreigners and effective usage of web based systems.

References

1. Brkić, M., Matetić, M.: VoiceXML for slavic languages application development. In: Proc. IEEE Conference on Human System Interaction, Cracow, Poland, pp. 147–151 (2008)
2. Genzel, D.: Creating algorithms for parsers and taggers for resource-poor languages using a related resource-rich language. PhD thesis, Brown University, Department of Computer Science, Providence, Rhode Island (2005)
3. Ipšić, I., Matetić, M., Martinčić-Ipšić, S., Meštrović, A., Brkić, M.: Croatian speech technologies. In: Proc. ELMAR Conference, Zadar, pp. 143–146 (2007)
4. Jurafsky, D., Martin, J.H.: Speech and language processing: an introduction to natural language processing, computational linguistics and speech recognition. Prentice Hall, Engelwood Cliffs (2000)
5. Kukka, P.: Utilization of VoiceXML in speech applications. Archit. Solut. Commun. Netw. Converg. 43(A), 21–27 (2002)
6. Meng, H., Yuk-Chi, L., Tien-Ying, F., Kon-Fan, L., Ka-Fai, C., Tin-Hang, L., Man-Cheuk, H., Ching, P.C.: Bilingual chinese/english voice browsing based on a VoiceXML platform. In: Proc. IEEE Intern. Conference on Acoustics, Speech and Signal Processing, Quebec, vol. 3, pp. 769–772 (2004)
7. Nepper, P., Treu, G., Küpper, A.: Adding speech to location-based services. Wireless Pers. Commun. 44(3), 245–261 (2008)
8. Nuance Café. VoiceXML tutorial, `http://cafe.bevocal.com/docs/tutorial/index.html` (accessed November 18, 2008)
9. Ondáš, S.: VoiceXML interpreters. In: Proc. 5th Students' Conference FEI TU, Košice, pp. 97–98 (2005)
10. Ondáš, S.: VoiceXML-based spoken language interactive system. In: Proc. 6th PhD Students' Conference and Scientific and Technical Competition of Students of FEI TU, Košice, pp. 97–98 (2006)
11. Ondáš, S., Juhár, J.: Dialog manager based on the VoiceXML interpreter. In: Proc. 6th Intern. Conference DSP-MCOM, Košice, pp. 80–83 (2005)
12. Rouillard, J.: Web services and speech-based applications around VXML. J. Netw. 2(1), 27–35 (2007)
13. Trninić, D.: Primena VXML aplikacija u poštanskim i telekomunikacionim kompanijama. IX telekomunikacioni forum TELFOR, Beograd (2001) (in Croatian)
14. Tsai, M.Y.: VXML dialog system of the multimodal IP-telephony - the application for voice ordering service. Expert Sys. with App. 31(1), 684–696 (2006)
15. Villarejo, L., Hernando, J., Castell, N., Padrell, J., Abad, A.: Architecture and dialogue design for a voice operated information system. Appl. Intell. 24(3), 253–261 (2006)
16. W3C, Voice Extensible Markup Language (VoiceXML) 2.0 (2004), `http://www.w3.org/TR/voicexml20/` (accessed December 16, 2008)

17. W3C, Voice Extensible Markup Language (VoiceXML) 2.1 (2007), http://www.w3.org/TR/voicexml21/ (accessed December 16, 2008)
18. VoiceXML Guide, http://www.vxmlguide.com/ (accessed December 15, 2008)
19. W3C Voice browser activity, http://www.w3.org/Voice/ (accessed December 14, 2008)

Polish Speech Processing Expert System Incorporated into the EDA Tool

A. Pułka and P. Kłosowski

Institute of Electronics, Silesian University of Technology, Gliwice, Poland
apulka@polsl.pl

Abstract. The work describes an application of the speech processing expert system into the electronic system design flow. The main idea of the system is presented. Some aspects concerning the proposed speech processing methodology bound with characteristic properties of Polish language are emphasized. The idea of dialog system menus is discussed. The inference engine based on AI techniques supporting natural language processing is proposed. The entire expert system architecture is introduced. Implementation and examples are discussed.

1 Introduction

A typical user of every-day electronic equipment wants to have programmable devices with many functions and simple handling, i.e. as flexible as possible. It presents a great challenge for engineers – who are forced to deliver sign-off products in a very short amount of time. Modern engineers would like to be supplied with very powerful and intelligent design tools and they treat a computer (machine) not only as a tool but sometimes almost as a partner guessing his intentions. During the design process an engineer performs many repeatable time-consuming actions, which not necessarily require creative thinking. That is why many EDA vendors supply their tools with mechanisms facilitating the design process. This paper proposes the design expert system supplied with the speech recognition module, dialog module with speech synthesis elements and inference engine responsible for data processing and language interpreting. The approach is dedicated to system-level electronic design problems and is a proposal of a tool for modeling tasks. It focuses on automatic generation of modules based on speech and language processing and on data manipulating.

2 The Idea and the Expert System Architecture

Problems of natural language processing are present in many practical applications [5, 6] and belong to hot topics investigated in many academic centers [11, 13, 14, 15, 16]. Rapid development of technology, multimedia and internet has shortened distances between countries evidently, so the automatic language translations systems are strongly demanded. Our main objective is to create an expert system that

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 281–293.
springerlink.com

aids the design process and enriches its abilities with speech recognition and speech synthesis properties. The proposed solution is intended to be an optional tool incorporated into the more complex environment working in the background. We show that such a philosophy simplifies the process of translation of the design models between subsequent levels of abstraction and facilitate the design components generation, too. The goal is to create a system that assists the working designer and suggests modules shells and templates, automatically generates signals, etc. We think that our approach is worth considering its application in a real, practical EDA tool.

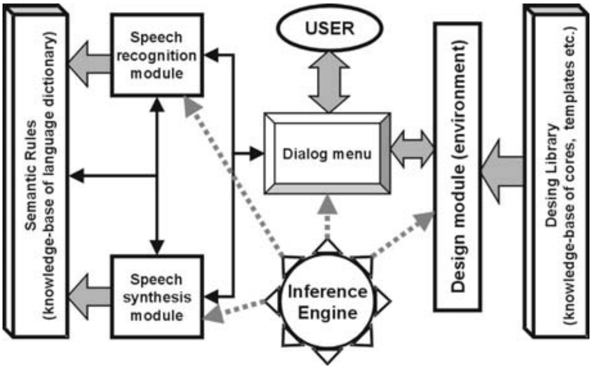

Semantic Rules
(knowledge-base of language dictionary)
Speech recognition module
Speech synthesis module
USER
Dialog menu
Inference Engine
Design module (environment)
Desing Library
(knowledge-base of cores, templates etc.)

Fig. 1. A block diagram of a design system based on menus

The above objectives can be met with the AI-based expert system that consists of the following components: speech recognition module, speech synthesis module, language processing module with knowledge base (dictionary and semantic rules), knowledge base of the design components and design rules and intelligent inference engine which ties together entire system, controls the data traffic and checks if the user demands are correctly interpreted. Fig. 1 presents the block diagram of the system, and the subsequent subsections address each part of it.

2.1 The Speech Recognition Module

Speech recognition is the process of finding the message information hidden inside the acoustic waveform [4]. The nature of speech recognition problem heavily depends on speaker, speaking conditions and message context. Usually, the speech recognition process is performed in two steps (Fig. 2). In the first step speech signal is processed by phonemes recognition system. As a result we obtain a sequence of phonemes or allophones. Phonemes are sound units which determine meaning of words. In phonetics, an allophone is one of several similar phones that belong to the same phoneme. A phone is a sound that has a defined wave, while a phoneme is a basic group of sounds that can distinguish words (i.e. change of one phoneme in a word can produce another word). This sequence of phonemes is processed by phonemes-to-text conversion unit with elements of speech understanding system.

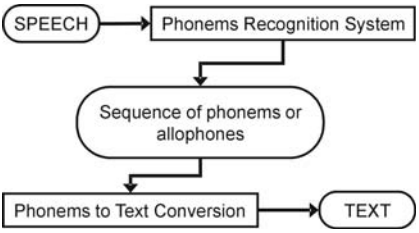

SPEECH
Phonems Recognition System
Sequence of phonems or allophones
Phonems to Text Conversion
TEXT

Fig. 2. Two steps of speech recognition process

2.2 Implementation of the Speech Recognition Module

Effective speech recognition process of Polish language is based on linguistic knowledge. The entire process of speech recognition can be improved if we make use of information stored in acoustic, phonetic, syntactic and semantic knowledge bases [6]. Implementation of the effective speech recognition system can be based on multilayer speech recognition system architecture [8]. Each layer moves the speech recognition process one step forward. There are: acoustic layer, articulation layer, phonetic layer, syntactic layer, semantic layer and application layer.

The first acoustic layer provides physical parameters of speech. Second articulation layer provides vectors of distinctive parameters of speech. Basing on these vectors phonetic layer generates sequence of speech phonemes. Syntactic layer using dictionary of pronunciation rules provide orthographical notation of speech. Semantic layer establishes the spelling meaning of the orthographical sequence of characters and provides sentences in Polish language. The task of application layer depends on destination of speech recognition system.

The presented method for speech recognition is based on detection of distinctive acoustic parameters of phonemes in Polish language. Distinctive parameters have been assumed as the most important selection of parameters, which have represented objects from recognized classes of phonemes. Improvement of phonemes recognition process is possible when using phonetics and phonology of Polish language [8]. Each phoneme is specified by vector of distinctive parameters of speech signal. First distinctive parameter is the class of the phoneme. Second corresponds to the place of the phoneme articulation. Third parameter denotes the method of the phoneme articulation.

Average number of distinctive parameters required for the recognition of one phoneme equals to 2.71, and can be estimated by the following formula:

$$Ns = \sum_{k=1}^{M} p_k \cdot N_k = \sum_{k=1}^{37} p_k \cdot N_k = 2.71 \qquad (1)$$

where: Ns is average number of distinctive parameters of speech, M number of phonemes, p_k - probability of k-th phoneme articulation, N_k – a number of distinctive parameters required for recognition of k-th phoneme. Table 1 presents set of distinctive parameters of Polish speech with articulation probability.

The most advanced speech recognition system is full speech dialog system with elements speech understanding based on AI techniques (Fig. 3).

Table 1. Distinctive parameters of Polish phonemes

k	Phoneme	Probability p_k	N_k	k	Phoneme	Probability p_k	N_k
1	b	0.013	3	20	ǯ	0.0005	3
2	p	0.027	3	21	č	0.010	3
3	d	0.019	3	22	ʒ´	0.002	3
4	t	0.038	3	23	c´	0.011	3
5	g´	0.001	3	24	I	0.034	3
6	k´	0.006	3	25	y	0.035	3
7	g	0.013	3	26	e	0.088	3
8	k	0.023	3	27	a	0.080	3
9	v	0.030	3	28	u	0.029	3
10	f	0.013	3	29	o	0.078	3
11	z	0.015	3	30	m	0.030	2
12	s	0.026	3	31	n	0.034	2
13	ž	0.010	3	32	ń	0.022	2
14	š	0.017	3	33	η	0.007	2
15	z´	0.002	3	34	r	0.007	1
16	s´	0.013	3	35	l	0.018	1
17	χ	0.009	2	36	j	0.039	2
18	ʒ	0.007	3	37	u̯	0.019	2
19	c	0.013	3				

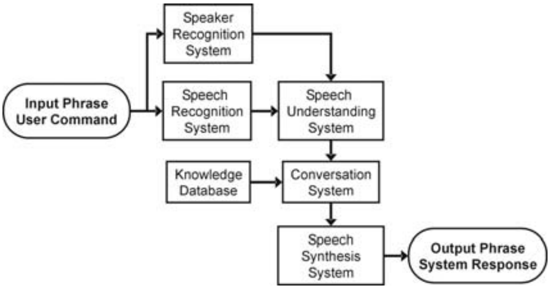

Speaker Recognition System
Input Phrase User Command
Speech Recognition System
Speech Understanding System
Knowledge Database
Conversation System
Speech Synthesis System
Output Phrase System Response

Fig. 3. The structure of speech dialog systems

2.3 Speech Synthesis Methodology

Today, the speech synthesis process is widely used in many practical applications, especially in telecommunication devices. The full TTS system converts an arbitrary ASCII text to speech. The first task of the system is extraction of phonetic components of the message performed by the text processing unit (Fig. 5).

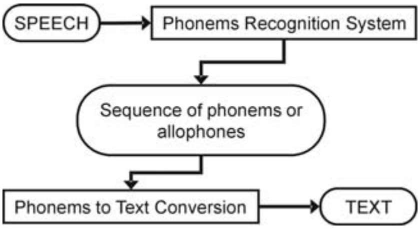

SPEECH
Phonems Recognition System
Sequence of phonems or allophones
Phonems to Text Conversion
TEXT

Fig. 4. The structure of TTS (Text-To-Speech) synthesis system

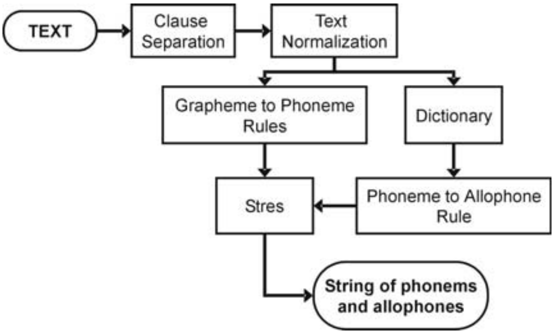

TEXT
Clause Separation
Text Normalization
Grapheme to Phoneme Rules
Dictionary
Stres
Phoneme to Allophone Rule
String of phonems and allophones

Fig. 5. The structure of TPU (Text Processing Unit)

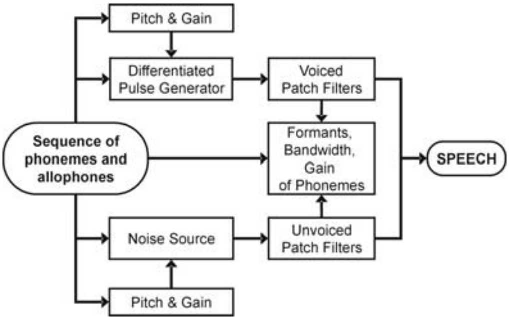

Pitch & Gain
Differentiated Pulse Generator
Voiced Patch Filters
Sequence of phonemes and allophones
Formants, Bandwidth, Gain of Phonemes
SPEECH
Noise Source
Unvoiced Patch Filters
Pitch & Gain

Fig. 6. The Speech Processing Unit structure

At the output of this stage we have a string of symbols representing sound-units (phonemes or allophones), boundaries between words, phrases and sentences along with a set of prosody markers (indicating the speed, the intonation etc.). The second part of the process consists of two steps – finding the match between the sequence of symbols and appropriate items stored in the phonetic inventory and binding them together to form the acoustic signal to be sent by the voice output device. This task is executed in speech processing unit shown in Fig. 6.

A combination of linguistic analysis is to be done during the first stage which involves: converting abbreviations and special symbols (decimal points, plus, minus, etc.) to a spoken form. Two generations of speech synthesizers for Polish language based on TTS technique have been developed [8] so far:

- SM10 text-to-speech system for Polish is the first speech synthesizer. It simulates the human vocal tract and is dedicated for blind persons. SM10 allows proper word pronunciation and word stress by means of full phoneme transcription. Speech synthesis has been made on the phoneme level.
- SM23, the next generation of speech synthesis system is based on allophonic level. The speech generated with this methodology is of better quality than speech obtained with phoneme-based technique in SM10. The new software provides natural-sounding and highly intelligible text-to-speech synthesis.

2.4 Examples of Polish Language Processing

We can distinguish two main tasks in language processing: a speech synthesis consisting of letter-to-phoneme and phoneme-to-sound conversions and a speech recognition divided into sound-to-phoneme and phoneme-to-letter conversions.

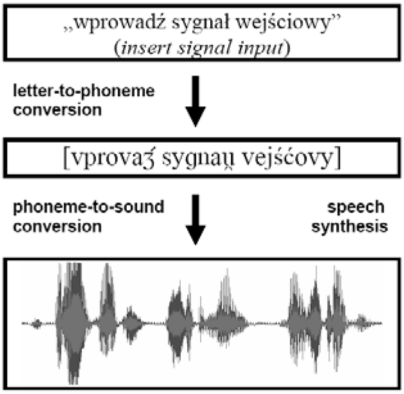

„wprowadź sygnał wejściowy”
(*insert signal input*)
letter-to-phoneme conversion
[vprovaʒ́ sygnau̯ vejśćovy]
phoneme-to-sound conversion
speech synthesis

Fig. 7. An example of letter-to-sound conversion in speech synthesis process

The letter-to-phoneme conversion changes ASCII text sequences to phoneme sequences. The phoneme-to-letter conversion performs reverse operations. It is based on implementation and employment of rule-based system and the dictionary for exceptions. This is very crucial fragment of the speech processing software.

Pronunciation of Polish language words is not very complicated. Even though the letter-to-phoneme conversion has more than 90 pronunciation rules, which requires an exception dictionary. Each phoneme is actually represented by a structure that contains a phonemic symbol and phonemic attributes that include duration, stress, and other proprietary tags that control phoneme synthesis. This scheme is used for handling allophonic variations of a phoneme. The term phoneme refers either to this structure or to the particular phone specified by the phonemic symbol in this structure. Fig. 7 and 8 present examples of this process.

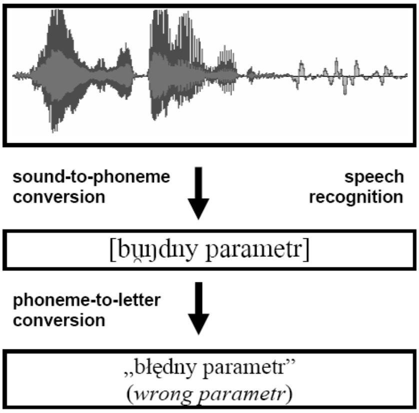


Fig. 8. Example of sound-to-letter conversion in speech recognition process

a)

```
listframe([model,2,decoder,`4BitGrayDecoder`]).
listframe([counter,`2BitCounter`,2]).
listframe([port,1,`Clock`,`in`,`std_logic`,1]).
listframe([port,1,`ParallelInputData`,`in`,`std_logic_vector(1 to 2)`,2]).
listframe([port,1,`CountEnable`,`in`,`std_logic`,1]).
listframe([portvariable,1,`CapacityVar`,``,`integer`,1]).
listframe([function,2,truthtable,`4BitGrayDecoder_tab`,
[`InputData`,`ChipEnable`,`OutputData`]]).
listframe([function,`4BitGrayDec_tab`,2,1]).
listframe([function,`4BitGrayDec_tab`,`('1' '-' '-' '-' '-' 'Z' 'Z' 'Z' 'Z' )`]).
(.......)
listframe([function,`4BitGrayDec_tab`,`('0' '1' '0' '0' '0' '1' '1' '1' '1' )`]).
listframe([delayedports,1,`InputDataA`,[`InputDataA_ipd`,]).
istframe([violationflag,1,`TviolRecovery_InputDataA_CarryIn`]).
listframe([aliasvariable,`CoutVar`]).
```

b)

```
gem([design_module,’my_design’,’MULT1’]).
gem([module_type,’MULT1’,multiplier]).
gem([design_module,’my_design’,’ADD1’]).
gem([module_type,’ADD1’,binary_adder]).
gem([design_module,’my_design’,’program3’]).
gem([module_type,’program3’,software]).
gem([design_module,’my_design’,’MUX1’]).
```

Fig. 9. Examples of PROLOG clauses: a) listframes representation of Vital hardware models [8] b) gems representation of system components

Allophonic rules are the last rules, which are applied to the phoneme stream during the letter-to-phoneme conversion process. In fact, they are the phonetic rules, because most allophonic rules are described as follows: "*if phoneme A is followed by phoneme B, then modify (or delete) phoneme A (or B).*"

2.5 System Knowledge Bases

We have divided the entire information stored within the system into two parts: the design knowledge-base and Polish semantic (dictionary) knowledge-base. The first part is a kind of design library [7], which consists of sets of virtual components, templates and cores with generic parameters, while the other is a typical dictionary with semantic rules [6, 9] (currently it is limited to the domain of electronic system design). The data stored in both parts is hierarchical – in the form of frame-lists and generic entity modules [7] (Fig. 9). This unification of both databases simplifies the management of the information within the system.

2.6 Semantic Rules of the System Menu Commands

The commands understood by the expert systems are grouped into the categories concerning the type of inserted information. We can distinguish global, editing and reviewing commands. Every command has its own permissible set of parameters; these parameters have a given form with other parameters or values, etc. This scheme creates a kind of semantic network within our system. Table 2 contains some examples of this structure.

Table 2. Examples of commands (in Polish)

TYPE	PARAMETERS
global	rozpocznij -> edycję -> danych
	start -> editing -> of data
	anuluj -> poprzedni -> moduł (port)
	cancel -> previous -> module (port)
editing	wprowadź -> sygnał -> wejściowy -> ..
	insert -> signal -> input -> ..
	popraw -> nazwę -> portu -> wejściowego -> Ain
	correct -> name -> of port -> input -> Ain
	dodaj -> sygnał -> wewnętrzny -> Tx
	add -> signal -> internal -> Tx
reviewing	pokaż -> ostatnio -> wprowadzony -> moduł
	show -> lately -> inserted -> module
	sprawdź -> połączenia -> między -> sygnałami -> wejściowymi
	check -> connections -> between -> signals -> input

2.7 Management of the Information in the System – Inference Engine

The dialog menu of the expert system works in the background – it resembles the even-driven behavior of a typical object-oriented environment. When needed (if the user calls a request) the speech interface starts and asks for the input information. The entire process can be repeated as many times as necessary. For control and management functions is responsible the special inference engine [7] that consists of two elements: PROLOG language [12] with backtracking search mechanism, inference engine borrowed from NVMG system [7] and based on a form of non-monotonic reasoning - default logic [1], but extended with elements of uncertainty and fuzzy information [10, 17]. PROLOG has proved its abilities and properties in the field of natural language processing [2, 3]. The non-monotonic logic models common-sense reasoning and increases the abilities of *'pure'* PROLOG.

The inference rules are based on Fuzzy Default Rules [17]:

$$\frac{\alpha : \beta_1, \beta_2 ... \beta_N}{\Phi^{\lambda}} \tag{2}$$

where: α, $\beta_1...\beta_N$ are *wffs* (well formed formulas) in a given propositional language $\boldsymbol{L}$ and Φ^{λ} is a Fuzzy Hypothesis(***FH***) [17] of the form:

$$\Phi^{\lambda} = \left\{ \left[h_1^{\lambda}, \mathrm{Tw}\left(h_1^{\lambda}\right) \right], \left[h_2^{\lambda}, \mathrm{Tw}\left(h_2^{\lambda}\right) \right], \ldots, \left[h_m^{\lambda}, \mathrm{Tw}\left(h_m^{\lambda}\right) \right], \right\} \tag{3}$$

where: $\boldsymbol{h_i^{\lambda}}$ ($\boldsymbol{i}$ = 1...$\boldsymbol{m}$) are wffs in propositional language $\boldsymbol{L}$, and $\boldsymbol{Tw(h_i^{\lambda})}$ denotes ***Trustworthiness***; i.e. one of the modality of generalized constraints in the Zadeh's sense [10] (bivalent, probabilistic, fuzzy, veristic etc.). For the simplest case the trustworthiness can be treated as a membership function or probability.

Additionally, we assume that prerequisite (like in Reiter [1]) represents strong information, while the possible uncertainty or missing of information is represented by justifications $\beta_1...\beta_N$.

```
infer():-
          fuzzy_default_rule(Alfa,[Beta1|BetaTail],[[H1,Th1]|HypothesesTail]),
          \+(negation(Alfa)), forall(member(X,[Beta1|BetaTail]),negation(X)),
          \+ temporal_hypothesis([[H1,Th1]|HypothesesTail]),
          assertz(temporal_hypothesis([[H1,Th1]|HypothesesTail])).
negation(Fact):-  Fact, !, fail.  /* modified negation*/
negation(Fact) :-  no(Fact),!.
negation(Fact) :-  \+ Fact, weak_neg(Fact).
weak_neg(Fact):-  weak_negation(Fact),!.
weak_neg(Fact):-  assertz(weak_negation(Fact)).
```

Fig. 10. PROLOG implementation of the inference rule

The interpretation of FDR is very similar to standard DL rule [1] ("*if prerequisite* α *is true and the negation of* $\beta_1,...\beta_N$ *(justifications) cannot be proved then infer hypothesis* Φ^{λ} *and treat it as a temporary conclusion (that could be rejected)*") except the form of the hypothesis (FH), which consists of different aspects (views, extensions) of the same problem and each of these sub-hypothesis has its own ***Tw*** coefficient reflecting the *significance* of the given solution. Elements of a given FH Φ^{λ}, i.e. are $\boldsymbol{h_1^{\lambda}}$, $\boldsymbol{h_2^{\lambda}}$,.., $\boldsymbol{h_m^{\lambda}}$ are mutually exclusive. The inference engine based on both parts navigates the process of speech (language) interpretation and because of it is based on non-classical deduction structure. Fig. 10 shows the Prolog implementation of the predicate infer and modified negation.

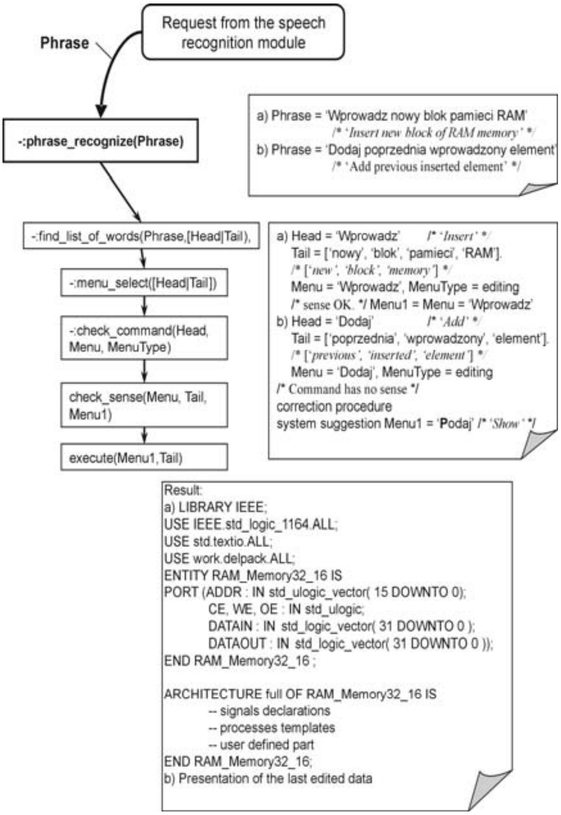
3 The Implementation and Examples

As mentioned above, the semantic rules of the recognition mechanism have been implemented together with the speech recognition system. The speech recognition system is based on Polish dictionary and contains many words, not necessarily dedicated to the design systems and it may happen that the recognized phrase is not correct. If the noise level or pronunciation of the user is not correct, the single words or the entire sentence could be misunderstood by the system. In consequence the command has no meaning for the design process. In such a case the inference engine based on FDL rules (2) starts its work trying to find the correct meaning of the command and returning the results to the design module. The semantic recognition of a given command is based on the database templates which constitute a kind of semantic patterns. The employment of such sophisticated inference engine enables making temporary hypotheses and their potential rejections in case of wrongly recognized sense. The latter process is known in AI techniques as revision of beliefs (the fragment of PROLOG implementation is given below).

```
phrase_recognize(Phrase):-
          find_list_of_words(Phrase,[Head|Tail), menu_select([Head|Tail]).
menu_select([Command|ListOfParameters]):-
          check_command(Command,Menu,MenuType),
          check_sense(Menu,ListOfParameters,Menu1),
          execute(Menu1,ListOfParameters).
/* If the command is not recognized system generates voice information to the user */
menu_select(_):-
          speech_generate(‘Polecenie nie rozpoznane’),
          speech_generate(‘Spróbuj jeszcze raz’).
          check_command(X,X,Type):- command(X,Type), !.
          check_command(Command,X,_):- infer1(X, Command).
```

Fig. 11. Some of the semantic checker predicated (PROLOG implementation)

The predicate menu_select is based on fuzzy default rule scheme and is extended by some additional operations not addressed in this paper. Predicates check_command, check_sense and check_parameter (not presented here) are responsible for command interpretation, and in case of fail, there are possible two different scenarios: generation of the returning information to a user about the error or start the dialog with the user. This interactive, dialog scheme is an idealistic solution not always possible. However, if we assume the finite number of possible actions at a given stage, it is possible to foresee context and construct a commonsense dialog scenario. Fig. 12 shows two examples of the expert system activity process based on speech commands. First diagram presents a correctly recognized command and subsequent steps from the recognition to the final description generation, while the other describes the situation, when the wrongly recognized command is corrected by the inference engine and the appropriate action is executed. The result of the first example is generation of HDL description, and the effect of the second example is presentation of the last entered module. The latter case is more interested because the system has recognized that the command has no meaning, i.e. it has no semantic value. The system has verified that the combination of the command ‘dodaj’ (*add*) with the rest of parameters has no sense, although such command exists within the system. The inference engine interprets it as a mistake and a given FDR finds command ‘podaj’ (*give*) that has similar phonetic characteristic, but completely different meaning which fits the parameters.


Phrase
Request from the speech recognition module
-:phrase_recognize(Phrase)
-:find_list_of_words(Phrase,[Head|Tail),
-:menu_select([Head|Tail])
-:check_command(Head, Menu, MenuType)
check_sense(Menu, Tail, Menu1)
execute(Menu1,Tail)

```
a) Phrase = 'Wprowadz nowy blok pamieci RAM'
                /* 'Insert new block of RAM memory' */
b) Phrase = 'Dodaj poprzednia wprowadzony element'
                /* 'Add previous inserted element' */
```

```
a) Head = 'Wprowadz'        /* 'Insert' */
   Tail = ['nowy', 'blok', 'pamieci', 'RAM'].
   /* ['new', 'block', 'memory'] */
   Menu = 'Wprowadz', MenuType = editing
   /* sense OK. */ Menu1 = Menu = 'Wprowadz'
b) Head = 'Dodaj'           /* 'Add' */
   Tail = ['poprzednia', 'wprowadzony', 'element'].
   /* ['previous', 'inserted', 'element'] */
   Menu = 'Dodaj', MenuType = editing
/* Command has no sense */
correction procedure
system suggestion Menu1 = 'Podaj' /* 'Show' */
```

```
Result:
a) LIBRARY IEEE;
USE IEEE.std_logic_1164.ALL;
USE std.textio.ALL;
USE work.delpack.ALL;
ENTITY RAM_Memory32_16 IS
PORT (ADDR : IN std_ulogic_vector( 15 DOWNTO 0);
        CE, WE, OE : IN std_ulogic;
        DATAIN : IN  std_logic_vector( 31 DOWNTO 0 );
        DATAOUT : IN  std_logic_vector( 31 DOWNTO 0 ));
END RAM_Memory32_16 ;

ARCHITECTURE full OF RAM_Memory32_16 IS
        -- signals declarations
        -- processes templates
        -- user defined part
END RAM_Memory32_16;
b) Presentation of the last edited data
```

Fig. 12. Examples of the program run

4 Conclusions and Summary

The paper presents the novel approach to the design process employing the speech dialog system. It combines the design of electronic systems issues with the speech processing and understanding techniques. The prototype PROLOG [12] implementation has been tested on the example set of the typical design procedures (commands) and Polish language dictionary [4]. The proposed expert system can be an interesting cover extending abilities of the existing EDA tools. Its implementation allows simplifying the design process. Thanks to the novel approach to Polish language recognition based on phonemes and allophones we have achieved very promising results. The speech recognition system works on huge language dictionary, while the design expert system is limited to only selected set of commands. Implementation of probabilistic techniques based on dictionary improves the effectiveness of the recognition process.

The described implementation gives new quality to the design process, allows the real conversation between the program and the user, helps handicapped persons to work as designers, and extends the quality of the work. The system is open, it can be extended with new rules (defaults) and commands and the thesaurus mechanisms that can give more flexibility to the designer.

References

1. Reiter, R.: A logic for default reasoning. Artificial Intelligence 13, 81–132 (1980)
2. Gazdar, G., Mellish, C.: Natural language processing in prolog. Addison Wesley, Wokingham (1989)
3. Lee, M.C., Leong, H.V.: NLUS – A prolog based natural language understanding system. In: Proc. IEEE Conference on Computing and Information, Toronto, Canada, pp. 204–207 (1992)
4. Ostaszewska, D., Tambor, J.: General information about phonetics and phonology of contemporary polish language. University of Silesia nr 488, Katowice, Poland (1993) (in Polish)
5. Manning, C.D., Schütze, H.: Foundations of statistical natural language processing. MIT Press, Cambridge (1999)
6. Jurafsky, D., Martin, J.H.: Speech and language processing: An introduction to natural language processing, speech recognition and computational linguistics. Prentice-Hall, Upper Saddle River (2000)
7. Pułka, A.: Modeling assistant - a flexible VCM generator in VHDL. In: Seepold, R., Martinez, N. (eds.) Virtual components design and reuse, pp. 171–182. Kluwer Academic Publishers, Boston (2000)
8. Kłosowski, P.: Improving of speech recognition process for polish language. Transactions on Automatic Control and Computer Science 47(61), 111–115 (2002)
9. Chou, W., Juang, B.H.: Pattern recognition in speech and language processing. CRC Press, Inc., FL (2002)
10. Zadeh, L.A.: Precisiated natural language (PNL). AI Magazine 25(3), 74–91 (2004)
11. Gu, L., Gao, Y., Liu, F.H., Picheny, M.: Concept-based speech-to-speech translation using maximum entropy models for statistical natural concept generation. IEEE Transactions on Audio, Speech and Language Processing 14(2), 377–392 (2006)
12. LPA Prolog, `http://www.lpa.co.uk` (accessed April 23, 2009)
13. Ammicht, E., Fosler-Lussier, E., Potamianos, A.: Information seeking spoken dialogue systems: part I and part II. IEEE Transactions on Multimedia 9(3), 532–566 (2007)
14. Infantino, I., Rizzo, R., Gaglio, S.: A framework for sign language sentence recognition by commonsense context. IEEE Transactions on Systems, Man, and Cybernetics, Part C 37(5), 1034–1039 (2007)

15. Neumeier, K., Thompson, C.: Parameterizing menu based natural language interfaces with location models. In: Proc. IEEE Conference on Integration of Knowledge Intensive Multi-Agent Systems, Waltham, MA, 236–240 (2007)
16. Wang, Y.: A systematical natural language model by abstract algebra. In: Proc. IEEE Conference on Control and Automation, Guangzhou, China, pp. 1273–1277 (2007)
17. Pułka, A.: FDL - A New Approach to Common-Sense Reasoning. Technical Report, Institute of Electronics, Technical University of Silesia, Gliwice, Poland (2008)

A Web-Oriented Java3D Talking Head

O. Gambino[1], A. Augello[1], A. Caronia[1], G. Pilato[2], R. Pirrone[1], and S. Gaglio[1,2]

[1] Università degli Studi di Palermo,Department of Computer Science, Palermo, Italy
{gambino,augello}@csai.unipa.it,
alessandrocaronia@inwind.it, pirrone@unipa.it
[2] Consiglio Nazionale delle Ricerche (CNR), Palermo, Italy
{pilato,gaglio}@pa.icar.cnr.it

Abstract. Facial animation denotes all those systems performing speech synchronization with an animated face model. These kinds of systems are named Talking Heads or Talking Faces. At the same time simple dialogue systems called chatbots have been developed. Chatbots are software agents able to interact with users through pattern-matching based rules. In this paper a Talking Head oriented to the creation of a Chatbot is presented. An answer is generated in form of text triggered by an input query. The answer is converted into a facial animation using a 3D face model whose lips movements are synchronized with the sound produced by a speech synthesis module. Our Talking Head exploits the naturalness of the facial animation and provides a real-time interactive interface to the user. Besides, it is specifically suited for being used on the web. This leads to a set of requirements to be satisfied, like: simple installation, visual quality, fast download, and interactivity in real time. The web infrastructure has been realized using the Client-Server model. The Chatbot, the Natural Language Processing and the Digital Signal Processing services are delegated to the server. The client is involved in animation and synchronization. This way, the server can handle multiple requests from clients. The conversation module has been implemented using the A.L.I.C.E. (Artificial Linguistic Internet Computer Entity) technology. The output of the chatbot is given input to the Natural Language Processing (Comedia Speech), incorporating a text analyzer, a letter-to-sound module and a module for the generation of prosody. The client, through the synchronization module, computes the time of real duration of the animation and the duration of each phoneme and consequently of each viseme. The morphing module performs the animation of the facial model and the voice reproduction. As a result, the user will see the answer to question both in textual form and in the form of visual animation.

1 Introduction

A talking head is a 3D animated model aimed at simulating a human head. The model is capable to emulate the articulation of word pronunciations, by synchronizing an audio flow with labial movements and reproducing emotional expressions of the face. The application fields of this technology include telecommunication [1, 2],

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 295–311.
springerlink.com

teaching [3, 4], speaking rehabilitation [5] and all systems requiring a friendly interaction with a human user. The problems related to a talking head realization can be summed up in three main tasks: viseme-phoneme association, animation and synchronization. The phoneme is the smallest unit of distinguishable sound: the concatenation of phonemes forms words and phrases.

Conversely, a "viseme" identifies the equivalent contractions of face muscles that produce the phoneme sound.

More phonemes can be expressed by the same viseme: the viseme-phoneme association task consists of detecting this connection. Furthermore, the visemes must be linked together in order to simulate the articulation of entire words pronunciation movements. The subsequent task is to smooth the lips movements in order to produce a fluent video stream. The most used techniques to accomplish this task are Keyframe Interpolation [6], Muscle-Based Facial Animation [7] and Direct Parametrization [8].

The last task, consisting in the correct synchronization between sound and animation flows, is obtained adapting the duration of the viseme for the whole duration of the current phoneme. This produces a pleasant audiovisual effect. This step can be realized using a hybrid architecture based on Hidden Markov Models and Artificial Neural Network(HNN/ANN) [9].

This method cannot be applied to realize a real-time synchronization since it requires heavy computational resources. In order to overcame this problem, we introduced a ”linear synchronization”. The talking head presented in this chapter is a real-time interface based on Java3D with full 3D capabilities and it is embedded into a simple conversational agent whose knowledge base is accomplished by pattern-matching rules. It uses the Java3D Morphing technique provided by the animation engine of Java3D and exploits the linear model to perform the Synchronization task.

2 State of the Art

In this paragraph we give an overview of the main works related to talking heads and conversion from text inputs to audiovisual streams. As an example, a 3D head model interactive interface to express a sequence of phonemes and emotions is presented in [10]. Rational Free Form Deformations simulate abstract muscle actions. This technique moves from Free Form Deformation (FFD) depicted in [11]; rational basis functions are included in the analytical formalisation. The face mesh is subdivided into regions and control points impose the deformation. A basic facial motion parameter is called Minimum Perceptible Action (MPA). Each MPA is associated to a set of parameters. Such parameters modify the face mesh. An MPA can be thought as an atomic action unit like the Action Unit (AU) of the Facial Action Coding System (FACS) [12], but it also includes non-facial muscle actions, such as eyes movements.

In [13] a 3D head model is constructed starting from a cylindrical acquisition performed with a Cyberware scanner. A generic face mesh is fitted on the scanner data thanks to an image analysis technique aimed at discovering local minima and maxima of the sampled data. Some facial features (like nose, eyes etc..) are detected so that the ones of the generic mesh fit the feature samples. They propose an accurate biomechanical model for facial animation which has been compared with the one depicted in [14]. Once the 3D model has been created, a dynamic model of facial tissue has been created, a skull surface has been estimated and the most important facial muscles have been inserted into the model.

In [15] the human co-articulation is modelled to improve the animation performance. Co-articulation is related to the articulation modifications of a speech segment in function of the preceding and upcoming segments. Their synthetic visual speech is driven by the gestural theory of speech production detailed in [16]. In this work is a dominance degree among speech segments is defined. This allows making a specific dominant segment more influent on the facial control. The co-articulation is mathematically modelled on the basis of the dominance of two contiguous speech segments.

In [17] a system able to convert input text into an audiovisual speech stream has been presented. It makes use of a wide image collection representing many mouth shapes. Each image is a "viseme" for the corresponding phoneme. The term "viseme" was introduced for the first time in [18] and it is the analogous of "phoneme" for the face. Visemes are obtained by means of a recorded visual corpus of a human subject enunciating one instantiation of each viseme. This text-to-audiovisual speech synthesizer is based on the merging of visemes. The merging task is performed using a morphing transformation. Such an optical flow transformation produces a smooth transition from a viseme to another. The video stream is synchronized with an audio stream thanks to the timing information obtained from a text-to-speech synthesizer.

In [19] some of the authors of the previous work presented a system based on machine learning techniques able to enunciate new utterances that were not recorded in the original corpus.

Video Rewrite[20] is a system able to create a new video of a subject saying words that she did not pronounce in a previously recorded video. The phonemes in the training data and in the new audio track are automatically labelled. Mouth images in the training video are reordered in order to match the phoneme sequence of the new audio flow. Video Rewrite approximates the closest phoneme to the unavailable one. The stitching process positions the sequence of mouth images compensating the head position and orientation between the mouth images and the background footage. Computer-vision techniques are used to track points on the speaker's mouth in the training video and the mouth gestures are merged thanks to morphing technique.

The system proposed in [21] is inspired by [20]. A talking person is recorded on a video and some samples are automatically extracted. These samples are stored in a library. The head is decomposed in a set of facial part. This task reduces the number of samples needed for the synthesis because each facial part can be animated independently. Such technique allows generating facial expression while speaking words.

In [22] the talking-head animations system is composed by two steps. The first is aimed to create an image samples library of facial parts are extracted from a video of a talking person. The second step needs phonetic transcript from a text-to-speech synthesizer (TTS) so that facial part samples can be reassemble into an animation. A coarse synthetic 3D head model of the recorded subject is created by a 3D polygon model and a set of textures. It is composed by few four-sided polygons approximating the face shape and the textures are the sampled facial parts.

In [23] is depicted a system devoted to reproduce on a 3D face model the face expressions of an actress. The system uses six studio quality video cameras to capture the actress face. Sample points on the face are acquired thanks to 182 colored marker points. While the geometric data is tracked, multiple high resolution images of the face are acquired. They are used to create the texture of the 3D face polygonal mesh. A novel compression method based on principal components analysis is used to compress the geometric data and MPEG4 codec is used to compress the texture sequence.

In [24] the concept of the morphable face model, that is a 3D deformable face model, is introduced. Starting from 200 registered 3D face scans (100 male and 100 female), they find a parametric description of faces using the statistics of a dataset. Also facial attributes, like gender, are parameterized. The face is divided into independent sub-regions that are independently morphed. An algorithm matches the deformable model with novel images or 3D scans of faces tuning the parameters of the morphable model. A method to obtain a morphable model from unregistered dataset of 3D face scans is also presented. The dimensionality of the morphable model can be increased adding new faces.

In [25] a generic face model is deformed to match with images of a subject, like in [24]. For this aim, multiple views of a human subject using regular cameras at arbitrary locations, while [24] and many other authors use Cyberware laser-based cylindrical scanners. The price to pay consists in some manual interventions during the process. Some points referred to typical face features, such as tip of the nose, mouth corners and so on, are manually placed on each photograph corresponding to a different view. These markers are used to compute the camera parameters for each view and the 3D marker positions in the space. These geometric positions are used to fit the face of the particular human subject. Other marker points could be necessary to refine the fitting process. A texture map is derived from the set of views. This process must be performed for each desired facial expression, given a human subject. The facial animations are produced by interpolation of contiguous 3D models while at the same time blending the textures.

In [26] is presented an animation approach based on the 3D face shapes during speech. A particular device (Eyetronic's ShapeSnatcher) acquires 3D face reconstructions. They are used to implement a robust face tracker without special markers. The face shapes are depicted by means of PCA. A space of eigen-facemasks is generated and visemes can be expressed in such space.

DECface [27] is able to synchronize the animation of a wireframe face model with the audio flow generated by a TTS. It executes the following operations: an ASCII text is presented as input data ; a phonetic transcription is created from the input text ; an audio flow is generated by speech synthesis ; a query is sent to the audio server and the current phoneme is determined by the speech playback; the mouth shape is computed from nodal trajectories; speech samples are synchronized with the graphics. It makes use of the DECtalk system to perform text-to-speech conversion, the phonemic synthesizer and vocal tract model. The mouth movements are simulated by motion of the vertexes polygonal mesh using a non linear law. The synchronization is performed by computing the duration of the mouth deformation in basis of the duration of an audio samples group.

In [28] an intelligent human-like character driven by input text is described. Its movements and voice intonation are controlled by linguistic and contextual information of the input text. A set of rules allows the mapping from text to body gestures and voice intonation. The set can be enriched with new rules related to new features. The system is accomplished in Java and XML modules. Two XML files encode the knowledge base: the fist one describes objects, while the second one depicts actions.

In [29] a web-based application based on client server architecture implementing a Talking head is outlined. A web browser, a TTS and a facial animation renderer are present on the client. The synchronization between the mouth movements and speech synthesis is based on co-articulation model [15]. Facial expressions are implemented by special bookmarks. The 3D head model is created using VRML.

In [30] the platform INTERFACE is implemented to develop LUCIA [31], a graphic MPEG-4 Talking Head. LUCIA is a female head 3D VRML model speaking Italian by means of FESTIVAL, a diphone TTS synthesizer [32]. The animation engine is based on a modified co-articulation model.

The SAMIR system [33] creates Web-based intelligent agents. A Behaviour Manager (BM), a Dialogue Management System (DMS) and an Animation Module (AM) compose the system. The DMS client/server accomplishes the communication between the user and the BM. BM creates also the most appropriate set of parameters encoding the emotional expressions. The corresponding facial expression is generated by the AM on the basis of that parameters. The facial expressions are implemented thanks to various morph targets [34]. The 3D face model was exported from FACEGEN in Shout3D file format. The morph targets are linearly interpolated by means of the Channel Deformer node in the Shout3D develop environment.

The work detailed in [35] is based on MPEG-4 standard. It allows defining audiovisual objects and provides a text-to-speech interface (TTSI). MPEG4 allows managing 3D VRML models also performing rigid body transform. A neutral face model, a set of feature points and Facial Animation Parameters (FAP) are also defined in the MPEG-4 standard. Face Animation Parameter Units (FAPU) describe the face geometry in such a way that FAP drive the animation.

In [36] is described a system able to adapt a generic wireframe model to a specific the somatic of a new model. The animation parameters and the deformation rules are properly suited by a calibration task. The deformation task is based on Multilevel Calibration with RBF (Radial basis functions) and Model Calibration with texture.

A generic renderer can become MPEG-4 compliant with ADI (Animation Definition Interface) [37]. It is based on wireframe deformation based on VRML IndexedFaceSet node. The system output provides animation FAP parameters similar to the ones in MPEG-4.

In [38] a talking head aimed to human-car means of communication simulating a human-human interaction is presented. It is accomplished by image based rendering driven by a TTS (Text-To-Speech). It consists of an images database where each image is phonetically labelled. Such a task is performed by audio-video recording of a subject reading text of a corpus. An Hidden Markov Model (HMM) recognizes and aligns the audio flow with phonemes. In this manner, each phoneme is associated with the corresponding video frames. A Unit Selection Engine (USE) provides the synthesis of the Talking Head. Given an input text, a TTS provides the speech synthesis, phonemes sequence and their durations. The USE performs the lip synchronization and smooth transitions between images. Its cost function takes also into account the co-articulation. This framework is quite similar to the one presented in [39], where the HMM is substituted with a Viterbi search on a graph. Such a graph is composed by candidate mouth images related to the corresponding phonemes. In [40] the Active Appearance Models (AAM) [41] are improved to accurately detect mouth feature points. These works move from [42] where the main framework was entirely present.

3 An Overview the Talking Head

3.1 Talking Head Features

Artists and scientists have studied the human organs shape and their functions through the centuries. The aim of digital face modeling involves specifically their realism and their natural movement. The artificial face can be subdivided in three main components: the facial bones, the facial muscles and the facial representation of a conversation.

Skeleton and facial muscles
The human face bony structure is composed of two main components: the skull and the facial skeleton. The skull contains and protects the brain, while several bones constitute the facial skeleton, but the lower jaw is the only mobile bone.

The muscles are directly attached to the facial skeleton. The facial muscle contractions give rise to the face movements and expressions. Tendons generate the energy for the muscle contractions. Each phoneme of the human speaking can be visually represented by a facial expression and a particular disposition of the mouth, called viseme.

Face Modeling
Geometric and texture description are fundamental for the facial model development. A face has a complex and flexible structure, due both to texture and color variety and expression imperfections. Even though most faces have a similar structure and characteristics, few facial differences allow discriminating between two similar individuals (for e.g. the monozygotic twins). One of the most relevant challenges of the facial animation consists in designing models able to perform minimal face variations. Some of the approaches oriented to facial geometry definition are reported below. A first class of approaches employs volumetric representation techniques; they can be

implicit and involve voxel (volume rendering), Constructive Solid Geometry (CSG) and Octrees (connected volume elements).

Another class of approaches for the facial geometry definition exploits surface representation techniques. The surface structure allows representing different facial configurations as modular surfaces, so many kinds of faces can be built.

Surface analytical description is typically defined as parametric surface, like B-Spline or NURBS, or polygon mesh.

Facial Component Characteristics

A facial model is also constituted by external parts, named facial components. They form the visible part of a face, like eyes, eyebrow, mouth, tongue, teeth and ears. The use of facial components enhances the naturalness of a face model.

The mouth is more realistic if it is composed by three modules: lips, tongue and teeth. The lips must be designed using curve surfaces with the capability to be flexible, so that the mouth can reproduce all the possible positions. The mouth is fundamental for the representation of emotions and for the speaking simulation; a phoneme should be represented by an appropriate ”viseme”.

Animation

Facial animation started with a sequence of hand-made slides. A present, many toolkits oriented to animation using different approaches are available. Human face animation can be produced using two different manners. The first one provides the desired animation using a technique named ”keyframe interpolation”. Starting from an initial image, a sequence of interpolated images is generated until the final image is reached; both the initial and final images are named ”keyframes”. Some key points both on the initial and final faces must be placed in automatic or semi-automatic manner in order to perform the interpolation of their intermediate positions on the frames. The second approach consists of a pseudo-muscular algorithm that surveys the real facial movements in order to arrange the model of face deformations. These approaches require input devices based on laser scanning or video in order to build the model of face deformations. The face expressions are generated using a restricted number of facial animation parameters (FAP).

Viseme

A viseme is the basic unit of visual field, correlated to a phoneme. It represents the face and mouth articulations that occur during the pronunciation of phonemes.

A viseme may represent more than one single phoneme. This approach reduces considerably the number of visemes that should be taken into account.

Speech Synthesis

The process of speech synthesis consists of transforming a written text in spoken text. In the context of speech synthesis systems, it is unthinkable to memorize all the words of a language, rather it is more suitable to use a voice synthesizer as a system for the automatic creation of speech through the phonetic transcription of the sentence that should be pronounced. The process of Speech synthesis must produce a talk that appears the most natural possible. In a TTS system the mechanism of speech synthesis is carried out through two components. The first component is the NLP (Natural Language Processing), capable of supplying a phonetic transcription of the text with information on prosody. The second component is the DSP (Digital Signal Processing) that transforms the symbolic information received from the NLP module into human voice. The NLP component consists of: a text analyzer that includes the functions of a pre-processor, a morphological analyzer, a contextual analyzer and syntactic-prosodic parser; a letter- to-sound module; and a module for prosody generator.

The pre-processor is deals with the input normalization. It transforms numeric values in words, identifies abbreviations and acronyms and transforms them into full texts. The morphological analyzer gives all possible categories of speech for each part of speech. The contextual analyzer takes account of the words present in the context in order to reduce the list of possible parts of speech. The syntactic-prosodic parser examines the remaining area of research and finds the structure of the text through a prosody-based approach. The letter-to-sound component deals with automatic setting of phonetic transcription of the text input. The module for the generation of prosody deals with the right emphasis of verses. Once that the syntactic-prosodic structure of a sentence has been deduced, we can exploit it to compute the length of each phoneme (and silences), as of intonation that should be applied. The transitions between phonemes are rather important for the understanding of speech. Two approaches, namely explicit and implicit, are available for the phonetic transition control. In explicit phonetic transition, a series of rules that formally describe the influence of adjacent phonemes are exploited. In implicit phonetic transition examples of phonetic transitions, articulations underlying interval of speech and acoustic unit nearly fundamental (diphones) are used. These two methodologies have generated different philosophies for the synthesis: one for the rules (synthesis-by-rule) and one for concatenation (synthesis-by-concatenation).

Conversational Agents

Conversational agents are often used to improve usability of human/computer interfaces. The use of these systems allows a fulfilling interaction and makes the system accessible either by inexpert users. The main issue of conversational agents implementation regards their natural language processing capabilities. In fact, natural language is characterized by several ambiguities which human beings can resolve through their own cultural experiences. Chatbots represent an alternative to advanced dialogue system, which analyze in depth the semantic and syntactic structure of the language. Chatbots can interact with the user using pattern-matching based rules. In the specific case the conversation is carried out by these kinds of agents looking for lexical matching between the user query and a set of question answer modules, called categories, stored in their own knowledge base. Chatbots can store user preferences and information, set up and change the conversation topics and trace dialogue history. In last years there has been an extensive use of chatbots, as interface to e-learning platforms, research engines, e-commerce web-sites.

4 The Proposed Architecture

The proposed system is characterized by a client-server architecture (see Figure 1). The user can dialogue with the chatbot writing in a chat form on the client side. The

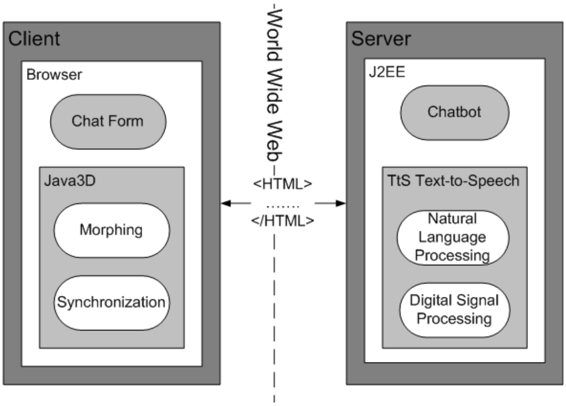


Fig. 1. The client/server system architecture

client side includes the Java3D module, composed of the Morphing and the Synchronization modules. The server side is composed of the Chatbot and the text-to-speech (TTS) modules. The TTS module includes the Natural Language Processing and the Digital Signal Processing modules. The implementation of a talking head involves issues regarding the definition of parameters, such as the shape of speech. Moreover the construction of a web-oriented talking head requires a simple installation, a visual quality, a fast download in order to reduce the size of the client system, involving in this way the development of a model of low complexity for interactivity in real time.

4.1 The Server Side

The Conversational Agent Module

The conversation is carried out implemented by means of chatbot technology. In particular the A.L.I.C.E. (Artificial Linguistic Internet Computer Entity) technology, an Open Source chatbot released under the GNU license, has been used.

A.L.I.C.E. knowledge base is composed of question-answer modules, called categories and described with AIML (Artificial Intelligence Mark-up Language) [42]. The main components of an AIML category are represented by the “pattern” corresponding to the user question, and the “template” corresponding to the chatbot answer. Pattern matching rules allow the dialogue progress. The AIML category is described by the tag <category>, while the pattern and the template are described respectively by the <pattern> and the <template> tags. Special symbols called wildcards allows for a partial matching between the user question and the AIML pattern. Specific AIML tags are used to enhance the dialogue capabilities of the chatbot, for example to set and get values of variables, to execute other programs, to recursively call the pattern matching rules. Table 1 shows an example of AIML code.

Table 1. An example of AIML category

```
<category>
 <pattern> What means the Alice acronym< /pattern >
 <template> It means Artificial Linguistic Internet Computer Entity
 </template>
</category>
```

TTS Module
The chatbot answer is analyzed by the Natural Language Processing (Comedia Speech) embedded into the TTS module. It incorporates a text analyzer, a letter-to-sound module and a module for the generation of prosody. To each sentence is associated a set of phonemes (SAMPA [44]) with their durations and pitches target (prosody). The English language has about 40 phonemes that give rise to about 1,600 diphones [45]. The synthesis is achieved through the frequency concatenation of successive diphones, which are acoustic segments including the transition between two consecutive phonemes. In particular, in English language there are about 1,600 diphones, rising from about 40 phonemes . In phonetics, diphones are acoustic segments that include the transition between two consecutive phonemes.

The information obtained from Comedia by the Natural Language Processing module is then used by the Digital signal processing (MBrola) [46] in order to produce human-like speech. All the obtained information, including concerning phonemes, prosody and talked generated are then sent to the client.

4.2 Client Side

The task of the client side is the evaluation of the real time of duration of animation and the duration of each phoneme and consequently of each viseme through the synchronization module. The synchronization module in particular sends the information of duration to the morphing module, which in turn performs the animation of the facial model and the related reproduction of voice. The final answer is returned to the user both in textual form and in the form of visual animation. The procedure is repeated until the chat session shutdown.

Phonemes-Visemes Relation
An appropriate association phoneme-viseme module has been implemented to add a voice system to the Talking Head. Our solution is based on the representation of text into phonemes with information on prosody. The visual animation has been obtained using visemes segmentation. Number of visemes used is 12, whereof one represents the silence. The 12 visemes illustrated in Fig. 2, have been chosen reviewing the studies made with different possible configurations of associations phonemes/visemes [47-49]. The value chosen has been the one supposed to be optimal among the number of visemes and association phonemes-visemes. We have used the CMU set of phonemes [50] containing 39 phonemes to which we added the phoneme silence to associate the corresponding viseme neutral. Finally a segmentation process, after receiving the input phonemes and information on prosody,

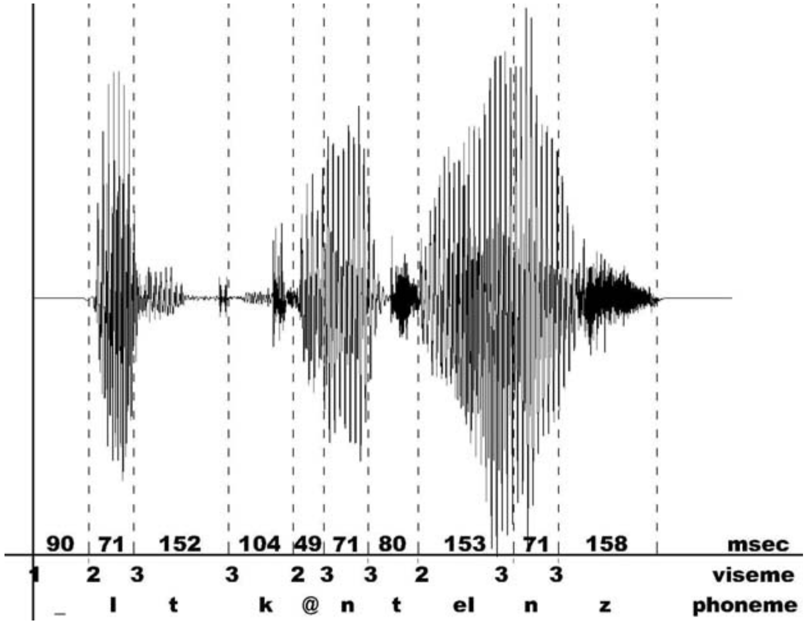


Fig. 2. Waveform and visemes and phonemes associated to the sentence ”It contains”

makes a parser of phonemes by SAMPA at CMU. The parser has been realized with [51] from SAMPA to TIMIT [52] and TIMIT to CMU. The CMU is a slight modification of TIMIT [53].

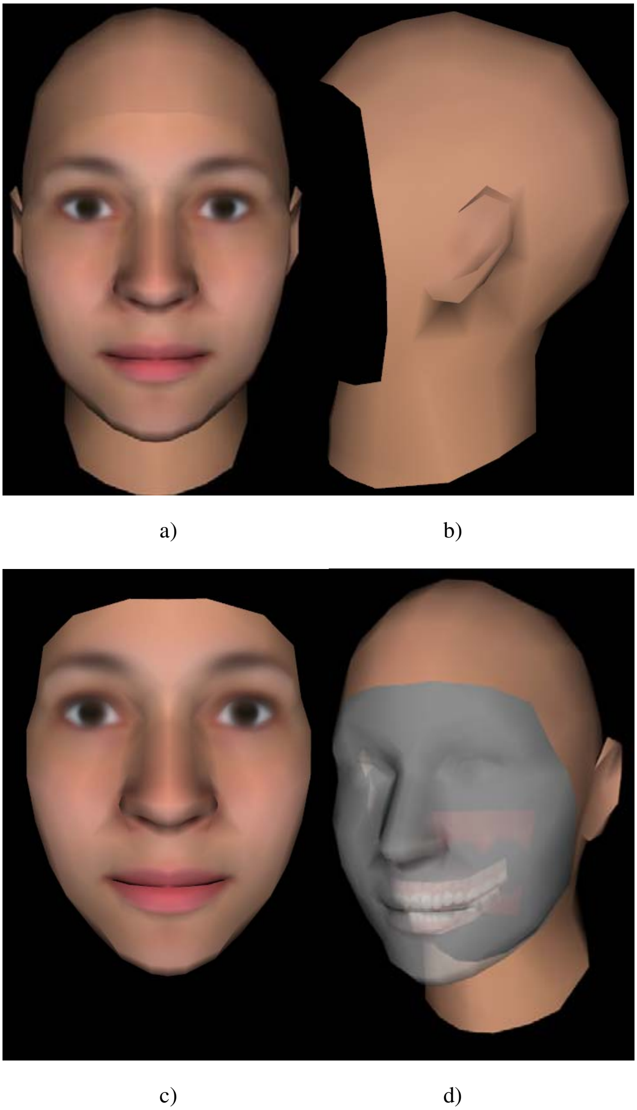
The Synchronization

We have performed a linear synchronization, because we suppose having a limited computational resource, since this process must be executed on the client side.

This solution allows the server handling multiple connections with many clients since this task is performed on each client. The synchronization of audio and graphics are obtained reading, at the beginning, the time necessary to perform the phonemes with the relevant information on prosody, adding it to the initial and final time of silence; then reading the actual duration of the audio file generated by the Digital Signal Processing MBrola. MBrola receives as input the information from the output of NLP. The length of each phoneme is revaluated by using a linear proportion. The real duration of audio entails the total length animation, while the duration of individual visemes has made to coincide with the phonemes duration.

Animation

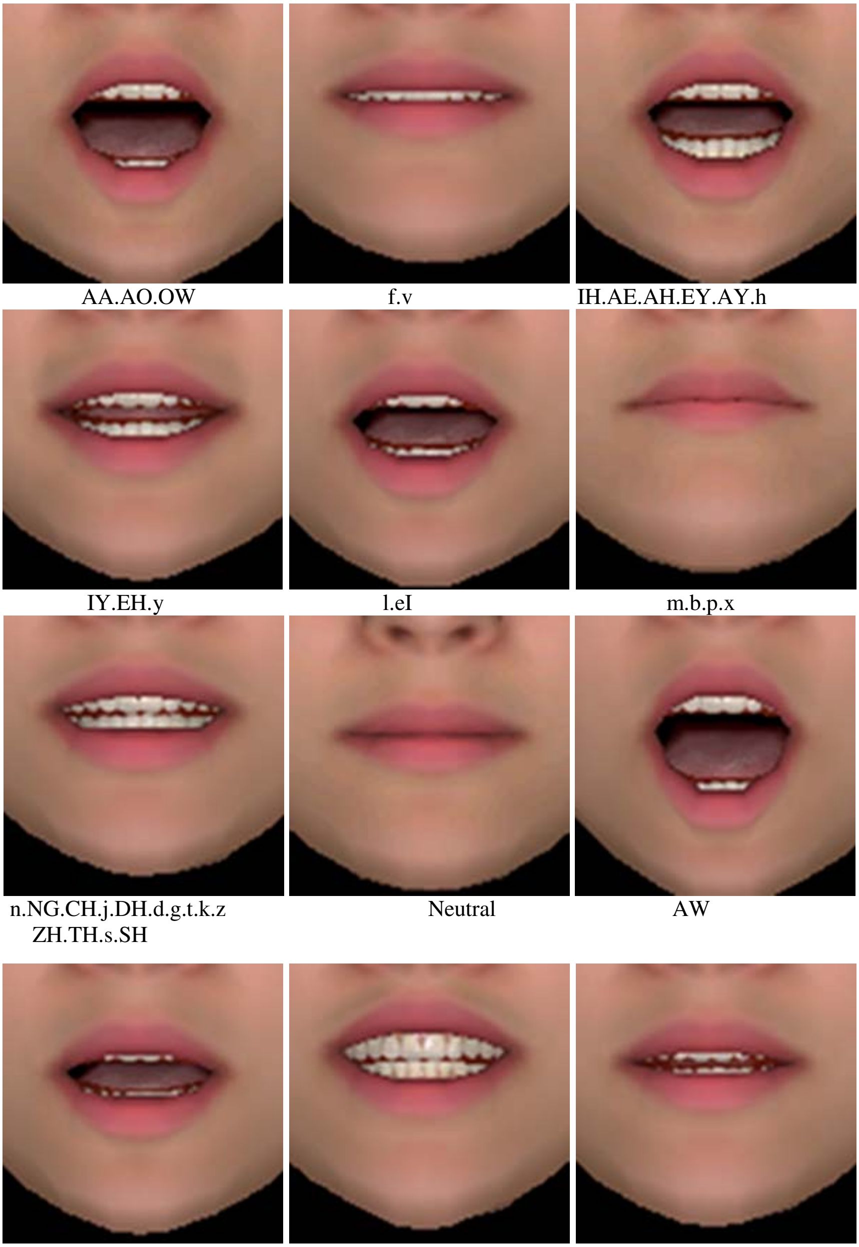
The animation has been obtained importing the Head 3D model from FACEGEN [54] (Fig. 3). Moreover FACEGEN has provided also the visemes associated to English phonemes (see Fig.4). We have modified each viseme to adapt it to the one of Annosoft, which can be considered the main reference in the state of art of the field. Even if the head is composed of 1004 triangular polygons, the face and


a) b)

c) d)

Fig. 3. a) the 3D Facegen model; b) the head; c) the face; d) the internal parts like tongue, mandible and maxilla are inside the volume

lips requires 712 of them. We have animated only the face discarding the rest of the head to obtain a higher speed. The animation is obtained by means of Java3D Morph class, that is a sort of keyframe interpolation. The face vertexes are translated from their initial position with a linear law in a way similar to the work reported in [33] until the reaching of their final position in the last frame. This way allows for the to achievement of a frames sequence.



AA.AO.OW

f.v

IH.AE.AH.EY.AY.h

IY.EH.y

l.eI

m.b.p.x

n.NG.CH.j.DH.d.g.t.k.z
ZH.TH.s.SH

Neutral

AW

r.ER

UW.UH.OY

w

Fig. 4. Each picture shows the mouth gesture of the corresponding viseme

5 Interaction with the System

The interaction is triggered by a user request in the chat form. After that, the chatbot searches in its knowledge base for the AIML pattern matching with the user query. The template in the corresponding category is then processed to obtain the chatbot answer. A set of phonemes and the information about the prosody are then associated to the answer. The prosody information allows a correct labial synchronization. Visemes associated to the phonemes are reproduced in a synchronized way with the audio produced by the vocal synthesizer. In the following example it is shown an interaction between the user and a chatbot acting as a museum guide in the Archaeological Museum of Agrigento [55]. The following AIML category allows the chatbot to answer at the starting of the dialogue.

The chatbot answer ”It contains 18 rooms” is then translated into a set of Sampa phonemes, with the prosody information, as shown in Table 2. In particular the first column represents the chatbot response, and each line belonging to the next columns contains a phoneme name, a duration (in ms), and a series of pitch targets. The pitch targets are composed of two float numbers corresponding to the position of the pitch target within the phoneme (in % of its total duration), and the pitch value (in Hz) at this position.

A possible interaction between the user and the chatbot is the following:

User: Can you give me information about the Archaeological Museum of Agrigento
Chatbot: It is located in the area of Valle dei Templi. It contains 18 rooms.
User: Who designed the museum structure?
Chatbot: The architect Minissi.
User: I need some information about the showcase 8
Chatbot: It contains an attic amphora.
User: Can you give me more information about it?
Chatbot: It belongs to Archaic Greek period and represents Athena

6 Conclusions

In this paper, after a review of the state of the art about talking heads, a web-based Talking Head embedded into a conversational agent has been illustrated. The system allows translating the answers generated by the conversational agent during the dialogue with the user into an appropriate 3D facial animation. In the produced animation the lips movements are synchronized with the sound given by a speech synthesis element. The main problems concerning the Talking Heads, such as viseme-phoneme association, animation and synchronization have been addressed for the system implementation. We have chosen to use the Java3D Morphing technique given by the Java3d animation engine. We exploit the linear synchronization algorithm to solve the synchronization problem between visemes and phonemes in a real-time environment. The system is founded on a web infrastructure.

The Web Infrastructure is realized using a Client-Server model delegating the Chatbot, the Natural Language Processing and the Digital Signal Processing services to the server, while the client is involved in animation and synchronization tasks. This solution allows the server to handle multiple requests from clients.

References

1. Stork, D., Henneke, M.: Speech-reading by humans and machine: Models, systems and applications. Springer, New York (1996)
2. Chen, T., Rao, R.: Audio-visual integration in multimodal communications. In: Proc. IEEE Conference, vol. 86(5), pp. 837–852 (1998)
3. Cosi, P., Magno Caldognetto, E.: E-learning e facce parlanti: Nuove applicazioni e prospettive. Atti delle XIV Giornate del GFS, 247–252 (2004)
4. Biscetti, S., Cosi, P., Delmonte, R., Cole, R.: Italian literacy tutor: Un adattamento all' italiano del colorado literacy tutor. In: Atti DIDAMATICA, Ferrara, Italy, pp. 249–253 (2004)
5. Cosi, P., Delmonte, R., Biscetti, S., Cole, R.A., Pellom, B., van Vuren, S.: Italian literacy tutor, tools and technologies for individuals with cognitive disabilities. In: Proc. STIL/ICALL Symposium, Venice, Italy, pp. 207–215 (2004)
6. Huang, C.F.: A solution for computer facial animation (2000), `http://graphics.csie.ntu.edu.tw/thesis/00MCFHuang.pdf` (accessed April 23, 2009)
7. Parke, F.I., Waters, K.: Computer facial animation. A K Peters Ltd., Wellesley (1996)
8. Rydfalk, M.: CANDIDE, a parameterized face. Report No LiTH-ISY-I-866, Dept. Electrical Engineering, Linköping University, Sweden (1987)
9. Tisato, G., Cosi, P., Drioli, C., Tesser, F.: INTERFACE: A new tool for building emotive/expressive talking heads. In: Proc. 9th European Conference on Speech Communication and Technology, Lisbon, Portugal, pp. 781–784 (2005)
10. Kalra, P., Mangili, A., Magnetat-Thalmann, N., Thalmann, D.: Simulation of facial muscle actions based on rational free form deformations. In: Proc. Eurographics Conference, Champery, Switherland, pp. 59–69 (1992)
11. Sederberg, T.W., Parry, S.R.: Free form deformation of solid geometric models. In: Proc. SIGGRAPH 1986 Conference, Dallas, TX, pp. 151–160 (1986)
12. Ekman, P., Friesen, W.: Facial action coding system: A technique for the measurement of facial movement. Consulting Psychologists Press, Palo Alto (1978)
13. Lee, Y., Terzopoulos, D., Waters, K.: Realistic modeling for facial animation. In: Proc. SIGGRAPH 1995 Conference, Los Angeles, CA, pp. 55–62 (1995)
14. Terzopoulos, D., Waters, K.: Physically-based facial modeling, analysis, and animation. Visualization and ComputerAnimation 1, 73–80 (1990)
15. Cohen, M.M., Massaro, D.W.: Modeling coarticulation in synthetic visual speech. In: Thalmann, M.N., Thalmann, D. (eds.) Models and techniques in computer animation, pp. 139–156. Springer, Heidelberg (1993)
16. Löfqvist, A.: Speech as audible gestures. In: Hardcastle, W.J., Marchal, A. (eds.) Speech production and speech modeling, pp. 289–322. Kluwer Academic Publishers, Dordrecht (1990)
17. Ezzat, T., Poggio, T.: MikeTalk: A talking facial display based on morphing visemes. In: Proc. IEEE Conference on Computer Animation, Philadelphia, PA, pp. 96–102 (1998)

18. Fisher, C.G.: Confusions among visually perceived consonants. J. Speech Hearing Res. 11, 796–804 (1968)
19. Ezzat, T., Geiger, G., Poggio, T.: Trainable videorealistic speech animation. In: Proc. 6th IEEE International Conference on Automatic Face and Gesture Recognition, Southampton, UK, pp. 388–398 (2004)
20. Bregler, C., Covell, M., Slaney, M.: Video rewrite: Driving visual speech with audio (1997), `http://cims.nyu.edu/~bregler/VideoRewrite.pdf` (accessed April 24, 2009)
21. Cosatto, E., Graf, H.: Sample-based synthesis of photo-realistic talking heads. In: Proc. IEEE Conference on Computer Animation, Philadelphia, PA, pp. 103–110 (1998)
22. Cosatto, E., Graf, H.P.: Photo-realistic talking-heads from image samples. IEEE Transactions on Multimedia 2(3), 152–163 (2000)
23. Guenter, B., Grimmy, C., Woodz, D., Malvary, H., Pighinz, F.: Making faces. In: ACM SIGGRAPH International Conference, Montreal, Canada (2006)
24. Blanz, V., Vetter, T.: A morphable model for the synthesis of 3D faces. In: Proc. Internatonal Conference on Computer Graphics and Interactive Techniques, pp. 187–194. ACM Press/Addison-Wesley (1999)
25. Pighin, F., Hecker, J., Lischinskiy, D., Szeliskiz, R., Salesin, D.H.: Synthesizing realistic facial expressions from photographs. In: Proc. ACM SIGGRAPH International Conference, Montreal, Canada, pp. 75–84 (2006)
26. Kalberer, G.A., Van Gool, L.: Lip animation based on observed 3D speech dynamics. In: Sabry, F., Gruen, A. (eds.) Proc. SPIE Conference. Videometrics and optical methods for 3D shape measurement, vol. 4309, pp. 16–25 (2001)
27. Waters, K., Levergood, T.M.: An automatic lip-synchronization algorithm for synthetic faces. In: Proc. 2nd ACM International Conference on Multimedia, San Francisco, CA, pp. 149–156 (1994)
28. Cassell, J., Vilhjálmsson, H.H., Bickmore, T.: BEAT: the behaviour expression animation toolkit. In: Proc. 28th Annual Conference on Computer Graphics and Interactive Techniques, Los Angeles, CA, pp. 477–486 (2001)
29. Ostermann, J., Millen, D.: Talking heads and synthetic speech: architecture for supporting electronic commerce. In: Proc. IEEE International Conference on Multimedia and Expo., New York City, NY, pp. 71–74 (2000)
30. Tisato, G., Cosi, P.: New interface tools for developing emotional talking heads. In: Proc. Lang. Tech. Conference, Rome, Italy, pp. 53–56 (2008)
31. Cosi, P., Fusaro, A., Tisato, G.: LUCIA: a talking-head based on a modified cohen-massaro's labial co-articulation model. In: Proc. Eurospeech Conference, Geneva, Switzerland, pp. 127–132 (2003)
32. Cosi, P., Tesser, F., Gretter, R., Avesani, C.: Festival speaks Italian! In: Proc. Eurospeech Conference, Aalborg, Denmark, pp. 509–512 (2001)
33. Abbattista, F., Catucci, G., Semeraro, G., Zambetta, F.: SAMIR: A smart 3D assistant on the web. Psychology J. 2(1), 43–60 (2004)
34. Fleming, B., Dobbs, D.: Animating facial features and expressions. Charles River Media, Hingham (1998)
35. Tekalp, M., Ostermann, J.: Face and 2D mesh animation in MPEG-4. Image Communication J. 15(4-5), 387–421 (2000)
36. Lavagetto, F., Pockaj, R.: The facial animation engine: Toward a high-level interface for the design of MPEG-4 compliant animated faces: circuits and systems for video technology. IEEE Transactions 9(2), 277–289 (1999)

37. Ostermann, J., Haratsch, E.: An animation definition interface: rapid design of MPEG-4 compliant animated faces and bodies. In: International Workshop on Synthetic/Natural Hybrid Coding and 3D Imaging, Orlando, FL, pp. 216–219 (1997)
38. Liu, K., Ostermann, J.: Realistic talking head for human-car-entertainment services. In: IMA 2008 Informationssysteme für mobile Anwendungen GZVB eV, Braunschweig, Germany, pp. 108–118 (2008)
39. Liu, K., Ostermann, J.: Realistic facial animation system for interactive services. In: International Conference Interspeech, LIPS 2008: Visual Speech Synthesis Challenge, Brisbane (2008)
40. Liu, K., Weissenfeld, A., Ostermann, J., Luo, X.: Robust AAM building for morphing in an image-based facial animation system. In: IEEE International Conference on Multimedia and Expo., Hannover, Germany, pp. 933–936 (2008)
41. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. IEEE Transactions on Pattern Recognition and Machine Intelligence 23(6), 681–685 (2001)
42. Cosatto, E., Graf, H.P., Ostermann, J.: From audio-only to audio-and-video text-to-speech. Acta Acustica united with Acustica 90(6), 1084–1095 (2004)
43. Artificial Intelligence Markup Language (AIML), http://en.wikipedia.org/wiki/AIML (accessed April 24, 2009)
44. Festvox, http://www.festvox.org/festtut/notes/festtut_7.html (accessed April 24, 2009)
45. Mbrola, http://tcts.fpms.ac.be/synthesis/mbrola.html (accessed April 24, 2009)
46. Lucey, P., Martin, T., Sridharan, S.: Confusability of phonemes grouped according to their viseme classes in noisye environments. Presented at 10th Australian International Conference on Speech Science & Technology, Macquarie University, Sydney (2004)
47. Annosoft LLC, http://www.annosoft.com/phoneset.htm (accessed April 24, 2009)
48. Chen, T.: Audiovisual speech processing. IEEE Signal Processing Magazine 19, 9–21 (2001)
49. The CMU Pronouncing Dictionary, http://www.speech.cs.cmu.edu/cgi-bin/cmudict#phones (accessed April 24, 2009)
50. Roch, M.: IPA/CMU/TIMIT phone mappings and American English examples. CS 696 Acoustic Modeling for Speech & Speaker Recognition
51. Garofolo, J.S., et al.: TIMIT Acoustic-Phonetic Continuous Speech Corpus. Linguistic Data Consortium, Philadelphia (1993)
52. Anderson, T.R.: Auditory Models with Kohonen LVQ for Speaker Independent Recognition SOFM and Phoneme. In: Proc. IEEE conference on Neural Network, pp. 4466–4469 (1994)
53. Singular Inversions Inc., FaceGen Modeller, http://www.facegen.com/modeller.htm (accessed April 24, 2009)
54. Pilato, G., Augello, A., Santangelo, A., Sorce, S., Genco, A., Gentile, A., Gaglio, S.: MAGA: a mobile archaeological guide at agrigento. IEEE Pervasive Computing 5(2), 54–61 (2006)

A VR-Based Visualization Framework for Effective Information Perception and Cognition

H.Y.K. Lau, L.K.Y. Chan, and R.H.K. Wong

Department of Industrial and Manufacturing Systems Engineering,
The University of Hong Kong, Hong Kong
hyklau@hku.hk, lkychan@hkusua.hku.hk, rockywonghk@gmail.com

Abstract. An information visualization framework is proposed that considers key human factors for effective complex data perception and cognition. Virtual reality (VR) technology with data transformation heuristics are deployed in building the framework where an interactive VR-based 3-D information visualization platform is developed. The framework is applied to develop a visualization system for an express cargo handling center where analysts are able to effectively perceive operation details and carry out timely decision making.

1 Introduction

Information processing is one of the key factors that affects human performance in complex information understanding. It involves perceiving data and transforming information to some human understandable form (e.g. cognition). Often, the efficiency of information processing is greatly affected by the complexity of information and the processing time may increase exponentially when the data set is large and complicated. With the proliferation of information technology, system analysts are facing an increasing challenge to process huge amount of data in order to perceive, understand and retain such information. In order to improve the understandability of data and shorten the system learning process, a VR-based information processing and representation framework, which includes a method and algorithms for information processing, plus an interactive user interface for data representation based on Virtual Reality (VR) technology is proposed in this research to transform complex system data (e.g. generated from system simulation) into information which both management and analysts would find easy to visualize, readily understand and get familiar with. In particular, the user interface capitalizes on the powerful features provided by virtual reality technology for 3-D stereoscopic visualization, immersive information perception, and multi-dimension sensation recreation that effectively implement user friendly features including fast-forwarding and rewinding of computer generated sceneries, user-definable viewing cameras configurations, and displaying fully-textured vivid computer graphic entities without sacrificing the processing speed. In addition, the data transformation algorithms of the framework reduce human error due to the misinterpretation of information.

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 313–332.
springrlink.   com   Springer-Verlag Berlin Heidelberg 2009

By adopting the proposed information processing framework, not only the system data can be vividly presented; the user interface enhances the effectiveness of system analysis and the system learning processes, and it is envisaged that the information processing time can be considerably shortened. Moreover, transforming data into a format that can be represented and visualized effectively and vividly is one of the key features provided by the proposed framework.

2 Literature Review

Information visualization is a popular research topic in recent years. Technology adopted to provide vivid visualization is advancing rapidly in both research and application domains, and different visualization systems are proposed and developed. As summarized by Tory and Möller [1], information visualization generally provides the following four features:

1. Data analysis can be enhanced when the data is visually represented for better understanding.
2. Concretize the user's mental model through the visual display; help verify the ideas and hypotheses, etc.
3. Mental model can be ameliorated if supporting and/or contradictory evidence for the hypotheses made can be obtained.
4. Organization and sharing of data can be achieved without huge effort.

However, most of the researches focused on achieving the first feature and they mainly concentrated on how data can be processed in a more efficient manner in order to provide the desired visualization effect in a very short time. Solely presenting information with well-textured graphical objects and fancy layout does not mean successful information visualization. The key to help users better perceive and understand information in order to make appropriate judgment relies on human perception and cognition, and more effort should be made to fulfill the other three features as listed above. In our proposed visualization framework, these four features are well-covered and therefore the system developed based on the framework is able to present data which users find it efficient to perceive and understand with reduced effort.

2.1 Perception of Complex Data

There are numerous researches undertaken regarding the perception difference between different kinds of information, such as text and image. Of these means, pictures (including graphs) and texts are commonly used. Pictures capture concrete and spatial information appositely while texts capture abstract and categorical information neatly. Each of these approaches suits well in a particular situation but not others. For example, a picture is better suited than texts to describe "How is a cat look like?", while texts are more suitable to describe something abstract like "What is peace?" To further categorize the different approaches, focused attentive processing is required for detailed information (e.g. text) while pre-attentive processing is needed for perceiving information in general (e.g. image). One will need more effort and mental energy to process detailed information, and the effect will be more significant when the complexity of data increases.

Based on this phenomenon, if the complex data sets can be transformed into a form such that less resource is required to process such information, and at the same time it facilitates the perception and cognition of the information. Information visualization is one of the approaches which serve this purpose. However, it is not surprising that there is certain degree of misunderstanding of visualization. Zhang et al. [2] pointed out that visualization does not mean to replace quantitative analysis. Instead, it complements the quantitative analysis and outshines the important parameters to be selected in order to perform appropriate analysis. In our proposed framework, a hybrid approach is adopted so that users can effectively capture the salient features and important scenario at a glance, identifies the field of interest quickly without being distracted and required to study in detail the complex information to start with. Besides, the advantages of different approaches can be integrated while the weaknesses can be complemented.

2.2 Current HCI Approaches for Data Visualization

There are number of information visualization approaches proposed by experts in the field of human factors. In terms of textual information visualization, researches such as spatial mappings [3-4] and geographic metaphors [5] were explored. Besides visualizing textual information, techniques for data visualization and analysis in different forms such as FilmFinder [6], Glyph-based visualization of an oceanography data set [7] and interactive visualization platform adopted by New York Stock Exchange (NYSE) [8] demonstrated that information visualization is a topic which is gaining increasing attention in the research community. Other information visualization related researches in the field can be found in [9-13].

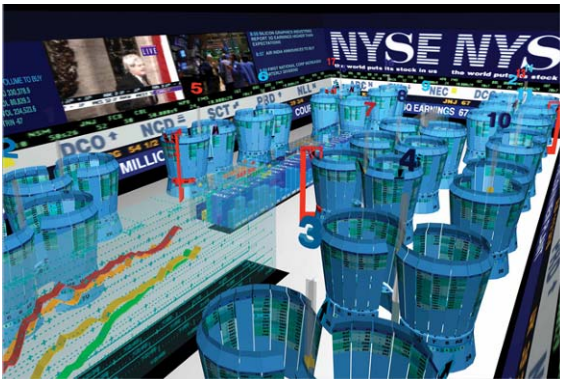


Fig. 1. Interactive visualization was adopted to help supervising the NYSE

Besides Virtual Reality, there are several areas where a lot of research effort has been made. As a result of the great advancement of computer and information technology, the Augmented Reality (AR) [14-16] platforms and immersive technology are now being actively researched and deployed.

AR technology refers to the real-time interaction between computer-generated data and the real physical world. Users are able to work with virtual data in real physical scenes and therefore AR provides a more intuitive and easy interaction between virtual entities and the real world objects/environment. AR application for information visualization is one of the fast-growing research fields for human-computer interaction. do Carmo et al. [17] developed an Augmented Reality environment that supports coordinated and multiple views, is one of the typical examples of the recent AR research development for information visualization.

For immersive technology, a classical example is the application of Cave Automatic Virtual Environment (CAVE) [18]. Typically, the CAVE produces stereoscopic 3-dimensional graphics that are rendered from the viewer's perspective. A traditional CAVE uses active / passive stereoscopic viewing devices to recreate the visual effective of stereo images and with the fully surrounding images, a completely immersive virtual environment can then be created. As CAVE provides a highly flexible platform for 3-D visualization, it has been adopted in various research and development projects. Some of these examples include the visualization of software objects and their relationships in complex software development [19], virtual exploration [20] and capturing of human skills for performance evaluation [21].

Although there are numerous information visualization applications developed, many of the current approaches are specific to particular application domains and not many general frameworks have been developed. To fill the gap, this paper proposes a generic framework for the effective visualization of information so as to facilitate the perception and cognition of complex information based on the technology of VR.

2.3 Minimizing Semantic Gap between the Data and the Reality

As stated by Tory and Möller [1], one of the key issues that must be addressed in information visualization is to concretize the user's mental model towards a scenario. There are plentiful of visualization platforms available in the field, such as virtual classroom [22], visualization of partial floor plan of Xerox PARC [23] and the UM-AR-GPS-ROVER Software [24]. All these application platforms have a common characteristic – to show how a system / object should look like. The purpose of this is to minimize the semantic gap between one's concept and the reality. If one's concept is not aligned with the actual fact correctly, misunderstanding of data or fact may occur. Therefore, the importance of minimizing the semantic gap between the concept and the reality should be emphasized, especially when large amount of complex data is presented to an analyst.

In the presentation of the generic visualization framework, this paper focuses on the development of a VR visualization system for the domain of logistics facility system performance analysis. Many of logistics facilities and systems consist of complex structure, such as large scale automated storage and retrieval systems (ASRS) and computer controlled conveyors systems that are found in distribution centers and logistics hubs in which complex interactions between different entities exist in the systems. Simply presenting figures or even graphs produced from system data may not be helpful for quick analysis and understanding of system performance and operation because of complexity and it may be difficult to appreciate the interrelationships between the information. For example, if a system is 80% utilized, it is sometime difficult for users to perceive the actual operation and physical scenario of such a system. Without any knowledge about the actual or physical system, management and even operators will find it very difficult to 'visualize' the situation. However, if the information can be understand and analyzed with the help of 'visualizing' the actual scenarios that occur in the physical facilities, the semantic gap can be significantly reduced and parties involved no longer need to interpret the meaning of the data with sheer vagueness. In addition, appropriate indication should be given in order to enhance the understandability of the information, such as visually annotating an entity, complementing numeric data/graph to a particular part of the system (for example, displaying the actual queue length of a long queue of goods which is difficult to count manually).

2.4 Focused Attention to Avoid Distraction

In the domain of logistics in a modern distribution center, by conducting a system simulation using state-of-the-art simulation tools such as AutoMod® [25] and FlexSim® [26], analysts are able to collect huge amount of data from these simulation systems. Given the complexity of such data sets, it is likely that analysts can easily be distracted with this volume of information, especially when some of these data are irrelevant or redundant. Moreover, additional mental workload is required to process the large amount of data which also degrades user performance. As such, attention issue should also be considered, and this issue has been studied widely by researchers including De Weerd [27], Duncan [28], Motter [29], Posner & Fernandez-Duque [30] and Rao [31]. In their research, attention is referred to as the fact that we can only process a limited amount of information actively from a huge amount of information available through our senses, stored memories, and other cognitive processes. Since it is common that system analysts need to handle large amount of complex data in a limited time period, in order to help the analysts to carry out the analysis in an effective and efficient manner, information visualization should be performed with the following characteristics:

- Unrelated information which is potentially distractive to the desired task should be filtered or hidden from the scene.
- Important information that should be focused on should be emphasized, so that the desired task can be conducted in an efficient manner.

2.5 Mental Workload

Gopher and Donchin [32] mentioned that mental workload is an attribute of the information processing and control systems that mediate between rules, stimuli, and responses. Mental workload level determines analyst's performance in certain extent. When the mental workload for a particular task is low or medium, one may not find the task to be difficult, therefore it often leads to satisfactory performance of such a task. However, if the mental workload is perceived to be very high, an analyst may not handle the task effectively and it leads to degraded performance.

As mentioned previously, system analysts deal with a large amount of numerical data in their everyday work. Without proper tools for data pre-processing, processing large amount of numeric data can be an exhaustive task, which implies a high mental workload to the analysts because a large amount of energy is required to process texts compared with images or other means. Therefore, proper tools for data pre-processing become necessary to reduce the mental workload to a reasonable level in order to improve analysts' performance. Our proposed framework aims to reduce the mental workload of complex data processing through several approaches such as reducing the need of creating conceptual linkage between data and user knowledge, and transforming part of text or number to graphical entities in VR for efficient and effective perception and understanding.

3 Proposed Visualization Approach

In order to vividly present the system data and enhance the effectiveness of information cognition and perception, a methodology based on VR technology is proposed to perform data transformation and visualization. Fig.2 shows the proposed general framework for information transformation and visualization that includes the following five main steps.

1. Preparation of raw data that may be generated from a simulation model of a system or other means
2. Classification of data
3. Transformation of system data into pre-defined data structures
4. Data manipulation in the VR system
5. Information visualization by the VR platforms

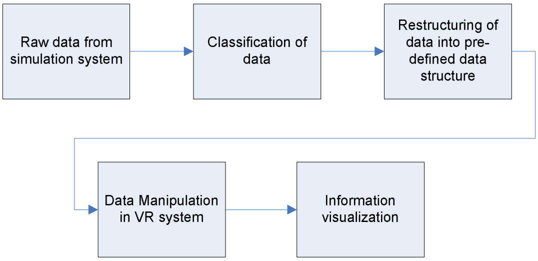

Raw data from simulation system
Classification of data
Restructuring of data into pre-defined data structure
Data Manipulation in VR system
Information visualization

Fig. 2. The proposed framework for information visualization

3.1 Preparation of Raw System Data

Data transformation process transforms raw system data into a format that can readily be utilized by the visualization platform. In generating these raw data, several means are adopted including taking references to operational records, collecting outputs of simulation system, etc. The data format is transformed into a compatible format with the VR visualization platform so that rapid transformation of data can be performed in real-time with least computation effort. In our case study, simulation system output files are used because most simulation systems are able to generate well-formatted data for subsequent data processing, and the formatted files are readily used by the VR visualization platform.

To further elaborate, if simulation approach is chosen to generate the data needed to be visualized in the visualization system instead of using operational records, simulation system developer should design carefully how this process should be carried out. Appropriate simulation software should be chosen and a simulation model should be developed. System behavior should be modeled correctly to ensure data accuracy logged in the system performance files because the VR visualization platform entirely relies on these files as the system input. If the simulation system is not modeled correctly, the visualization platform will present misleading visual information that does not reflect the true scenarios.

In the case study, we take the data generated by the simulation tool AutoMod® of an express air cargo handling center as the input to the visualization platform. Fig. 3 shows the screen shot of the simulation system developed using AutoMod® simulation tool.

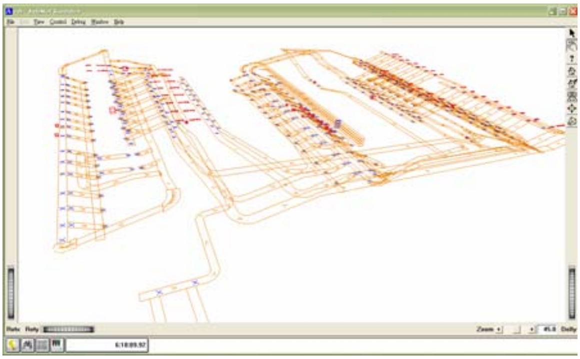


Fig. 3. A simulation system developed using AutoMod®

3.2 Data Classification

When the raw system data is prepared, data classification is then performed. In general, there are three main types of data that should be classified and grouped.

1. Data that can be directly visualized

These are information that can be visualized graphically. One of these examples in the case of the simulation of the operation of an express cargo center is the number of workers in a particular work zone. Based on the number of workers in the raw system data, corresponding number of workers can be allocated in the virtual environment to reflect the actual situation.

2. Data which cannot be visualized

Some of the information is very difficult to be visualized effectively but useful for complementing information for better understanding of the virtual environment. This kind of information should be kept in the visualization platform in different format, such as text. In the context of express cargo handling center, the cycle time of a particular parcel going through the system is very difficult to be visualized. However, it can be easily presented in numerical format. Such information should be grouped together and system designers should discuss with end user how the information is best presented.

3. Data which is not useful for visualization or analysis
In order to visualize information in a way that user will not get distracted because of too much information, data filtering is necessary to remove such data that is not useful for analysis no matter data can be visualized or not.

Data classification is crucial in information visualization because the purpose of the platform is to highlight/emphasize important parameters to be selected and eliminate non-useful information to avoid distraction. Fig. 4 show a General Data Classification Tree (GDCT) that provides guidelines for system developers to decide on what and how data should be presented in order to facilitate information perception and cognition.

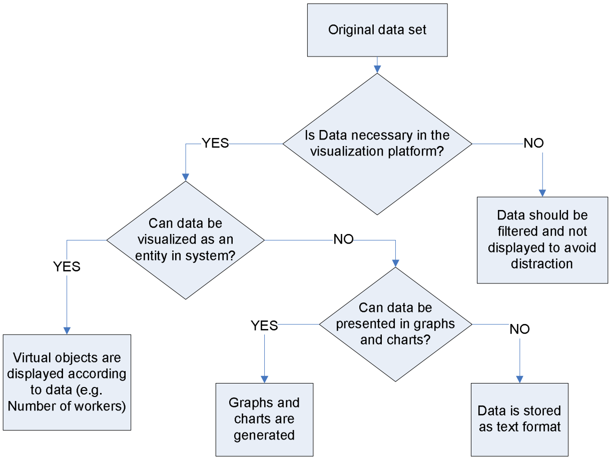

Original data set
Is Data necessary in the visualization platform?
YES
NO
Data should be filtered and not displayed to avoid distraction
Can data be visualized as an entity in system?
YES
NO
Can data be presented in graphs and charts?
YES
NO
Virtual objects are displayed according to data (e.g. Number of workers)
Graphs and charts are generated
Data is stored as text format

Fig. 4. A General Data Classification Tree

3.3 Transformation of System Data

After classifying the raw system data, transformation of data can be performed. The transformation mechanism consists of sets of predefined rules which determine what information should be generated as system output. The transformation rules serve two main purposes:

1. The transformation of data involves calculation of the key performance indicators such as utilization of resources, efficiency, time delay of a process, etc. of the system being analyzed. In the case of an automated material handling system in an express cargo handling center, the key performance indicators include system utilization, average queue length, cycle time, number of worker deployed, etc. The key performance indicators show the system behavior and performance at different time slots in a quantitative manner.

2. Besides calculating the key performance indicators, the transformation rules determine what and how graphical objects should be presented in the VR visualization platform. In the context of the visualization of simulation data that is obtained from a simulation study of an automated material handling system, if a particular part of the system has a queue containing 10 objects in a certain time period, the visualization platform should display a queue of 10 objects in the system to reflect the system status. With the presentation of data in a qualitative manner, efficient data perception can be achieved in a very short time.

In general, the construction of transformation rules varies for different application domains. Here are some guidelines that should be considered:

1. Location of where graphical objects will be displayed should be defined to mitigate misperception (e.g. Operators do not fly in the air).
2. Data might have to be conditioned (e.g. rounding or truncation) so that objects can be presented in a discrete manner. In the case of express cargo handling center, if there are 7.68 pieces of parcel stacked in a work zone in average, 8 pieces of parcel should be presented instead because displaying 7.68 pieces of parcel which may confuse the system user. However, the original data should be kept in the system for subsequent processing.
3. If entities are correlated, data integrity and consistency have to be ensured. Relationship between different objects should be added if it is not addressed before transformation.
4. Data extrapolation may have to be performed if data is not sampled continuously. Better data perception can be expected because the data presented will change in a more natural manner; again, it is reminded that original data set should be kept.

3.4 Visualization of Transformed Information in the VR Platform

To present the information in the virtual environment, a fully-textured virtual system model should be created using 3D graphics development tool such as Autodesk® 3ds® Max [33] or Autodesk® Maya® [34]. This virtual model should be developed based on the actual dimensions and construction of the physical facility in such a way that user will have better experience and understanding of how the system looks like without visiting the physical facilities because the correct dimensions and scale enable more natural perception. Besides the facilities, entities of the system, such as operators and loads, should also be modeled.

When the data transformation is completed, the information will be imported into the virtual environment. In the case of the express cargo handling center, corresponding computer graphics will be prepared in the virtual facility, such as number of operators in certain part of the system, number of loads in a queue, etc. Besides, transforming the numerical data into graphs and charts is a classical yet useful approach adopted in the framework. Purely visualizing the numerical data in virtual environment can only provide a general concept of how the situation is. It provides only a quick conceptualization but the level of details is not high enough for system analysis. To overcome this shortfall, figures and graphs can be used. Charting complementing the visualization in virtual environment offers an excellent means for system analyst to capture valuable information in a relatively very short time effectively.

In design of the VR-based user interface, in addition to having an aesthetic presentation, color contrast technique is used in enhancing the understandability of simulation result. Color contrasting is one of the common approaches that facilitate the process of learning and understanding. By highlighting an object with bright color, human attention can be easily drawn and the information searching time can be greatly shortened. Different colors used in the same object carry different meanings, in the case presented, using solid color in the objects represents average loading while using transparent color represents the maximum loading. Coloring also provides a convenient means to distinguish a group of objects from the others, and it can shorten the identification of areas of interest for system analysis to focus on.

4 Case Study – Implementation of the Framework to Visualize the Operation of an Express Cargo Handling Center

Based on the conceptual framework of the proposed information transformation and visualization methodology, a case study was conducted in visualizing the simulation data of an express cargo handling center. The system performance data is generated by a simulation model of the express cargo handling center according to several system objectives including the deployment of automated material handling facilities, machine specifications, human resources allocation plan and processing time of different operations. Detailed time studies of the system and its corresponding operations were carried out to enable accurate generation of simulation results. The assumptions made in the development of simulation model were collaboratively defined with logistics system administrators and management. In order to arrive at an accurate model, exact dimensions of the facility as defined by the CAD drawings are used by the simulation model as well as the virtual model. The AutoMod® simulation tool was used to build the simulation model and to conduct the simulation exercise to generate the operation data whereas Maya was used to develop the 3-D objects including the facility, human operators, vehicles, express cargoes that are salient to the physical system.

When the development of the simulation model and the virtual model were completed, the VR data transformation platform was developed. As stated in the previous section, the visualization platform is an interactive VR application that analysts can interact with. Virtools® [35] is used for the application development because it supports interactive features in 3-D application plus other desirable properties that are found in a typical virtual reality system. When the 3-D objects created using Maya were imported into Virtools® environment, developers are able to “give lives” to the objects and system users can interact with these objects with different kinds of input peripherals such as a mouse and keyboards. Raw data contained in data files generated from the AutoMod’s system were used as the source for data visualization, and the 3-D objects created using Maya were regarded as entities in the virtual facility that is associated to the corresponding entities in the simulation model.

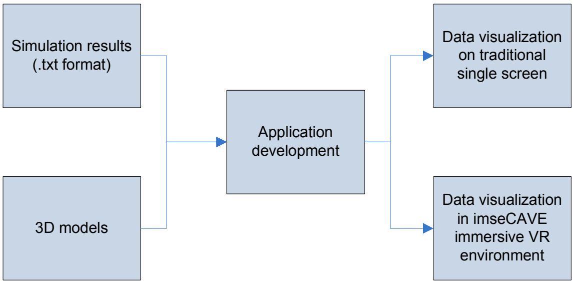

Simulation results (.txt format)
3D models
Application development
Data visualization on traditional single screen
Data visualization in imseCAVE immersive VR environment

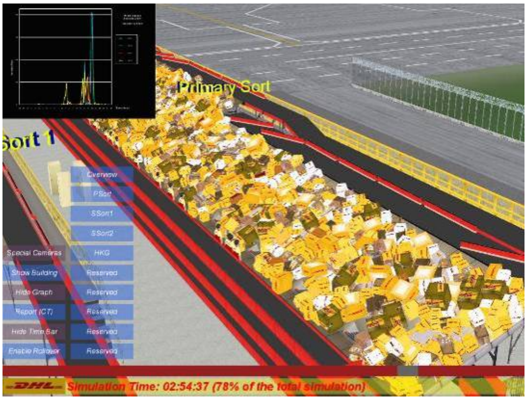
Fig. 5. Flow of information in the visualization platform

In the Virtools® application development environment, information grouping was first performed. As there were number of system files that contain information concerning the operation and performance of the express cargo center, grouping of information is required to identify the relevant data that is needed to be represented in the VR information visualization platform. In this case, information such as number of objects on the mechanized conveyors and inclined cargo ramps (slides), number of operators in different work zones and number of loads (express cargo) queuing to be handled were grouped and sorted. Besides, calculation of key performance indicators is performed, such as calculating the utilization of the inclined cargo ramps based on its maximum capacity and number of parcels stacked. Other information that is not readily visualized due to their intrinsic meaning such as cycle time of each individual load and the utilization of operators were grouped into another category where they are presented quantitatively in the form of charts and actual figures.

After importing the raw information, the system determines the objects that need to be displayed in the corresponding format. In the proposed system, some of the information are statically represented (e.g., maximum number of operators available) and some of the information is dynamically represented (e.g., level of loads or cargo queue length in different time slots). For those objects that change over time such as number of parcels shown on the slide, the system will update the scene and reflect the change accordingly as the time slot changes.

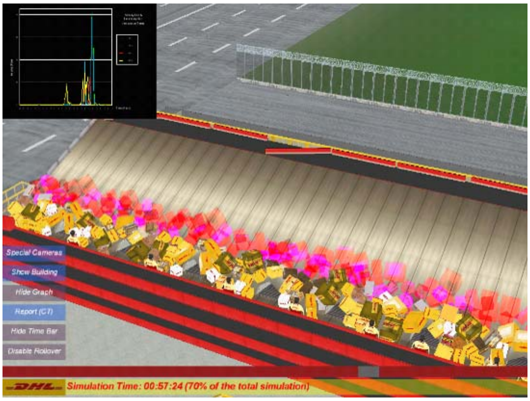
It is believed that by combining actual numerical information and graphical presentation can effectively enhance the understandability of information [1]. To achieve the best visualization for human perception, 2-D charts are included in the visualization platform so that users can analyze the scenario by complementing the "actual quantitative" status of the system and aligning the scenes with the corresponding charts at the same time.

In order to review the system status at different moment in time, a special feature of forwarding and rewinding was implemented in the VR information visualization platform. By dragging the time bar at the bottom of the screen, user is able to see the status of the system at a particular time slot. An indicator is also shown in the performance chart in the upper left corner when user drags the time bar button, and users can select the particular time slot accurately.



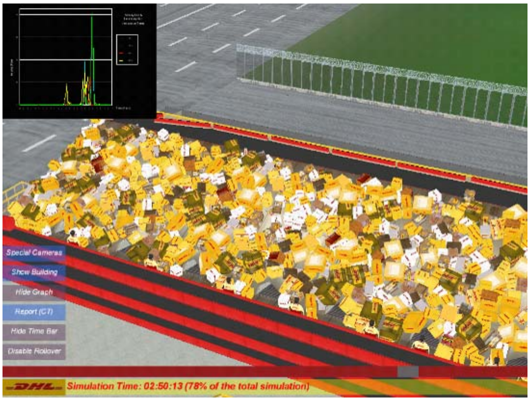
Simulation Time: 02:54:37 (78% of the total simulation)

Fig. 6. System status complementing with actual performance figures that are given by a time graph (upper-left corner)



Simulation Time: 00:57:24 (70% of the total simulation)

Fig. 7. The representation of low utilization level of a cargo slide



Simulation Time: 02:50:13 (78% of the total simulation)

Fig. 8. The representation of high utilization level of a cargo slide

The ability to 'fly through' a facility in 3-D is one of the very important features that tremendously helps users understand a complex physical facility such as in the case of the express cargo handling center in this case study. Users can navigate and travel freely around in the virtual environment by using an intuitive positioning device such as a mouse or joystick. This feature enables the users to effectively understand the layout and operation flow of the whole facility. In addition, the visualization platform is equipped with user definable camera views for efficient navigation to specific locations in the cargo center such as those illustrated in Fig. 6 – 8 and Fig. 10 – 11. Different camera positions and orientations can be defined by the users and can be stored in the system for subsequent reference. Once these camera views are defined in the system, users can select the corresponding on-screen buttons and the system will navigate directly to the view points of these dedicated locations with the desired viewing angles. Analysts who are not familiar with the visualization system can operate the system in an efficient manner by using the predefined camera views.

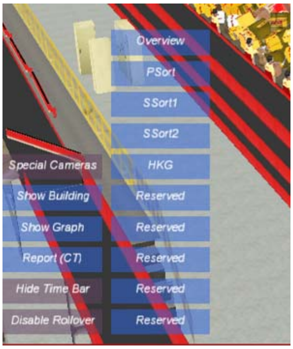


Fig. 9. Specific camera views can be defined by the users

In order to effectively represent the different 3-D objects in the visualization system, a coloring scheme was designed so that the 3-D objects that appear in the virtual scene carry corresponding visual information. For example, the express cargoes on the slides have two major colors. Cargoes in solid color refer to the average level of the loading while cargoes in semi-transparent red represent the maximum level of the loading on a slide in a particular time slot. This coloring scheme provides extra information that helps users appreciate the system performance holistically. In addition, color contrast technique was adopted to show grouping of objects. When the mouse pointer rolls over one of the operators in a particular work cell, all the operators belonging to the same work cell will be highlighted. Moreover, there are annotations above the workers showing the work group a particular worker belongs to. This feature makes the identification of objects in the same work group much easier and clearer.

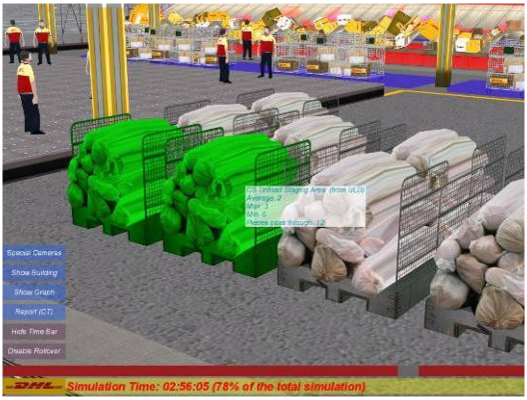


Fig. 10. Cargoes in the same work group are highlighted when they are selected

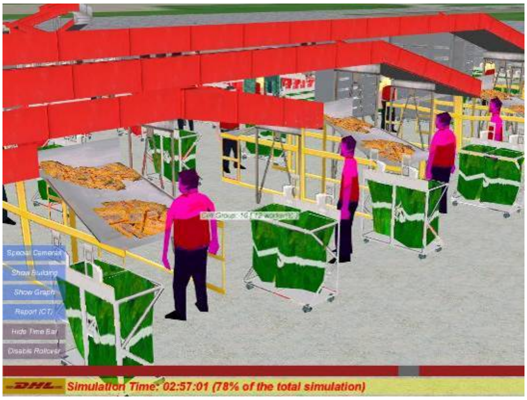

Cell Group: 16 [12 worker(s)]
Special Cameras
Show Building
Show Graph
Report (CT)
Hide Time Bar
Disable Rollover
DHL
Simulation Time: 02:57:01 (78% of the total simulation)

Fig. 11. Operators who are working in the same work group are highlighted when they are selected

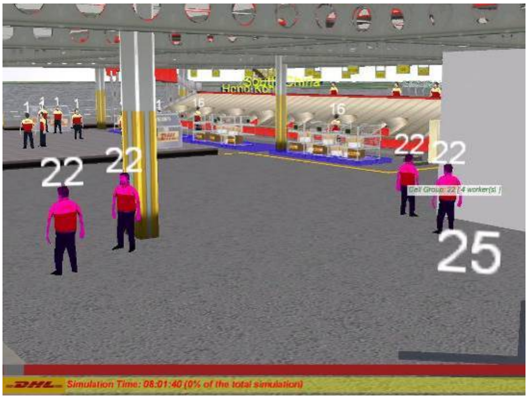

1
1
1
1
1
16
16
Hong Kong
South China
22
22
22
22
Cell Group: 22 [4 worker(s)]
25
DHL
Simulation Time: 08:01:40 (0% of the total simulation)

Fig. 12. Annotating objects with specific ID for easy object identification

5 Information Visualization Using the ImseCAVE VR System

The imseCAVE system in the Department of Industrial and Manufacturing Systems Engineering at The University of Hong Kong was developed based on the concept of the CAVE. The imseCAVE consists of three 10 foot by 8 foot projection walls and a 10 foot by 10 foot silver screen as the floor projection screen (Figure 13). With the use of networked computers, high resolution LCD projectors, and specially designed polarizing lenses, a high performance passive stereoscopic projection system is obtained. To enable computer generated models and images to be simulated, a cluster of networked PCs are used as the virtual reality engine and these PCs are linked with a dedicated high bandwidth Ethernet network. User interfaces including wireless joysticks, tracking devices, etc. are integrated to provide an ergonomic means to interact with the virtual reality models in real-time in a fully immersive virtual environment [36].

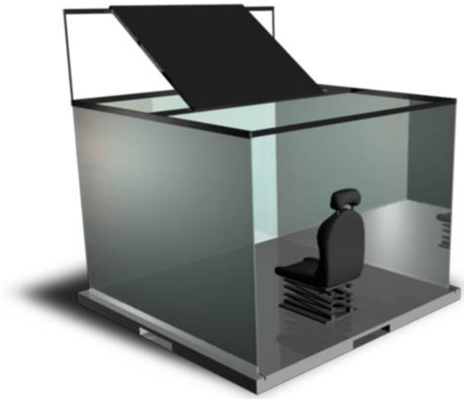


Fig. 13. An illustration of the configuration of imseCAVE

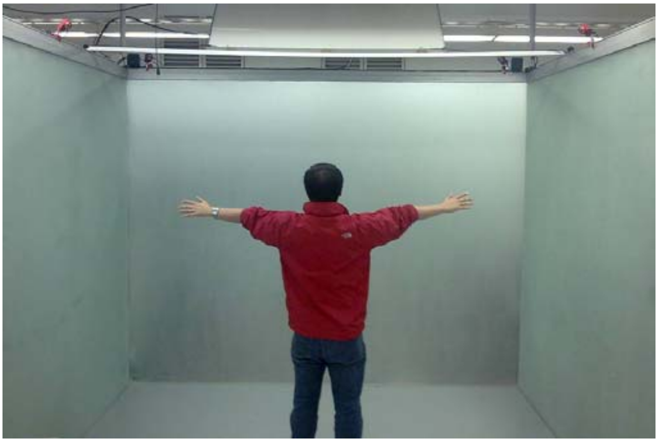


Fig. 14. The actual setting of the imseCAVE. Infra red cameras are attached at the top of the imseCAVE

Besides the VR-based visualization platform that is adopted in the single screen application, imseCAVE is also adopted and an immersive visualization platform is resulted. Since the imseCAVE provides an immersive environment and the user is essentially inside the 3D model, a user will need to operate the visualization software in imseCAVE which the operation mode is different comparing with using single screen application. The user using the imseCAVE wears a pair of polarized glasses to visualize stereo images and flies around the express cargo center by using a wireless joystick. The user can also fly to a pre-defined set of locations displayed in the virtual scene by pressing corresponding buttons on the joysticks. The pre-defined set of locations is defined based on the salient locations that are determined by the stakeholders, including operators, system analysts and the management. In addition, an adjustable slide on wireless joystick is used to switch to different epochs in time in the simulation period. Movement of the slide is directly mapped onto the movement of the time bar. In order for a user to precisely highlight or select virtual objects, an optical tracking system is developed. Reflective dots (Fig. 15) are attached on the polarized glasses and the position of these dots is tracked in real-time by a set of high resolution and high frame rate infra-red cameras, which are attached to the top of the imseCAVE for effective head movement tracking. (Fig. 14)

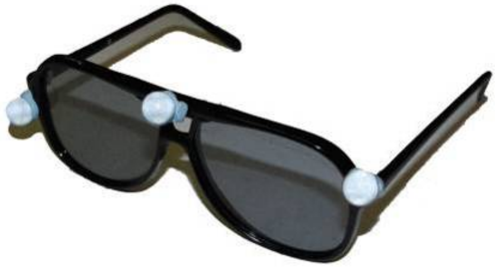


Fig. 15. Polarized glasses with reflective dots attached

By analyzing the images captured by different cameras, accurate spatial movement of the user's head can be obtained. Using the captured 3-D position and orientation of the analyst's head, the system displays the viewing direction in the imseCAVE by emitting a virtual ray with green cursor which shows the viewing direction of the analyst. When the virtual ray intersects with an object in the scene, the corresponding object's information will be displayed on the floating board. In the single screen version of the visualization platform, there is a feature which shows the 2-D chart for the system objects, such as worker utilization and system queue length. The same feature has also been incorporated in the imseCAVE version. As there are several screens in the imseCAVE, there is sufficient space to display the 2-D charts. In the implementation, the left screen of the imseCAVE was chosen for 2-D chart display.

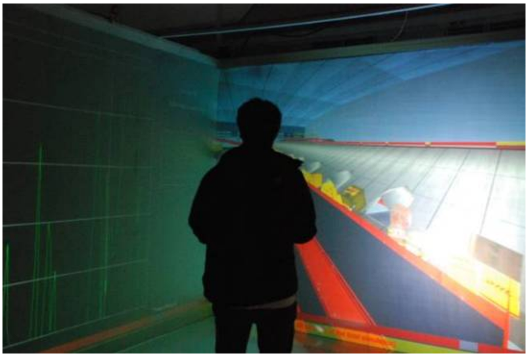


Fig. 16. Left screen of the imseCAVE presenting the graphs of the utilization of the cargo slide, and the other screens display the situation of the express cargo handing facility in a VR environment

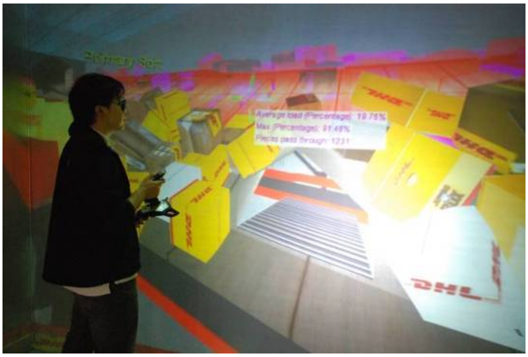


Fig. 17. Head-motion tracking enables easy object picking

6 Discussion and Future Work

An implementation of the proposed VR-based information visualization platform has been successfully undertaken under two different hardware platform, i.e., with a single screen display and using the imseCAVE based on an express cargo handling center. The system was demonstrated to users from a wide spectrum including personnel from the express cargo center, people in the logistics industry, engineers, university students and researchers. The feedback obtained was very positive and in particular, the management of the express cargo center found the system helpful in facilitating the understanding of the overall operation of the system within a very short time comparing to studying raw data. Originally, system analysts needed to spend a lot more time in processing the raw data so that a detailed analysis can be performed. With the help of the proposed system, information of key performance indicators can be obtained within minutes. This will enable, management to better perceive the performance of the express cargo handling operation with the use of a combination of 2-D charts and 3-D virtual sceneries. Users of the information visualization platform agreed that they can now perform decision-making more effectively and efficiently because of the proficient perception and cognition of the information.

Moreover, different versions of the applications have their own strength. The single screen version of the platform is highly portable and ready to use in most of the computer as many of the computers have equipped keyboards and mice. The implementation of the proposed visualization system using the imseCAVE provides another experience to the analysts. Viewing the surrounding environment in the imseCAVE makes analysts feel that they are physically located in the facility. With the use of polarized glasses, stereoscopic view of the virtual entities can be perceived and analysts can visualize the real facility in 3-D. The spatial sensation effect has been further enhanced, and analysts would have better understanding of the facility.

Although the framework proposed has proven success in the case study described, a generic framework which serves a wider range of users and domain is our target in future research. With this direction in mind, an interactive AR visualization framework is under development. It is expected that such a framework can contribute new finding in the field of AR, which further enhance the performance and capabilities of human-computer interaction.

Acknowledgment

The authors would like to express their gratitude to DHL Aviation (Hong Kong) Limited for providing the screen captures of the information visualization system for illustration.

References

1. Tory, M., Möller, T.: Human factors in visualization research. IEEE Transaction on Visualization and Computer Graphics 10(1), 72–84 (2004)
2. Zhang, L., Tang, C., Song, Y., Zhang, A., Ramanathan, M.: VizCluster and its application on classifying gene expression data. Distributed and Parallel Databases 13(1), 73–97 (2003)
3. Rennison, E.: Galaxies of news: An approach to visualizing and understanding expansive news landscapes. In: Proc. 7th Annual ACM Symposium on User Interface Software and Technology, Marina del Rey, CA, USA, pp. 3–12. ACM Press, New York (1994)
4. Rennison, E.: Personalised galaxies of information. In: Proc. Conference on Human Factors in Computing Systems: Demonstration Companion, Denver, CO, USA, pp. 31–32. ACM Press, New York (1995)
5. Harve, S., Hetzler, B., Nowell, L.: ThemeRiver: Visualizing theme changes over time. In: Proc. IEEE Symposium on Information Visualization, Salt Lake City, USA, pp. 115–123. IEEE Computer Society Press, Los Alamitos (2000)
6. Ahlberg, C., Shneiderman, B.: Visual information seeking: Tight coupling of dynamic query filters with starfield displays. In: Proc. ACM Conference on Human Factors in Computing Systems, Boston, MA, USA, pp. 313–321. ACM Press, New York (1994)

7. Healey, C.G.: On the use of perceptual cues and data mining for effective visualization of scientific datasets. In: Proc. Graphics Interface, pp. 177–184 (1998)
8. Asymptote, NYSE 3D trading floor (1998), `http://www.asymptote.net` (accessed on April 27, 2009)
9. Ferreira de Oliveira, M.C., Levkowitz, H.: From visual data exploration to visual data mining: a survey. IEEE Transactions on Visualization and Computer Graphics 9(3), 378–394 (2003)
10. Keim, D.A.: Information visualization and visual data mining. IEEE Transactions on Visualization and Computer Graphics 8(1), 1–8 (2002)
11. Jankun-Kelly, T.J., Ma, K.-L., Gertz, M.: A model and framework for visualization exploration. IEEE Transactions on Visualization and Computer Graphics 13(2), 357–369 (2007)
12. Lau, A., Moere, A.V.: Towards a model of information aesthetics in information visualization. In: Proc. 11th Conference on Information Visualization, pp. 87–92 (2007)
13. Chittaro, L.: Visualizing information on mobile devices. Computer 39(3), 40–45 (2006)
14. Gausemeier, J., Fruend, J., Matvsczok, C.: AR-planning tool: designing flexible manufacturing systems with augmented reality. In: Proc. of the Workshop on Virtual Environments, Barcelona, Spain. Proc. ACM Conference Series, vol. 23, pp. 19–25 (2002)
15. Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., Maclntyre, B.: Recent advances in augmented reality. IEEE Computer Graphics and Applications 21(6), 34–47 (2001)
16. Takacs, G., Chandrasekhar, V., Gelfand, N., Xiong, Y., Chen, W.C., Bismpigiannis, T., Grzeszczuk, R., Pulli, K., Girod, B.: Outdoors augmented reality on mobile phone using loxel-based visual feature organization. In: Proc. 1st ACM Conference on Multimedia Information Retrieval, Canada, pp. 427–434 (2008)
17. do Carmo, R.M.C., Meiquin, B.S., Pinheiro, S.C.V., Almeida, L.H., Godinho, P.I.A., Meiquins, A.S.G.: Coordinated and multiple views in augmented reality environment. In: Proc. 11th Conference on Information Visualization, Zurich, pp. 156–162 (2007)
18. Cruz-Neira, C., Sandin, D.J., DeFanti, T.A., Kenyon, R., Hart, J.C.: The cave automatic virtual environment. Communications of the ACM 35(2), 64–72 (1992)
19. Bonyuet, D., Ma, M., Jaffrey, K.: 3D visualization for software development. In: Proc. IEEE Conference on Web Services, pp. 708–715. IEEE Computer Society, Washington (2004)
20. Vote, E., Feliz, D.A., Laidlaw, D.H., Joukowsky, M.S.: Discovering Petra: archaeological analysis in VR. IEEE Computer Graphics & Applications 22(5), 38–50 (2002)
21. Kim, J.H., Hayakawa, S., Suzuki, T., Hirana, K., Matsui, Y., Okuma, S., Tsuchida, N.: Capturing and modeling of driving skills under a three dimensional virtual reality system based on PWPS. In: Proc. 29th Conference of the IEEE Industrial Electronics Society, vol. 1, pp. 818–823 (2001)
22. He, D., Banerjee, P.: Enhancing simulation education with a virtual presentation tool. In: Proc. 36th Conference on Winter Simulation, Washington, DC, pp. 1734–1739 (2004)
23. Robertson, G.G., Card, S.K., Mackinlay, J.D.: Information visualization using 3D interactive animation. Communications of the ACM 36(4), 57–71 (1993)
24. Behzadan, A.H., Kamat, V.R.: Visualization of construction graphics in outdoor augmented reality. In: Proc. 37th Conference on Winter Simulation, Orlando, Florida, pp. 1914–1920 (2005)

25. Applied Materials, Inc. AutoMod Simulation Software Solutions - Applied Materials (2007), http://www.brookssoftware.com/pages/245_automod_overview.cfm (accessed April 27, 2009)
26. Flexsim Software Products, Inc. Flexsim – General Purpose Simulation Software (2008), http://www.flexsim.com/software/flexsim (accessed April 27, 2009)
27. De Weerd, P.: Attention, neural basis of attention. In: Nadel, L. (ed.) Encyclopedia of cognitive science, vol. 1, pp. 238–246. Nature Publishing Group, London (2003)
28. Duncan, J.: Attention. In: Wilson, R.A., Keil, F.C. (eds.) The MIT encyclopedia of the cognitive sciences, pp. 39–41. MIT Press, Cambridge (1999)
29. Motter, B.: Attention in the animal brain. In: Wilson, R.A., Keil, F.C. (eds.) The MIT encyclopedia of the cognitive sciences, pp. 41–43. MIT Press, Cambridge (1999)
30. Posner, M.I., Fernandez-Duque, D.: Attention in the human brain. In: Wilson, R.A., Keil, F.C. (eds.) The MIT encyclopedia of the cognitive sciences, pp. 43–46. MIT Press, Cambridge (1999)
31. Rao, R.P.N.: Attention, models of. In: Nadel, L. (ed.) Encyclopedia of cognitive science, vol. 1, pp. 231–237. Nature Publishing Group, London (2003)
32. Karwowski, W.: International encyclopedia of ergonomics and human factors. CRC Press, Inc., Boca Raton (2006)
33. Autodesk. Autodesk - Autodesk 3ds Max (2007), http://usa.autodesk.com/adsk/ser-vlet/index?id=5659302&siteID=123112 (accessed April 27, 2009)
34. Autodesk. Autodesk - Autodesk Maya (2007), http://usa.autodesk.com/adsk/ser-vlet/index?siteID=123112&id=7635018 (accessed April 27, 2009)
35. Virtools. Virtools, A Dassault Systèmes Technology (2007) , http://www.vir-tools.com (accessed April 27, 2009)
36. Lau, H., Chan, L.: Interactive visualization of express cargo handling with the imse-CAVE. In: Proc. Conference on Virtual Concept, Biarritz, France (2005), http://www.virtualconcept.estia.fr/2005/index.php?page=13 (accessed May 15, 2009)

Part IV
Robots and Training Systems

From Research on the Virtual Reality Installation

F. de Sorbier de Pougnadoresse, P. Bouvier, A. Herubel, and V. Biri

Université Paris-Est, Laboratoire d'Informatique de l'Institut Gaspard Monge (LABINFO-IGM), France
{fdesorbi,bouvier,herubel,biri}@univ-mlv.fr

Abstract. This article presents a new virtual reality installation inspired by the CAVE system, but in effect non expensive and transportable. Moreover, special attention was focused on the feeling of presence. This means user should be able to feel his self-presence in the virtual environment and the presence of the virtual objects and entities. Keeping in mind to increase the feeling of presence, we finally developed a dedicated game based on an immersive environment, consistent interactions and emotion.

1 Introduction

Virtual reality (VR) aims at plunging one or more users at the heart of an artificial environment where they are able to feel and interact in real time thanks to sensor-motor interfaces. The key-point of a VR experiment is that users respond through real actions and emotions to stimuli which materialize the virtual environment (VE) and the virtual events.

VR is a leading area which is mainly exploited in high-end fields like simulators, scientific research, army and industry. Reasons why VR is not more generalized are mainly the cost of the system, the large room needed and the skills required. However, there is a growing demand for the use of VR in fields like games, artistic installations and education. That's why we design a non-expensive (7000€) *CAVE*-like system [1] which is also easily transportable in order to be deployed in school or during show.

Our system use four screens of size 3mx3m, a 3D sound rendering with headphones and various interaction tools like the *Nintendo Wiimote* or a 3mx3m home-made sensitive carpet. After an overview of the concepts related to presence, we will describe our VR installation. Finally, we will outline the results of the evaluation of the system.

2 Virtual Reality and Presence

2.1 The Role of Presence

VR experience makers often intend to create the conditions required to make users feel presence. This feeling of presence could be seen shortly as a strong feeling to

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 335–345.
springrlink.com

exist in the VE. This means, user's *being* in the VE has to be the most natural as possible in the sense without the awareness of the virtuality of the situation [2]. Even if it has not been yet strongly established that presence improve the efficiency of the experiment [3, 4], presence remains an important aims to tend toward in that it motivates the researches on several levels: new interaction devices, new algorithms etc.

The feeling of presence can be seen from three angles [5]. First, the spatial presence (or *being there* [6]) which is to feel like at a different place. Self-presence, here it's not *being there* but just *being*, that is projection of user's ego in the role he is supposed to incarnate in the VR application. Social presence is the *being with*, it relates to the presence of other intelligences.

One understands that to tend toward such unstable psychological state, it is crucial that users accept to get caught up in the experiment. Then, our work is to provide an experiment sufficiently credible -but not necessarily realistic- in order to delude user's sense and critical thinking. For this purpose we identified five inter-related technical and psychological pillars on which we can lean on: human cognition knowledge as a foundation, immersion, interaction, consistency of the sensori-motor loop and emotions. These pillars, that are inclined to be sufficiently generic to fit with any kind of application field, have to be seen as guidelines for VR designers. We point the fact that there is no hierarchy between the pillars, some are technical, other are more human side. Sometimes they work on their own, sometimes they work in conjunction, the aim still remains to arouse presence in a synergistic manner.

2.2 The Pillars of Presence

As we already said human must be at the heart of a VR system that's why it is central to understand how we perceive our environment and other humans, how we construct a mental representation of the environment etc. Moreover this may also permit to understand “the mental and neural processes that may underlie presence” [2]. That's why we consider human cognition knowledge as a root for other pillars. We could lean on the headways in neuropsychology, social cognition, psychology of emotions etc. This knowledge may guide the VR system design (how to delude user's senses?) but also the experiment itself (scenario, atmosphere).

The second pillar is immersion. It is achieved through the stimulation of user's senses in order to generate sensations which enable, sometimes thanks to an illusion, the perception of the virtual environment. From this perception will ensue a proper comprehension of the virtual environment and consequently its appropriation. At the stage of immersion there is absolutely no reference to presence or *being there*. The immersion is measurable: does the system provide a stereoscopic display? What about 3D sound spatial-propagation? Use a sensory substitution? Show aesthetic high-defined textures etc. It is obvious that immersion will be more complete [7] if several senses are stimulated in coherence. Last point, most of the time the virtual environment is rendered as it would be if no user were there. We forget that, if someone is plunging in a virtual environment, by his simple inactive presence he has an effect on this environment. So we claim that immersion must be bi-directional, this will improve the credibility of the environment.

The third pillar, interaction has to fulfill two tasks: to acquire information (for example via a tactile system) and to communicate the information (for example to point out or manipulate an object, to modify the environment, to navigate etc.). Moreover, by inducing that user and virtual entities exist in the VE because they interact, interactions possibilities will help to improve the feeling of presence. If we refer to the concept of "perceptual illusion of non mediation" from [8] we understand that interaction devices must be as transparent and natural as possible until being forgotten. Moreover, direct body interactions could improve presence because these latter enable to match virtual actions with real or at least expected kinesthetic feedback. It could be the case if a virtual movement corresponds to a real movement like walking for example. Finally, until now we only emphasized communication between user and the environment but of course that can include a communication with autonomous agents or other human.

The fourth pillar consists in maintaining an action-perception or sensor-motor loop consistent. It will be necessary to respect two main points. Firstly, we will have to take care not to break the causality link between user's actions and the system's feedback. That means we have to implement real-time algorithms and provide high frequency displays. Secondly, we must maintain the time and place consistency between various sensory stimulations associated with the same event or virtual object. This fourth pillar constitutes a link between the last two ones but can also include some works related to cross-modal illusion [9, 10] and multimodality [11].

Even by providing some high quality immersion and interaction which respect the consistency of the action-perception loop, it will probably remain a distance between users and the role they are suppose to incarnate in the experiment. This distance which links user to reality may have several causes: the system's shortcomings, the real world distractions [12] or just because the experiment is not motivating. We argue that technology can not, on its own, create the feeling of presence [13, 14]. Here is the role of the emotion pillar. By introducing emotion in the experiment we want to cancel this distance and encourage users to forget these disturbances and accept to get caught up and stay focused on the experiment. Finally, we think that emotions and presence self support themselves in a virtuous circle: firstly, emotion enables to reach presence more easily and then presence permits to feel more intensely emotions [14].

In section 3 we will explain how we put some of these ideas in concrete form through different choices.

2.3 Virtual Reality Installations

Head mounted displays [15] (*HMD*) are often associated with the concept of virtual reality because they seem to be well-suited for immersive purposes. That kind of device which is quite intrusive because of weight is composed of two screens allowing stereoscopic display. Strength and weakness of a *HMD* is that users are roughly and completely cut off from the real environment. Therefore, the loose of fixed points contributes to faze user who will be unable to keep a motionless position. Moreover, its cost will restrain its use to single user's applications. Finally, it is necessary to operate a tracking of user's position to maintain the consistency between the virtual content and the orientation of the viewpoint. For these reasons, we prefer to concentrate on screen-based installations.

The *CAVE*-system [1] relies on large screens to increase the feeling of presence, placing user at the center of a cube. Each face is a screen on which stereoscopic images are back-projected. That way, user is visually plunged into the virtual world (and drawbacks of *HMD*'s are almost removed). However, like *starcave* [16], *SAS*, *blue-c* [17] or A*llosphere* [18], a *CAVE*-like system requires large space to ensure a correct display area. Furthermore, such an installation is hardly transportable due to its size and meticulous settings that are required. Finally, this system may be expensive because of devices used, maintenance cost and framework needed.

These limits have led to propose some alternatives. The area taken up by the VR installation can be reduced by replacing the projective devices with well-adapted ones. For example, *The Varrier* [19] suggests the use walls from autostereoscopic screens instead of the back-projected screens, but this solution is very expensive. Trying to ease the displacement of an installation, *MiniVR* [20] and *Cyberdome* [21] suggest solutions based on a small screen or a curved screen.

To cut prices, some researchers (e.g. [22, 23] and *HIVE* [24]) have followed a self-made approach with consumer grade components. In addition, these installations are easily transportable because they are light and simple to assemble. Concretely, these VR installations bring flexibility and control of the system.

3 About the Developed Virtual Reality Installation

3.1 Outlook

Here are the constraints we identified in order to create a VR system less expensive and easily transportable. Software has to be home-made or free open-source, bulk of our system has to be designed with consumer grade components and if no low-cost solution exists for a device we create it. Moreover, the framework has to be as light as possible, easy to assemble, a medium room like a class room must contain it and the whole system has to be transportable in a commercial vehicle.

Despite these constraints we still want to provide an experiment which enables to feel presence. So, we will have to take care of several points according to each pillar. For immersion: the quality of the video-projector (luminosity, resolution, frequency, size of the projection plane), multi-sensorial rendering (sound, tactile). For interactions, we provide natural devices which are not too much intrusive but offer a good precision (*Wiimote*, a sensitive carpet for the management of user's moves). For the consistency of the sensor-motor loop: low latency system based on a client-server architecture. Finally, in order to find some subterfuges in the application, to make user forget the shortcomings of the installation.

3.2 Description of the Installation

The Screens

To visually immerse the user we choose to use a *CAVE*-like display system but using only the four vertical faces of the cube. The upper face is not used because the installation is designed to be used in rooms with classical height (roughly 3.50m). Each screen is hand-made using tracing paper which is resistant and allows back-projection. A video-projector is placed behind each screen and is calibrated as for the image perfectly matches its surface. This layout prevents user to cast his shadow when working close to the screen.

Currently, the visual installation is monoscopic but can easily be evolved to stereoscopic display. The installation can be upgraded with twice more video-projectors and special screens which support a polarized projection system [25]. However the number of required devices does not allow us to experiment stereoscopic projection yet. Active devices could also be used but need to be synchronized and are still expensive.

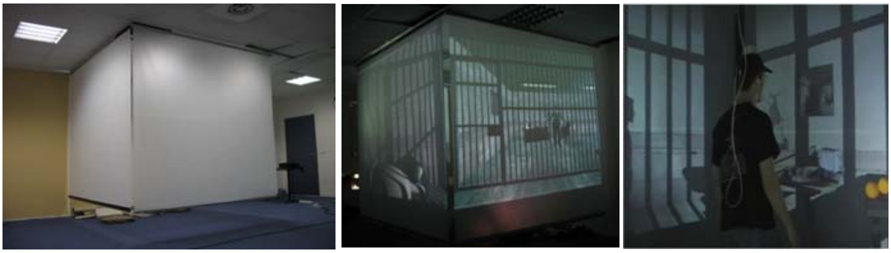


Fig. 1. Different views of the virtual reality installation

Sound Spatialization

To render the virtual acoustic environment two kinds of devices can be used: headphones or loudspeakers each having their advantages and shortcomings. Headphones are generally considered as more intrusive than loudspeakers what hinders the immersion. But headphones have the significant advantage of isolating the listener from external sounds (spectators, other installations, etc.). As our installation is transportable, it can be installed in halls where the acoustic cannot be controlled and thus phenomenon of reverberation could appear and perturb the listener's perception. That is why we decide to use headphones as auditory display and a simple implementation of the *HRTF* associated to a spherical head model [26].

Interaction Devices

One way to interact with the virtual environment is to capture user's position and head's orientation. Keeping in mind the cost constraint, two devices have been created.

The first device is a multi-zonal sensitive carpet used to capture user's position. User's weight enables the contact of two aluminum sheets isolated by perforated foam (Fig. 2, right). The floor area of the carpet is 9m² and contains 64 square zones meaning that the precision of the device is 14cm². Data transmission between the carpet and the server is provided by a MIDI interface. The multi-zonal sensitive carpet transmits the approximate user's feet position and some actions can be detected like user's jump or when user is hitting the ground with one of his feet. User's average posture can be deduced when these latter information are coupled with the compass data.

The second device is indeed an electronic compass that gives user's head orientation. The lightness of the sensor allows placing it on a cap worn by user so its impact on the feeling of presence is small. The device is depicted on Fig. 2.

A *Nintendo Wiimote* is used to let the user directly interacts with the content of the virtual environment. The benefits are its low price and the capacities of that device to increase immersion. Moreover, some actions can be triggered according to user's position captured by the sensitive carpet.

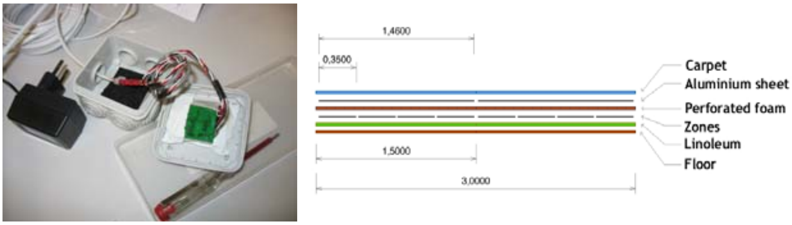


Fig. 2. The electronic compass (left) and a cross section of the sensitive carpet (right)

Software Architecture

Achieving real-time is a key point for an immersive installation and software architecture is the corner stone of this accomplishment. Such installation runs numerous processing threads sometimes very complex like sound processing, captor acquisitions and 3D rendering on multiple displays. Those processes are mostly independent from each other except for synchronization data. Such characteristics call for a distributed architecture over network. Assigning processes between several machines reduces per machine computational overhead and assures that every process gets the needed level of preemption. However this solution does have structural problems, so we stated that different processes need synchronization data. As an example, sound processing needs to be synchronized with user's position. Network accesses, despite being nearly as fast as hard disk accesses, are significantly slower than RAM and carrying data between each machine can be a huge bottleneck. The lack of hard real-time virtual reality platforms is another problem. Hard real-time ensures that two machines A and B receiving a signal from a machine C will act exactly the same at the same time. To address this problem A and B should share a clock signal to synchronize their reactions to an event.

Our implementation uses a client/server architecture pattern with a double logical network topology. Our server hosts several processes and handles the central control of the installation. An operator can diagnose problems and possibly stop or

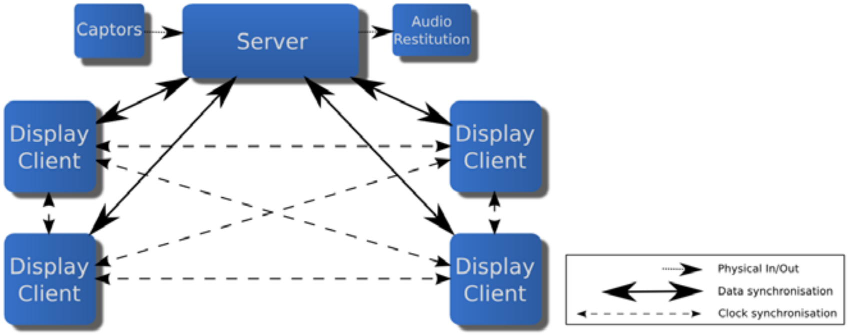


Fig. 3. Software architecture

restart the experiment. In order to avoid multiple synchronizations between captors, we centralized their acquisitions on the server.

The server hosts the sound processing and runs threads as well. Sound data should not be carried through the network since it may significantly decrease network performances. Then, only sound controls are carried over the network through the sound thread, which may process and play a locally stored sound as a consequence. Our server is a completely homemade *C++* program using the *Qt* library for inputs/outputs and the GUI. In our installation, the most resource intensive process is the multiple screen 3D rendering. Any 3D rendering process is a client to our server and is hosted on a separate machine. In our installation a display client is a *Ogre3D* application, but most of 3D engines such as *OpenSceneGraph* could be used as a display client. Each client is able to render on one screen and can be informed of captors of data. It can order the server to play or stop a particular sound. Synchronizing these events with the server is achieved using a star logical network. Signals from the server are broadcasted to each client while client orders directly reach the server. We stated that when receiving the same signal, clients can act slightly differently due to the lack of hard real-time. We resolve this problem with a full mesh logical network topology which connects every client. Each client broadcast a clock signal on the network to other clients. The clock is a signal count. If a client detects that another one has received more signals it will wait to receive the same signal before triggering more events. Else, if a client detects that the others didn't receive a signal, it will wait for them. This simple mechanism avoids de-synchronization between clients.

3.3 Importance of the Content

As described previously, one of the pillars concerns emotions. Most of virtual reality installations include computing units and others devices which lead to create visual and audio discomforts. More the user feels involved in the world presented to him more he forgets these imperfections. Generating emotions helps captivating user to focus his attention on the virtual environment rather than the material. To that purpose, we create a scenario taking advantage of the emotion's pillar to involve users.

Several video games studies [27-28] have shown the importance of integrating emotions in content to increase presence feeling. For example, fear and anxiety are widely used for stimulating user's emotions in video games implementing dark atmosphere or dangerous place. Considering the scenario, it should be a help for the generation of emotions [29].

For these reasons, our VR installation has been developed considering using emotions within a scenario. We created a story for first person game taking place in a prison. The user is playing the role of a prisoner who is trying to escape with a fellow prisoner (NPC) who gives some advice. That scenario has several advantages. First it allows user to evolve in the installation as he is moving in the virtual environment (for example the floor area fit with the jail area). Secondly, the atmosphere of prison is associated with strong feelings of stress and fear. These two latter points contribute to induce the feeling of presence, by generating emotions that involve user in the virtual world, and by increasing immersion thanks to similarities between real and virtual spaces.

Bi-directional aspect of immersion is also increased by displaying user's shadow and his image on reflective surfaces like mirrors. This point is important for self-presence as user has to feel existence in the virtual world. For that purpose, we use a technique named *visual hulls* [30] which allows to reconstruct the geometry of an object according to a set of images captured from calibrated cameras. For each image, the silhouette of the object is computed and projected in the form of a cone centered on the related camera. The intersection of the cones produces an envelope approximating the body of the object. In the same way, we use visual hulls to create an avatar similar to the user which will be used to compute shadow or reflections in a mirror as in figure 4.

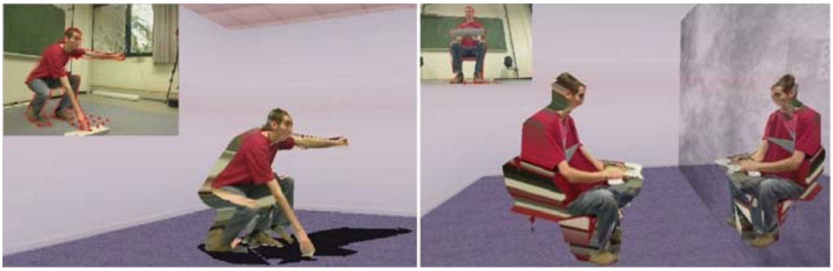


Fig. 4. Using a visual hulls of the user increase self presence

3.4 Financial Details

As previously explained, the main disadvantage of most VR installations is their cost. Especially, *CAVE*-like systems are expensive (from 23,000€ to 300,000€). These reasons led us to propose a low-cost installation made with consumer grade components. The manufacturing cost of our installation is around 7,500€ divided as follows:

- Display (4 screens and 4 video-projectors): 3,800€
- Computer hardware (5 PCs with Linux and network): 3,300€
- Small equipments (carpet, electronic compass and headphones): 420€

4 Evaluation

We have initiated an evaluation procedure of the efficiency of the installation concerning the feeling of presence and to validate several of our choices detailed in section 3. Our study had 30 participants who are all graduates students in computer sciences. Most of them regularly play video games. For this evaluation we use post-exposure questionnaire using a seven point Likert item. We firstly evaluate the perceived feeling of presence through its different angles (cf. section 2.1). Question 1 about spatial presence: *During the game, how much did you feel like you were really there in the virtual environment?* (1=there, 7=not there at all). Question 2 about self-presence: *During the game, how much did you feel you really be the prisoner?* (1=was, 7=was not me at all). Question 3 about environment presence: *During the game, how much did you feel that the virtual environment was a real place? (*1=real, 7=not real at all). Question 4 about social presence: *During the game, how much did you feel that other prisoners were existing person?* (1=existing, 7=not existing at all). Here are the results (average of the quote) for these 4 questions: Q1: 3.5, Q2: 3.9, Q3: 4.1, Q4: 5.1. The results were quite conformed to our expectation, but we estimate them too unfinished to draw conclusions.

We also evaluate the relevance of our interaction tools choices. Question 5: *Does the video game you just played provide high quality play control?* (1=high quality, 7=low quality). We had a quite surprising bad result for this question which may show that people expect more from a VR game than an interaction tool like *Wiimote*.

In the future experiment we will evaluate the impact of each pillar on presence, independently and in conjunction. We will test if self-presence decrease without the virtual shadow or the reflection in the mirror. We will also do a comparative study between playing the game in front of a PC or in our installation.

5 Conclusions

We have presented a low-cost (~7000€) and transportable version of a *CAVE*-like virtual reality installation. We also take care to keep enough quality to arouse feeling of presence. We have described the different parts of the installation which are screens, sound, interaction devices and software architecture. We have also argued the importance of the content of the application to maintain the feeling of presence. Finally, we have briefly presented the results of our studies based on a questionnaire which were aimed at checking if user feel to be present in the virtual world.

These latter evaluations have pointed out the lack of efficiency concerning the pillar of interaction. So we are thinking to more natural solutions of interactions by using non-intrusive devices. One of our goals is to propose a direct interaction thanks to the tracking of user's hands. We would like to establish some eye contact during the experiment which will signify that you are existing for someone else and then will increase your feeling of self-and co-presence. We also would like to add several fans in the installation to simulate air displacement to increase immersion. Finally, we want evaluate the impact of our four screens display on the feeling of presence by comparing it with a standard PC display.

References

1. Cruz-Nera, C., Sandin, D.J., Fanti, T.A., Kenyon, R.V., Hart, J.: The CAVE: audio visual experience automatic virtual environment. Communications of the ACM 35(6), 64–72 (1992)
2. IJsselsteijn, W.A.: Elements of a multi-level theory of presence: Phenomenology, mental processing and neural correlates. In: Proc. Intern. Conference PRESENCE 2002, pp. 245–259 (2002)
3. Welch, R.B.: How can we determine whether the sense of presence affects task performance? Presence 8(5), 574–577 (1999)
4. Bormann, K.: Subjective performance. Virtual Reality 9(4), 226–233 (2006)
5. Lee, K.M.: Presence, explicated. Communication Theory 14, 27–50 (2004)
6. Minsky, M.: Tele-presence. Omni, 45–51 (1980)
7. Stein, B., Meredith, M.: The merging of the senses, Cambridge, MA (1993)
8. Lombard, M., Ditton, T.: At the heart of it all: The concept of presence. Journal of Computer-Mediated Communication 3(2) (1997)
9. Ijsselstein, W.A., de Kort, Y.A., Haans, A.: Is this my hand I see before me? The rubber hand illusion: In reality, virtual reality and mixed reality. Virtual Environment 15(4), 455–464 (2006)
10. Biocca, F., Inque, Y., Polinsky, H., Lee, A., Tang, A.: Visual cues and virtual touch: role of visual stimuli and inter-sensory integration in cross-modal haptic illusions and the sense of presence. In: Proc. Intern. Conference PRESENCE 2002 (2002)
11. Väljamäe, A., Larsson, P., Västfjäll, D., Kleiner, M.: Travelling without moving: Auditory scene cues for translational self-motion. In: Proc. Intern. Conference on Auditory Display (2005)
12. Wang, Y., Otitoju, K., Lu, T., Kim, S., Bowman, D.: Evaluating the effect of real world distraction on user performance in virtual environments. In: Proc. ACM Symposium on Virtual Reality Software and Technology, pp. 19–26. ACM Press, New York (2006)
13. Banos, R.M., Botella, C., Liano, V., Guerrero, B., Rey, B., Alcaniz, M.: Sense of presence in emotional virtual environments. In: The 7th Annual International Workshop on Presence, Valencia (2004)
14. Riva, G., Mantovani, F., Capideville, C.S., Preziosa, A., Morganti, F., Villani, D., Gagioli, A., Botella, C., Alcaniz, M.: Affective interactions using virtual reality: The link between presence and emotions. Cyberpsychology and Behavior 10, 45–56 (2007)
15. Sutherland, I.E.: A head-mounted three dimensional display. In: Seminal graphics: Pioneering efforts that shaped the field, pp. 295–302. ACM, New York (1968)
16. DeFanti, T.A., Dawe, G., Sandin, D.J., Schulze, J.P., Otto, P., Girado, J., Kuester, F., Smarr, L., Rao, R.: The StarCAVE, a third-generation CAVE and virtual reality OptIPortal. Future Generation Comp. Sys. 25(2), 169–178 (2009)

17. Spagno, C., Kunz, A.: Construction of a three-sided immersive tele-collaboration system. In: Proc. IEEE Conference on Virtual Reality, pp. 22–26 (2003)
18. Hollerer, T., Kuchera-Morin, J., Amatriain, X.: The allosphere: A large-scale immersive surround-view instrument. In: Proc. 2007 Emerging Display Technologies Workshop, San Diego, CA (2007)
19. Peterka, T., Sandin, D.J., Ge, J., Girado, J., Kooima, R., Leigh, J., Johnson, A., Thiebaux, M., DeFanti, T.A.: Personal varrier: Autostereoscopic virtual reality display for distributed scientific visualization. Future Gen. Comp. Sys. 22(8), 976–983 (2006)
20. Fairen, M., Brunet, P., Techmann, T.: MiniVR: A portable virtual reality system. Computers & Graphics 28(2), 289–296 (2004)
21. Shibano, N., Hareesh, P., Hoshino, H., Kawamura, R., Yamamoto, A., Kashiwagi, M., Sawada, K.: Cyberdome: PC-clustered semi spherical immersive projection display. In: Proc. Intern. Conference on Artificial Reality and Tele-existence, pp. 1–7 (2003)
22. Ohno, R., Ryu, J.: Development of a portable virtual reality system for disaster education. In: 8th EAEA Conference, Moscow (2007)
23. Grimes, H., McMenemy, K.R., Ferguson, R.S.: A transportable and easily configurable multi-projector display system for distributed virtual reality applications. In: Proc. SPIE, pp. 68040G-68040G-9 (2008)
24. Cliburn, D.C., Stormer, K.: The HIVE: Hanover immersive virtual environment. Journal of Computing Sciences in Colleges 20(4), 6–12 (2005)
25. Bouvier, P., Chaudeyrac, P., Loyet, R., Piranda, B., de Sorbier, F.: Immersive visual and audio world in 3D. In: Proc. 9th International Conference on Computer Games, pp. 159–165 (2006)
26. Cheng, C.I., Wakefield, G.H.: Introduction to head related transfer functions (HRTFs): Representation of HRTFS in time, frequency and space. Journal of the Audio Engineering Society 49(4), 231–249 (2001)
27. Grimshaw, M.: Sound and immersion in the first-person shooter. International Journal of Intelligent Games & Simulation 5, 1 (2008)
28. Schneider, E.F., Lang, A., Shin, M., Bradley, S.D.: Death with a story: How story impacts emotional, motivational, and physiological responses to first-person shooter video games. Human Communication Research 30, 361–375 (2004)
29. Aylett, R., Louchart, S.: Towards a narrative theory of virtual reality. Virtual Reality 7, 2–9 (2003)
30. Laurentini, A.: The visual hull concept for silhouette-based image understanding. IEEE Transactions on Pattern Analysis. Machine Intel. 16(2), 150–162 (1994)

High-Level Hierarchical Semantic Processing Framework for Smart Sensor Networks*

D. Bruckner[1], C. Picus[2], R. Velik[1], W. Herzner[2], and G. Zucker[1]

[1] Institute of Computer Technology, Vienna University of Technology, Vienna, Austria
{bruckner,velik,zucker}@ict.tuwien.ac.at
[2] Department Safety & Security, Austrian Research Centers, Vienna, Austria
{cristina.picus,wolfgang.herzner}@arcs.ac.at

Abstract. Sensor networks play an increasing role in domains like security surveillance or environmental monitoring. One challenge in such systems is proper adaptation to and interpretation of events in the sensing area of an individual node; another challenge is the integration of observations from individual nodes into a common semantic representation of the environment. We describe a novel hierarchical architecture framework called SENSE (smart embedded network of sensing entities)[1] to meet these challenges. Combining multi-modal sensor information from audio and video modalities to gain relevant information about the sensed environment, each node recognizes a set of predefined behaviors and learns about usual behavior. Neighbor nodes exchange such information to confirm unusual observations and establish a global view. Deviations from “normality” are reported in a way understandable for human operators without special training.

1 Introduction

Today, surveillance has become a relevant means for protecting public and industrial areas against terrorism and crime. For both keeping privacy of irreproachable citizens as well as enabling automated detection of potential threats, computer-based systems are needed which support human operators in recognizing unusual situations. For that purpose the SENSE project implements a hierarchical architecture of semantic processing layers in a network of SENSE nodes [1], [4], where the goal of these layers is to learn the "normality" in the environment of a SENSE network, in order to detect unusual behavior or situations and to inform the human operator in such cases [5]. SENSE consists of a network of communicating sensor nodes, each equipped with a camera and a microphone array. These sensor modalities observe their environment and deliver streams of mono-modal events to a reasoning unit, which derives fused high-level observations from this information, which again is exchanged with neighbor nodes in order to establish a global view about the commonly observed environment. Detected potential threats are finally reported to the person(s) in charge. The first application area of SENSE is an

* This work is partially funded by the European Commission under contract No. 033279 within the Sixth Framework Program for Research and Technological Development.

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 347 – 358.
springrlink.com

airport; therefore, the special interests of the airport security staff are considered as typical user requirements for the taken approach.

Though most of the methods used in the particular layers are widely used in e.g. observation systems and many other applications, to our knowledge no other system uses a combination of them in order to let the messages of the system really appear smart and meaningful to the user.

This article is structured as follows: Section 2 outlines the system architecture, while Section 3 describes the individual layers in more detail, and Section 4 contains a conclusion and an outlook.

2 Architecture Overview

SENSE adopts an 8-layer data processing architecture, in which the lower layers are responsible for a stable and comprehensive world representation to be evaluated in the higher layers (Fig. 1).

First, the visual low-level feature extraction (layer 0) processes frame by frame from the camera in 2D camera coordinates and extracts predefined visual objects like “*person*”, “*group*”, or “*luggage*”. At the same time, the audio low-level extraction scans the acoustic signals from a linear 8-microphone array for trained sound patterns of predefined categories like "*scream*" or "*breaking glass*". Due to limited processing capabilities, this layer can deliver significantly unstable data. In case of unfortunate conditions for the camera (many persons, they exchange positions in the crowd, bad light conditions, etc), detected symbols can change their label from person to object to person-group and back for the same physical object within consecutive frames. The size of detected symbols can change from small elements like bags to large groups of persons covering tens of square meters and including previously detected single persons and other objects. Consequently, the challenge for higher levels is to filter out significant information from very noisy data.

The modality events of layer 0 (low-level symbols in Fig. 1) are checked at layer 1 for plausibility – e.g. the audio symbols with respect to position and intensity, the video symbols with respect to position and size. In the second layer, symbols which pass the first spatial check are subdued to a further check, regarding their temporal behavior. The output of this layer is a more stable and comprehensive world representation including unimodal symbols. Layer 3 is responsible for sensor fusion: the unimodal symbols are fused here to form multi-modal symbols.

Layer 4 is the parameter inference machine in which probabilistic models for symbol parameters and events are optimized. The results of this layer are models of high-level symbols (HLSs) and features that describe their behavior. In layer 5, the system learns about trajectories of symbols. Typical paths through the view of the sensor node are stored. Layer 6 manages the communication to other nodes and establishes a global world view. The trajectories are also used to correlate observations between neighboring nodes. In layer 7, the recognition of unusual behavior and events takes place using two approaches: the first one compares current observations with learned models by calculating probabilities of occurrence of the

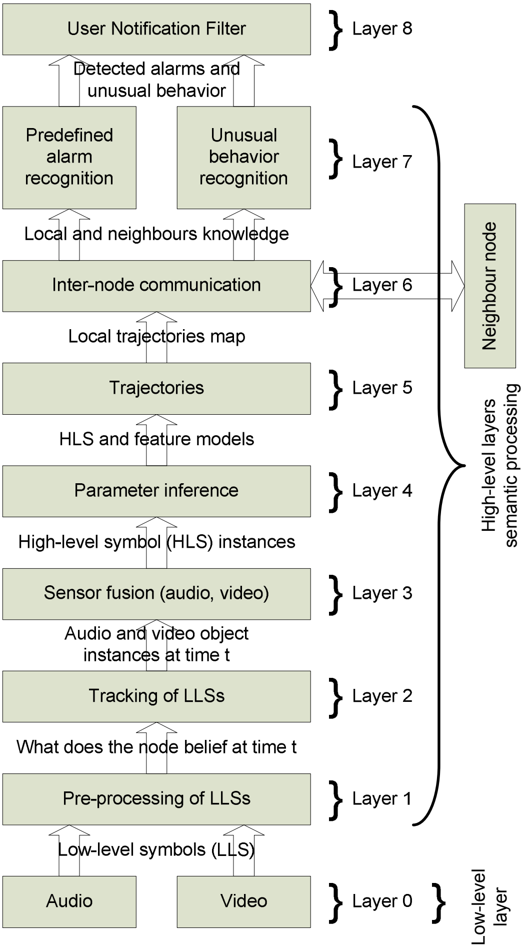

User Notification Filter
Layer 8
Detected alarms and unusual behavior
Predefined alarm recognition
Unusual behavior recognition
Layer 7
Local and neighbours knowledge
Inter-node communication
Layer 6
Neighbour node
Local trajectories map
Trajectories
Layer 5
HLS and feature models
Parameter inference
Layer 4
High-level symbol (HLS) instances
Sensor fusion (audio, video)
Layer 3
Audio and video object instances at time t
Tracking of LLSs
Layer 2
What does the node belief at time t
Pre-processing of LLSs
Layer 1
High-level layers semantic processing
Low-level symbols (LLS)
Audio
Video
Layer 0
Low-level layer

Fig. 1. High semantic processing software architecture

observations with respect to their position, velocity, and direction. It also calculates probabilities of the duration of stay of symbols in areas, probabilities of the movement along trajectories, including trajectories across nodes. Observations with probabilities below defined thresholds raise "unusual behavior" alarms. The second part of this layer is concerned with the recognition of predefined scenarios and the creation of alarms in case predefined “threat” conditions are met. Examples for predefined scenarios are "unattended luggage" or "running person". Finally, layer 8 is responsible for the communication to the user. It generates alarm or status messages and filters them if particular conditions would be announced too often or the same event is recognized by both methods in layer 7.

3 Description of Layers

3.1 Low-Level Feature Extraction

Basic Description
Layer 0 is responsible for low-level feature extraction. It provides the high-level layers with unimodal data streams, which describe observations of the sensors of a particular SENSE node at the time *t* in the environment where the node is embedded. There are extracted audio and video low-level symbols (LLS) representing defined primitives: person, person flow, luggage, etc. for video data; steps, gun shooting, etc. for audio data.

Function Principle
A description of the functionality of this layer lies outside the scope of this contribution. An extensive coverage can be found in [22] and [23].

3.2 Preprocessing Including Plausibility Checks

Basic Description
The task of the visual feature extraction is to match a set of templates with the current frame. In order to find objects of various sizes, the templates are scaled to allow matching within a range of sizes in every possible location in the frame. To filter unrealistic LLSs from the data stream, we first learn the average size of LLSs in dependence of their type and position in the camera field. Second, a plausibility check is done based on bounding boxes. Bounding boxes of symbols are used to determine if a person or object is blended into a larger object. In such a case, the count of smaller and larger symbols decides, which kind of symbol is most likely. The selected type is then used for further processing. Additionally to the size of symbols, the system also learns the average speed and usual direction of movement. This information is furthermore used for symbol tracking.

Function Principle
To determine the parameters of symbols, Gaussian or mixture of Gaussian models are used. One model is utilized for pixel clusters in order to translate the 640x480 camera pixels to 20x15 pixel clusters, each of them having a size of 32x32 pixels. Each pixel cluster contains models for each type of symbol. The models need different parameters depending on the number of persons, because people behave differently depending whether they are alone, in pairs, or in groups. As symbols can change their type from frame to frame, all symbols are kept – even symbols with low probability of occurrence – and marked accordingly. After having performed all plausibility checks, a voter decides which symbols are handed over to the next layer.

To construct models about the directions of movement (the angle), a split and merge algorithm for Gaussian mixture models is used [24]. More primitive variants of this algorithm are also used in order to process other parameters (size, speed), which possess only one component and therefore lack indices, posteriors, and priors.

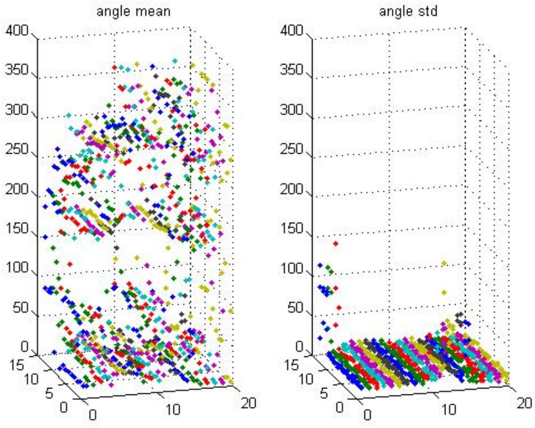


Fig. 2. Means and standard deviations of Gaussian mixture models of the direction of moving objects in the camera view

Results

Fig. 2 shows the result of the application of the algorithm to data from a realistic video. The video duration is 207.3 seconds and it consists of 2074 frames (10 frames per second). In sum, 924 different objects moved in the frames with different appearance times, ranging from 100 milliseconds to some minutes.

In the figure, three main directions for each cluster emerge. Most of the angles are grouped around the following directions: between 0° to 100°, around 200°, and between 250° to 350°. The σ value for each section is smaller than 20° and only at the left top and bottom right location there can be found bigger σ values. The reason for their occurrence is that in these locations, objects occur very seldom and the few sampled angles are not enough to construct a reliable statistics. Because of this, the σ values stay bigger (as they are initialized) before the first split and merge can be applied. During evaluation, these values need to be omitted.

3.3 Tracking

Basic Description

The tracking layer uses particle filter techniques to track the preprocessed symbols. The aim is to record the trajectories of targets over time and to get a correct, unique identification of each target.

In the area of particle filters, “objects” are tracked, not “symbols”. Therefore, this term is used in the following. Traditionally, multiple objects are tracked by multiple single-object-tracking filters. While the usage of independent filters is computationally tractable, the result is prone to frequent failures. Each particle filter samples in a small space. The resulting “joint” filter’s complexity is linear in the number of targets. The problem is that in cases where targets interact, as it occurs in many of our scenarios, single particle filters are susceptible to failures when interactions occur. In a typical failure mode, several trackers start tracking the single target with the highest likelihood score.

Function principle

To track the preprocessed symbols, a particle filter specifically designed for tracking interacting objects [2] is used. To address tracker failures resulting from interactions, a motion model based on Markov random fields (MRFs) [11] is introduced.

When targets pass close to one another or merge as persons do in a crowd, tracking multiple identical targets becomes especially challenging. Recently, an approach was developed that relies on the use of a motion model, able to adequately describe target behavior throughout an interaction event [2]. This approach contains a motion model that reflects the additional complexity of the target behavior.

The reason for using this approach is that the number of LLS in the observation model can change in each time frame. E.g., if several persons walk through a corridor, the visual feature extraction algorithms might detect a number of single persons in one image and just a group of persons in the consecutive one. Under bad conditions, the detection for the same physical object can change often within short time periods. Details about how to use the method described in [2] for our purposes, can be found in [19]. Its advantage is that it minimizes the computational effort in comparison to joint particle filters for tracking of multiple objects at once and it also minimizes the fault detection rate compared to a set of single-object-tracking particle filters for individually tracking each object simultaneously.

Results

The effectiveness of the method is shown in Fig. 3. Two of the big advantages of the tracker are that it omits spurious visual errors only appearing very shortly and

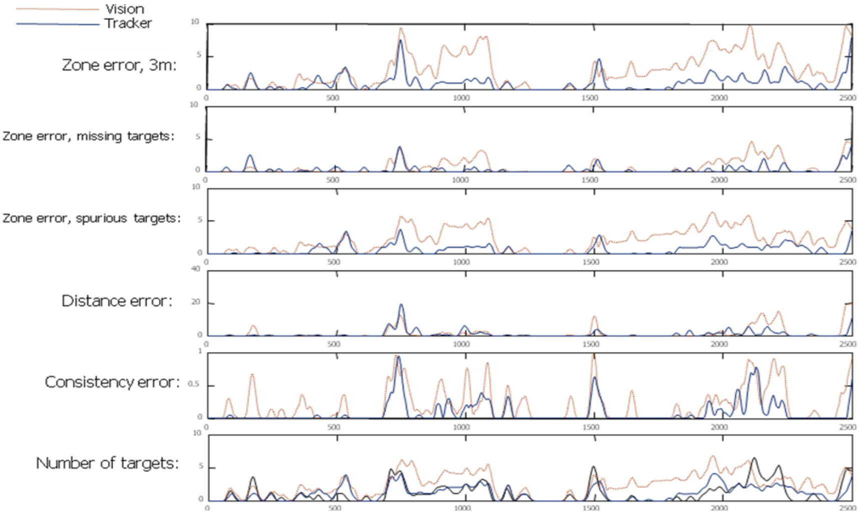


Fig. 3. Result of the tracking layer. Shown are zone errors (if a tracker target is within a 3m radius circle in world coordinates around an original person), distance error (accumulated distance between original persons and closest targets), and consistency error (if a target is associated to the same original person over time)

that it achieves much better consistency (as can be seen in Fig. 3, the “consistency error” is much smaller), which is necessary for constructing a model of trajectories of the tracked objects. Both, vision alone and tracker are compared to a “ground truth”, which was created manually by pointing at the position of each person in each frame. In the bottom sub figure, the real number of targets is additionally shown as darkest line.

3.4 Sensor Fusion

Basic Description

The inputs of this layer are the unimodal symbols with smoothed trajectories. Its task is the fusion of audio and video symbols. For this purpose, factor analysis [10, 12] is used to determine the correlation between audio and video LLSs. The output of the layer is a symbolic representation of the real world as a collection of multi-modal symbols [13].

Function Principle

Fusion of the audio and the video data is done using correlations between the provided data streams. Based on time correlation of low-level symbols (LLS), features that are used for this purpose are: direction of arrival, position, loudness, power spectrum, and size of the video symbol in pixels.

3.5 Parameter Inference

Basic Description

The task of the parameter inference layer is to set up a local semantic description of the events sensed by each node from the set of all the incoming lower level data. Besides the data representation, this layer also establishes neighborhood connections between data representatives automatically within the node’s view. For learning the data representation, a vector quantization approach is used based on the Growing Neural Gas (GNG) algorithm [17]. In particular, it shall be referred to [18] and [16], in which the GNG method is used for learning high frequency paths in visual surveillance scenarios.

Function Principle

Getting symbols from the sensor fusion layer, the task of this layer is to infer the parameters of the underlying probabilistic model (Mixture of Gaussians [3]). To generate a spatial or – depending on requirements – spatio-temporal model of the events, unsupervised learning is used. The GNG algorithm is applied as a data-driven vector quantization approach that iteratively generates a codebook of prototypes in order to represent the data distribution, following the criterion of minimizing the overall network quantization error. By data representation in terms of prototypes, a full covering of the data is provided. This means that even isolated small clusters of data are included into the data representation. Besides this, the number of prototypes is arbitrary, while in other methods, e.g. the K-means clustering, the number of clusters is a parameter to be defined beforehand.

Results
Ground-truth data of the CAVIAR project from the PETS 2004 workshop [25] is used to generate high-level symbols (HLS). The dataset includes 49 sequences, which correspond to two different camera views for a total of approximately 100 trajectories. Persons walk along a corridor, which contains entrances to several shops. The available ground-truth provides the bounding box and the IDs of the objects for each frame. The reference point used for the trajectories of the training dataset is the head-top. Fig. 4 shows the case that two-dimensional HLSs are super-imposed to the two camera views.

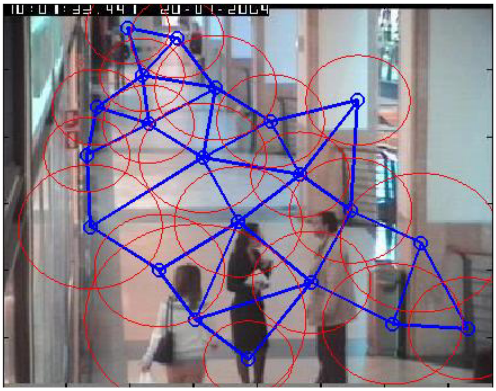


Fig. 4. Front view of the corridor from the CAVIAR dataset. HLSs in 2D are represented by the ellipses. In blue, a representation of the graph can be seen, including prototype centers and edges

3.6 Trajectories

Basic Description
Trajectories in a node are derived by use of a learned transition matrix, which consists of transitions between HLSs. For this purpose each HLS keeps a list of all local trajectories to which the symbol belongs. At the time *t*, when an observation is associated with that symbol, it is also associated to all trajectories to which the symbol belongs. The time sequence of symbol activation is used to compute the most probable trajectory for the univocal association of the observation to the trajectory. Besides this, trajectories will also be active in neighboring nodes with correlated symbols. Global trajectories are learned over multiple nodes, by association of local trajectories sharing correlated symbols between pair of nodes. The activation of a less probable local or global path triggers an alarm.

Function Principle
Each node keeps a lookup table of all possible local and global trajectories. The matrix is built using the information deriving from the local transition matrix and the correlation matrix learned between symbols in different nodes. The knowledge about the pair-wise association between neighboring nodes is exchanged among nodes in order to set up the global view.

3.7 Inter-Node Communication

Basic Description
Inter-node communication is based on the Loopy-Belief Propagation (LBP) algorithm [6, 7]. Its primary task is to detect neighborhood relations between nodes by estimating correlations between symbols in different nodes, using simultaneous activations. This knowledge is used for both improving local views through feedback from neighbors and establishing the global view (e.g. trajectories).

Function Principle
A description of the functionality of this layer is outside the scope of this paper. Details can be found in [20].

3.8 Alarm Generator

Basic Description
The layer responsible for alarm generation detects predefined alarms and unusual behavior. The major difference between predefined alarms and unusual behavior is the following: a predefined alarm can be associated with a predefined, human-readable text, like "scream noise" or "person dropped luggage". Unusual behavior in any situation – not only a predefined one – is detected as “unusual that deviates from normality”.

Function Principle 1: (Predefined) Scenario Recognition
The input for this method is the symbolic representation on the level of the modality related symbols. A rule-base is used to combine these symbols in a hierarchical way in order to get symbols with higher (more abstract) semantic meaning from symbols with lower semantic level.

Predefined alarms like “screaming person” merely rely on information available from low-level symbols of the audio modality. In contrast, alarms like “unattended luggage” require a symbolic processing of different information sources (see also [14]).

Function Principle 2: Unusual Behavior Recognition
The recognition of unusual behavior is operated on the base of the learned models of normal behavior. Each event is proven for normality: if it is classified outside the learned models, a deviation from normality is reported to the next layer. The criteria for normality are two: 1) activation of learned HLSs; 2) activation of a

learned sequence of HLSs. The activation corresponds to a probability of the event being classified in the given model higher than a threshold.

3.9 User Notification Filter

Basic Description
The task of this layer is the connection of the high-level sensor processing and the user interface. It delivers alarms to the user interface and can be asked about the status of one or several nodes. It filters identical alarms occurring e.g. when the same lurking person is reported several times from the predefined scenario recognition. When a reported unusual behavior can be matched with a predefined alarm, only the latter will be delivered. Furthermore, the user can choose filtering rules to omit or prolong alarms via the GUI.

Function Principle
A filtering mechanism is applied to avoid the sending of the same alarm several times or the sending a predefined alarm and an unusual behavior for the same object. Furthermore, an additional rule-base with user preferences is also considered (see e.g. [21]).

4 Conclusions

This article presents a hierarchical processing architecture for smart sensor networks. The innovative aspect lies in the step-by-step processing, during which low-level information becomes more and more meaningful to a human operator in charge. Expected results are described for the deployment in an airport environment, whereby all layers of the hierarchical processing framework are described in order to understand the idea behind the architecture. Particular results from layers are depicted in order to give the reader better impressions of the overall performance of the approach.

References

1. `http://www.sense-ist.org` (SENSE project website) (accessed April 1, 2009)
2. Khan, Z., Balch, T., Dellaert, F.: An MCMC-based particle filter for tracking multiple interacting targets. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(12), 1960–1972 (2006)
3. Bishop, C.M.: Neural networks for pattern recognition. Oxford University Press Inc., New York (1995)
4. Zucker (né Pratl), G., Frangu, L.: Smart nodes for semantic analysis of visual and aural data. In: Proc. IEEE Conference INDIN, pp. 1001–1006 (2007)
5. Sallans, B., Bruckner, D., Russ, G.: Statistical detection of alarm conditions in building automation systems. In: Proc. IEEE Conference INDIN, pp. 6–9 (2006)

6. Crick, C., Pfeffer, A.: Loopy belief propagation as a basis for communication in sensor networks. In: Proc. UAI Conference on uncertainty in Artificial Intelligence, pp. 159–166 (2003)
7. Yedidia, J.S., Freeman, W.T., Weiss, Y.: Characterization of belief propagation and its generalizations (2001), http://www.merl.com/publications/TR2001-015/ (accessed April 1, 2009)
8. Lee, D.S.: On-line adaptive Gaussian mixture learning for video applications. In: Proc. 8th Conference on Computer Vision, Prague, Czech Republic (2004), http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.29.2197 (accessed March 29, 2009)
9. Vlassis, N., Likas, A.: A greedy EM algorithm for Gaussian mixture learning. Neural Proc. Letters 15(1) (2002), http://www.cs.uoi.gr/~arly/papers/gem-npl.ps.gz (accessed March 30, 2009)
10. Ghahramani, Z., Beal, M.J.: Variation inference for Bayesian mixtures of factor analyzers. In: Advances in neural information processing systems. MIT Press, Cambridge (2000)
11. Ueda, N., Nakano, R., Ghahramani, Z., Hinton, G.E.: SMEM algorithm for mixture models. Neural Computation Archive 12 (2002)
12. Ghahramani, Z., Hinton, G.E.: The EM algorithm for mixture of factor analyzers. Technical Report CRG-TR-96-1, University of Toronto (1996)
13. Lombardi, P.: Study on data fusion techniques for visual modules. Technical Report, University of Pavia (2002)
14. Burgstaller, W.: Interpretation of situations in buildings. Dissertation thesis, Vienna University of Technology (2007)
15. Kindermann, R., Snell, J.L.: Markov random fields and their applications. AMS Books Online, ISBN: 0-8218-3381-2 (1980), http://www.ams.org/online_bks/conm1 (accessed March 24, 2009)
16. Bauer, D., Brändle, N., Seer, S., Pflugfelder, R.: Finding highly frequented paths in video sequences. In: Proc. 18th Intern. Conference on Pattern Recognition (ICPR 2006), Hong Kong, China, pp. 387–391 (2006)
17. Fritzke, B.: A growing neural gas network leans topologies. In: Advances in Neural Information Processing Systems, vol. 7, pp. 625–632 (1995)
18. Pflugfelder, R.P.: Visual traffic surveillance using real-time tracking. TR PRIP, Vienna University of Technology (2002), http://www.icg.tu-graz.ac.at/pub/pubobjects/pflugfelder2002 (accessed March 25, 2009)
19. Bruckner, D., Kasbi, J., Velik, R., Herzner, W.: High-level hierarchical semantic processing framework for smart sensor networks. In: Proc. IEEE Conference on Human System Interaction, Cracow, Poland, pp. 668–673 (2008)
20. Picus, C., Cambrini, L., Herzner, W.: Boltzmann machine learning approach for distributed sensor networks using loopy belief propagation inference. In: Proc. 7th Intern. Conference on Machine Learning and Applications, San Diego, CA, pp. 344–349 (2008)
21. Wide, P.: The electronic head: a virtual quality instrument. IEEE Transactions on Industrial Electronics (48), 766–769 (2001)
22. Simo, J., Benet, G., Andreu, G.: Embedded video processing for distributed intelligent sensor networks. In: Proc. 3rd Intern. Conference From Scientific Computing to Computational Engineering, Athens, Greece (2008)

23. Tsahalis, D., Nokas, G., Tsokas, K., Photeinos, D.: The use of decision tree classifiers for the detection of sound objects using microphone array filtered data. Part I and II. In: Proc. 3rd intern. conference From Scientific Computing to computational Engineering, Athens, Greece (2008)
24. Yin, G.Q., Bruckner, D., Zucker, G.: Statistical modeling of video object's behavior for improved object tracking in visual surveillance. In: Proc. IEEE ICIT Conference, Churchill, Victoria, Australia (2009)
25. Anon. CAVIAR dataset, http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/ (accessed March 26, 2009)

A Human-System Interaction Framework and Algorithm for UbiComp-Based Smart Home

C.L. Wu[1] and L.C. Fu[2]

[1] Department of Computer Science & Information Engineering,
National Taiwan University, Taipei, Taiwan
f89922042@ntu.edu.tw

[2] Department of Computer Science & Information Engineering and
Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan
lichen@ntu.edu.tw

Abstract. Current smart home is a ubiquitous computing environment consisting of multiple agent-based autonomous spaces, and it brings both advantage and challenge that a service interacting with users can be performed with multiple choices of configuration in space, hardware, software, and quality. In addition, a smart home should also satisfy the “home” feeling when interacting with its inhabitants. In this work, we analyze the relationship between services, spaces, and users, and then propose a framework as well as a corresponding algorithm to model their interaction, thus for a smart home to behave both smart and as a home when interacting with its inhabitants.

1 Introduction

Current smart homes have transited from a centralized home automation system (HAS) integrating many hardware (HW) devices with software (SW) applications to a distributed system composed of cooperative agents integrating heterogeneous ubiquitous-computing (UbiComp) HW/SW to provide services [1], which can be seen as a UbiComp-based environment consisting of multiple autonomous spaces. Here we define that an autonomous space is a service area handled by agents, and each space equips HW/SW to provide various services. The greatest advantage of UbiComp-based smart home is that a service can be performed with multiple choices of configuration of at which space by what HW/SW in what level of quality, but this is also a big challenge to make the best configuration.

However, the human requirement for home, generally involved with “comfort”, “convenience”, and “security”, can not be satisfied simply by advanced computer technology mentioned above. Therefore, for a smart home to be more than technology-based smart, it is very important to fulfill “comfort + convenience + security” when choosing the configuration of a service interacting with inhabitants. In detail, we define “comfort” to be the quality of services (QoS) and the way to provide services, “convenience” to be the relationship between the user and the space where services are provided, and “security” to be the privacy issue. In this work,

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 359–370.
springrlink.com

we will study the above issues, and propose a framework with a corresponding algorithm as the solution.

The rest of this paper is organized as follows. Section 2 overviews related works, and section 3 describes the overall system. Section 4 and section 5 addresses the proposed framework and algorithm, and section 6 analyzes and discusses the design issues. Results of experiments and evaluations are shown in section 7, and conclusions are made in section 8.

2 Related Works

Reference [2] develops a system which can adapt its QoS according to the environment changes and user preference. However, it considers neither the issue of multiple spaces nor the issue of multiple users, and the privilege issues that a user may have different privilege at different spaces and different users may have different privileges in the same space is also not considered. The privacy issue and interaction requirement of services are also ignored. In addition, this system is user-initiative, thus lacking of the situation where interaction is system-initiative, which is very important for UbiComp-based environment [3].

Reference [4] proposes a contextual notification model for system to adapt its interaction according to "Message Priority," "Usage Level," and "Conversation". It considers the relation between user and system, however, it is for single user only, and issues of privacy, privilege, and multiple spaces, are also ignored.

To further consider the relation between services and user, we study from [5] and [6]. In [5], the authors proposed a framework to categorize the notification systems according to their design objectives in "interruption," "reaction," and "comprehension". In [6], the authors extend the model from [7] to categorize the interaction way between user and system as "foreground" and "background" according to whether or not requiring user's attention. In this work, we combine the framework of [5] and the concept of [6] to represent the interaction requirement of services, as well as the relation between services, spaces, and users.

As for the privacy issue, we study from [8] which addressed that a system should provide users three modalities of control – solitude, confidentiality, and autonomy – to complete the construction of an integrated privacy. We apply this viewpoint to consider the interaction privacy in three corresponding perspectives – physical, information, and freedom. Details will be shown in later section.

There are still several related works. Reference [9] proposes a framework which envisions a smart home as a user centric multimedia system. Reference [10] and [11] propose similar concept, which is that the next generation human-computer interaction should utilize and enhance existing appliances to embed the interaction information, so that inhabitants can naturally access them. Reference [12] proposes a mobile device based personal home server to achieve spontaneous interaction, personalization, privacy protection, and interoperability. However, all these works do not touch our focus, which is that fully utilizing the resources of a smart home to fulfill the "comfort + convenience + security" requirement when interacting, based on the relationship between human and spaces in a smart home.

3 System Overview

The interaction flowchart of a smart home system can be simply described as Fig.1(a). Via the backbone platform, all the components are connected and can communicate with one another. Smart home system gathers statuses of the environment and its inhabitants via intelligent sensing HW/SW, receives user input from smart human-computer interface (HCI), and then sends all information to inference mechanism to infer what services to be provided. According to the inference results, smart home system will perform the parts in dotted line, including interacting with inhabitants via interface, manipulating devices around users, and changing the environment status via the integrated control handled by HAS.

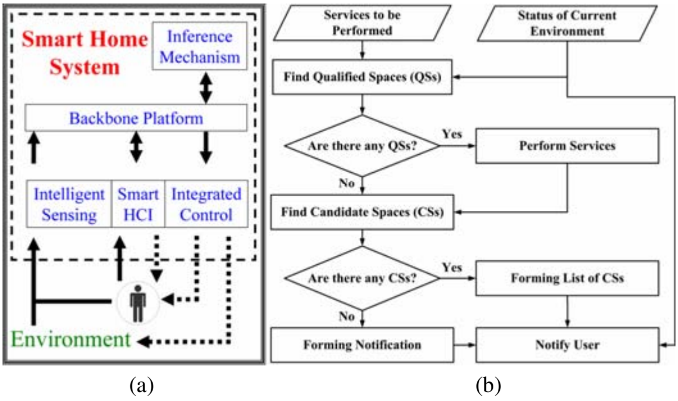


Fig. 1. (a) the interaction flowchart. (b) the algorithm flowchart to provide services

The focus of this work is how to utilize the UbiComp-based environment to fulfill "comfort + convenience + security" when performing the dotted line parts, and the flowchart is shown in Fig.1(b) and detailed as follows. When a smart home is going to perform a service for a user, it will compare the requirements of the service with the status of current environment, to find out if there are some qualified spaces (QSs), whose status and resources are both capable of performing this service. If such spaces exist, the smart home will perform the service at one of the QS according to the requirements of the service; otherwise, the smart home will continue to find out if there are some candidate spaces (CSs), whose resources are capable but status is not ready. If such spaces exist, the smart home will form a CS list with a notification to indicate user that a service is waiting to be performed; otherwise, the smart home will form a notification to indicate user that a service can not be provided at current environment. In either situation, the smart home will notify the user. The idea behind is to fully utilize the spaces in the smart home so that services are provided at appropriate spaces according to the relation between user, service, and space. After the notification, if the user agrees this service to be performed, he/she can choose one of the CSs and change its status for the smart home to perform the service, or override the smart home by manually initiating the service at any space, but this is out of the scope of this work.

There are several points needing to be taken care of as follows. The first point will be dealt with in section 4, and the other points will be addressed in section 5.

1. Define the requirements of services and the status of environment.
2. Define the conditions for a space to be a QS or CS.
3. Determine at which QS by what HW/SW in what level of QoS.
4. Determine how to rank CSs to form the CS list.
5. Appropriately notify the user according to the status of environment.

4 Framework for Human-System Interaction

4.1 The Requirements of Services

According to [2], each service should specify its required resources and corresponding HW/SW. But we further propose that these requirements should be classified into several QoS levels. Requirements in the lowest level represent the least constraints to perform this service, whereas requirements in the highest level mean that the Quality of Service (QoS) will not be higher even richer resources and HW/SW are provided.

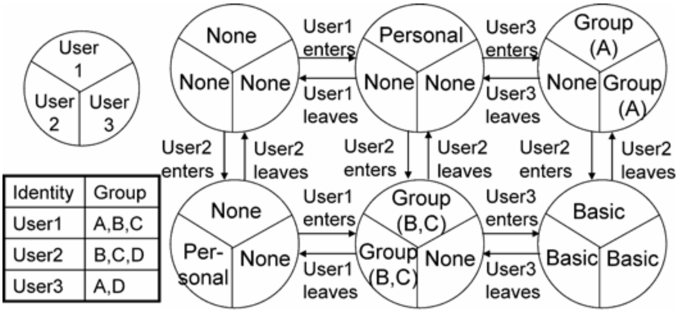
Learning from [6], service can be classified according to whether requiring user attention. However, inspired by [5], we propose that this attention-based classification way should be further improved as that each service should specify how much attention it requires, because while background services do not require user attention, foreground services may prefer full, some, or little user attention.

Inspired by [4] and [8], we propose that each service should specify its requirement for priority and privacy. The former can be set as “normal” or “high” and the latter can be set as “personal”, “group”, or “basic”. Priority is used to adjust the initial interaction scopes of services and notifications, whereas privacy is a constraint for selecting spaces to perform services. Although there may be multiple groups in a smart home, we assume that if a service requires “group” privacy, it only needs to specify one of them as its privacy requirement.

All these requirements mentioned above should be assigned by the service developers at the design phase, except that the requirement for priority and privacy may be dynamically determined by the service content.

4.2 The Status of Environment

First, we define that the privacy level of a space is determined as “personal”, “group”, “basic”, or “none”, by all involved users. And a space can specify multiple groups as its group privacy level according to involved users. A simple example is described as follows, and is shown in Fig.2. If there is only one user in the space, then its privacy level is personal for that user. If another user joins this space and these two users belong to some groups, then its privacy level goes down to “group” and includes all their common groups. Otherwise, if they do not belong



Identity	Group
User1	A,B,C
User2	B,C,D
User3	A,D

Fig. 2. An example for transition of privacy level in a space

to any same group, then its privacy level goes to basic. If the space is not involved with some user, then for him/her, its privacy level is none, the lowest one.

Then corresponding to the attention requirement of services, we first classify the spaces as two types, background space (B.Space) and foreground space (F.Space), according to their interaction level. The interaction level of a space is defined as whether there are foreground interaction between user and this space recently. To model that the user attention will go lower as time goes by or the user leaves the space, another space type, inactive foreground space (IF.Space), is proposed. And to represent the spaces which can be utilized but currently do not belong to the user, we propose another space type, free space (Free.Space). At last, to further deal with the privacy issue caused by multiuser, we define another space type, restricted space (X.Space), as that if a space is a F.Space for user$_1$ but not for user$_2$, then this space is an X.Space for user$_2$. Further, according to whether user$_2$ is in the X.Space or not, X.Space is classified as restricted background space (XB.Space) and restricted free space (XFree.Space). A simple example for the transition of spaces is described as follows, and is shown in Fig.3.

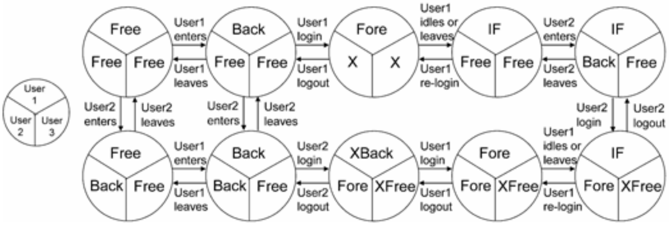


Fig. 3. An example for transition of interaction level in a space

According to [2], each space should specify its own resources and its equipped HW/SW. However, to further deal with the privilege issue, we propose that each space also has to specify how much resources and what HW/SW can be allocated for each user according to the user identity.

There are still three other statuses for each space but not related to service requirements. The first one is a unique name for each space. The other two are its service area, and its topology with other spaces. The former is used to determine whether a user enters a space, and the latter is used to calculate the distance between spaces thus later to rank the CS. When defining a new space, its service area is first specified and registered in the smart home system, and then its topology is represented by edges, each of which links to a space whose service area intersects or neighbors with. An example for topology of spaces is shown in Fig.4.

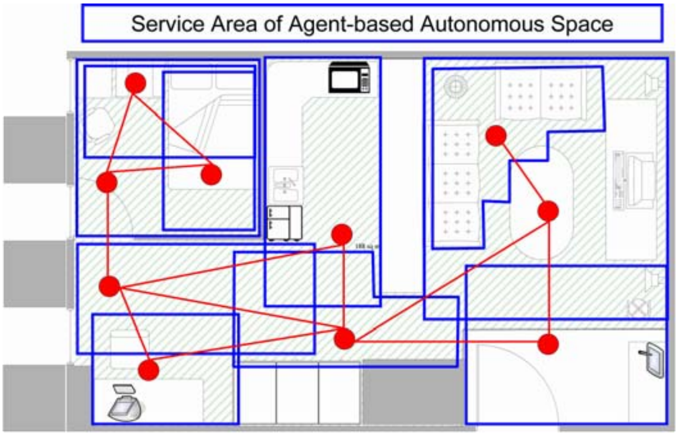
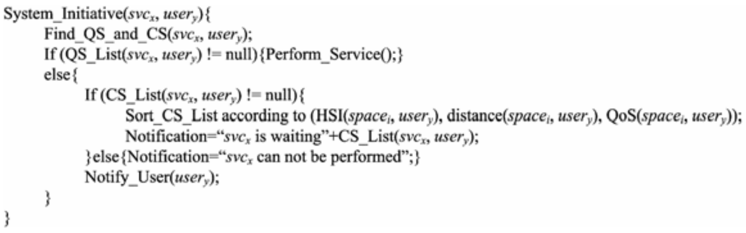


Fig. 4. An example of smart home which consists of multiple spaces with their topology

The first two statuses, privacy level and interaction level, are dynamically determined by the interaction between users and spaces, whereas the other statuses are defined and fixed at the beginning of space creation.

5 Algorithm for Human-System Interaction

The main algorithm shown in Fig.5 mainly formulates Fig.1(b), and several sub-algorithms in Fig.5 will be described later. If the smart home initiates a service, *svc_x*, for a user, *$user_y$*, then each space, *$space_i$*, will be determined whether it can be a QS or CS by comparing the requirements of *svc_x* with all its space statuses.



```
System_Initiative($svc_x$, $user_y$){
    Find_QS_and_CS($svc_x$, $user_y$);
    If (QS_List($svc_x$, $user_y$) != null){Perform_Service();}
    else{
        If (CS_List($svc_x$, $user_y$) != null){
            Sort_CS_List according to (HSI($space_i$, $user_y$), distance($space_i$, $user_y$), QoS($space_i$, $user_y$));
            Notification="$svc_x$ is waiting"+CS_List($svc_x$, $user_y$);
        }else{Notification="$svc_x$ can not be performed";}
        Notify_User($user_y$);
    }
}
```

Fig. 5. The algorithm for a smart home to initiate services

5.1 Find QS and CS from the Environment

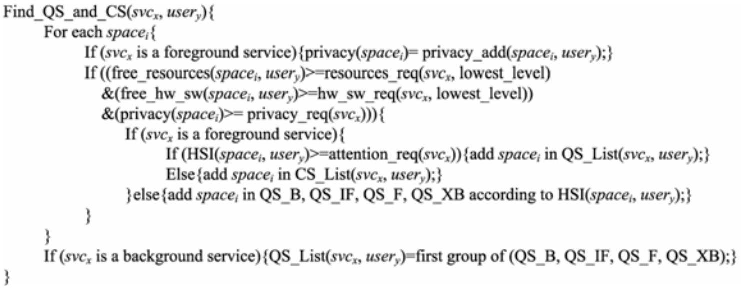
If $space_i$ is a QS for svc_x for $user_y$, it fulfills the following conditions:

1. Its remaining resources and HW/SW for $user_y$ are richer than the least resources constraints of svc_x.
2. Its privacy level for $user_y$ is not lower than the privacy required by svc_x.
3. Its interaction level for $user_y$ is not lower than the attention level specified by svc_x.

According to the last condition, QSs for foreground services can only be F.Space with high enough interaction level, whereas QSs for background services can be any type of spaces other than Free.Space and XFree.Space. But the QSs for background services will be first grouped and sorted by the order of B.Space, IF.Space, F.Space, XB.Space, and then all the QSs in the group listed first will be chosen as the true QSs.

The idea of CS is to deal with the situation that currently there is no QS for the system-initiative foreground service. To achieve the purpose of foreground services which is to draw user attention, all suitable spaces in the environment will be found out for this service to interact with the user. And since background services do not require user attention, there is no CS for them. CSs are those spaces which only do not fulfill the last condition, as well as those Free.Space or XFree.Space which fulfill the first two conditions if $user_y$ has entered them.

This part of algorithm is shown in Fig. 6.



```
Find_QS_and_CS($svc_x$, $user_y$){
    For each $space_i${
        If ($svc_x$ is a foreground service){privacy($space_i$)= privacy_add($space_i$, $user_y$);}
        If ((free_resources($space_i$, $user_y$)>=resources_req($svc_x$, lowest_level)
          &(free_hw_sw($space_i$, $user_y$)>=hw_sw_req($svc_x$, lowest_level))
          &(privacy($space_i$)>= privacy_req($svc_x$))){
            If ($svc_x$ is a foreground service){
                If (HSI($space_i$, $user_y$)>=attention_req($svc_x$)){add $space_i$ in QS_List($svc_x$, $user_y$);}
                Else{add $space_i$ in CS_List($svc_x$, $user_y$);}
            }else{add $space_i$ in QS_B, QS_IF, QS_F, QS_XB according to HSI($space_i$, $user_y$);}
        }
    }
    If ($svc_x$ is a background service){QS_List($svc_x$, $user_y$)=first group of (QS_B, QS_IF, QS_F, QS_XB);}
}
```

Fig. 6. The algorithm to find Qs and CS for a service

5.2 List of CS

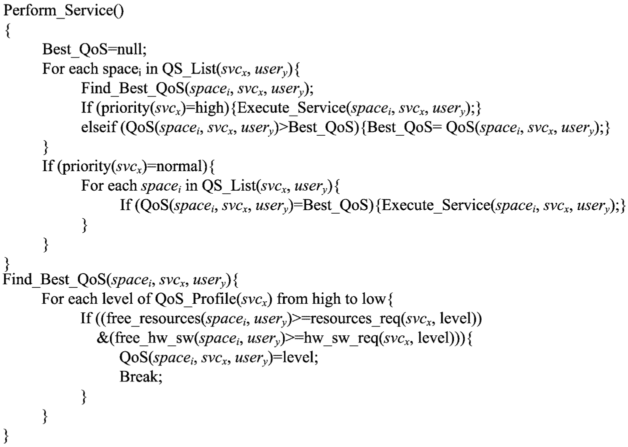
All the CSs will be grouped and sorted by the order of F.Space, IF.Space, B.Space, Free.Space, XB.Space, XFree.Space. After that, CSs in each group will be further sorted by its distance from $user_y$ according to its topology and by its best QoS level. For the user to identify each CS, the CS list includes the following information: name, type, distance, QoS level. This process is included in Fig. 5.

5.3 Configuration of Services

The algorithm in Fig. 7 decides at which QS by what HW/SW in what QoS level svc_x will be performed. The priority of svc_x is for adjusting the interaction scope:

- **Service requiring high priority:** All the QSs will initiate svc_x at the best QoS level they can afford with related HW/SW specified by svc_x.
- **Service requiring normal priority:** All the QSs will calculate the best QoS level they can afford, and all the QSs with the best QoS level are chosen to perform svc_x with specified HW/SW.

If there are multiple QSs initiating svc_x, user can choose one of them, and others will stop svc_x. If user does not make the decision, the situation will remain.

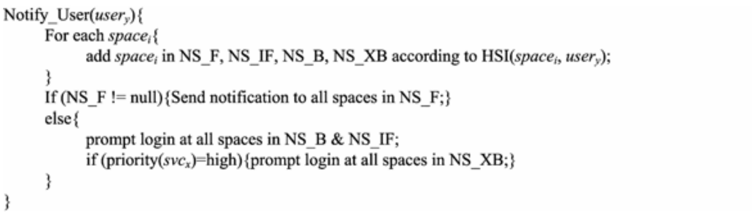


```
Perform_Service()
{
    Best_QoS=null;
    For each space$_i$ in QS_List($svc_x$, $user_y$){
        Find_Best_QoS($space_i$, $svc_x$, $user_y$);
        If (priority($svc_x$)=high){Execute_Service($space_i$, $svc_x$, $user_y$);}
        elseif (QoS($space_i$, $svc_x$, $user_y$)>Best_QoS){Best_QoS= QoS($space_i$, $svc_x$, $user_y$);}
    }
    If (priority($svc_x$)=normal){
        For each $space_i$ in QS_List($svc_x$, $user_y$){
            If (QoS($space_i$, $svc_x$, $user_y$)=Best_QoS){Execute_Service($space_i$, $svc_x$, $user_y$);}
        }
    }
}
Find_Best_QoS($space_i$, $svc_x$, $user_y$){
    For each level of QoS_Profile($svc_x$) from high to low{
        If ((free_resources($space_i$, $user_y$)>=resources_req($svc_x$, level))
          &(free_hw_sw($space_i$, $user_y$)>=hw_sw_req($svc_x$, level))){
            QoS($space_i$, $svc_x$, $user_y$)=level;
            Break;
        }
    }
}
```

Fig. 7. The algorithm to perform and configure a service

5.4 Notification

For background services, because there is no CS for them, it will only notify users "svc_x can not be performed currently". As for foreground services, besides the above notification content, it can also notify users "svc_x is waiting to be performed" attached with the CS list. The related algorithm is shown in Fig.8.



```
Notify_User($user_y$){
    For each $space_i${
        add $space_i$ in NS_F, NS_IF, NS_B, NS_XB according to HSI($space_i$, $user_y$);
    }
    If (NS_F != null){Send notification to all spaces in NS_F;}
    else{
        prompt login at all spaces in NS_B & NS_IF;
        if (priority($svc_x$)=high){prompt login at all spaces in NS_XB;}
    }
}
```

Fig. 8. The algorithm to notify user

Since the purpose of notification is to draw user attention about some events of svc_x, it can be seen as a kind of special foreground service. Therefore, the possible types of spaces to perform notification can only be F.Space, IF.Space, B.Space, XB.Space. Next, the priority of svc_x will decide the notification scope.

- **Service requiring high priority:** If there are some spaces in F.Space group, notification will be directly performed in all these spaces. Otherwise, for privacy, login will be prompted first in all the remaining spaces, and then the notification will only be performed in the space where after the user login.
- **Service requiring normal priority:** The process is the same as the one in high priority except that the notification will not be performed at XB.Space.

6 Analysis and Discussion

In this section, we will analyze and discuss how we design our work to make a smart home to fulfill "comfort + convenience + security" when interacting.

6.1 Comfort

"Comfort" is related to QoS and how to perform services. For the former, our proposed work lets smart home system manage to perform services at the best QoS level each space can afford. As for the latter, our work dynamically expands the interaction scopes of services and notifications. And whenever there is a notification for a user, if the user is in some F.Space, it means that the user attention level is high enough, so we assume that the user is willing to be notified, and then the notification is directly performed. Otherwise, if the user is not in any F.Space, it means that user attention level may not be high enough, or the user is not willing to be interrupted, so the smart home will prompt login first. If the user agrees to interact, he/she can simply login for further information. If the user does not want to be interrupted, he/she can just ignore the notification to preserve privacy. This causes no harm since the user will only miss the notification that "a background service can not be performed currently" or miss the foreground services the smart home initiates, which is not preferred by the latter user currently.

6.2 Convenience

"Convenience" is related to the relationship between the user and the space where services are provided. We arrange foreground services at highly interactive spaces so user will not miss them. Relatively, we arrange background services at spaces with low attention, thus saving resources of highly interactive space for foreground services. In addition, sorting CSs by their distance can minimize the necessary move for the user, thus improving his/her convenience.

6.3 Security

"Security" is related to the privacy issue, and its complexity comes from the multiuser situation. We define that the privacy level of a space will be lower as involving more users, and by defining "none" in privacy level of space, we prevent service being automatically provided in the spaces where the user is not there. Further, IF.Space is related to the control modality of autonomy in [8], since in our work, the user can leave F.Space and thus keep attention-required foreground services out. This mechanism shows the freedom for a user to control his/her privacy. As for X.Space, it is related to the control modality of solitude in [8], which addressed the physical privacy. By arranging X.Space at the last position either in the CS list or in the notification order, we prevent the interruption between each user as much as possible by allocating them different spaces. Finally, when a F.Space transits into IF.Space, all the foreground services whose attention requirements are higher than current interaction level will be paused until the user re-login to resume them. On the other hand, when a F.Space transits into B.Space, all the foreground services will be ceased. This achieves the control modality of autonomy and confidentiality in [8], by keeping the user-related services private.

7 Experiment Results and Evaluations

First, experiments are conducted to verify that human-system interaction can be improved by matching foreground service in foreground space rather than background space. And then we evaluate our proposed work by interviewing 54 people with application scenarios before and after applying our framework and algorithm.

7.1 Experiments

The experiment data are collected from 3 users. In our first experiment, we ask the user to focus on his/her desktop PC, which is F.Space. Aside the user, a mobile PC which can not be interacted directly is set as B.Space. A service is randomly initiated at F.Space or B.Space every 3~5 minutes, along with a message randomly instructing the user to accept/reject this service. The user is asked to follow the instruction, thus to make sure this service having user attention and to simulate it as a foreground service. The response time is calculated as the period from the moment initiating the service to the moment receiving user response. The results in Table 1 verify that the user responds faster when service is initiated in F.Space.

Table 1. Response time for a randomly initiated foreground service

Scenario	Average	Standard Deviation	Data Amount
In F.Space	2924 ms	814 ms	268 samples
In B.Space	6752 ms	965 ms	257 samples

The environment for our second experiment is the same, but in addition to the randomly initiated service, the user is asked to continuously interact with another foreground service in F.Space. We measure the response time for the continuous foreground service at three scenarios: the random service appears at F.Space,

Table 2. Response time for a continuous foreground service

Scenario	Average	Standard Deviation	Data Amount
Normal	3990 ms	914 ms	1141 samples
In F.Space	9407 ms	1863 ms	125 samples
In B.Space	14122 ms	3015 ms	131 samples

B.Space, or does not appear. The results in Table 2 also verify that the user returns to his/her original focused tasks faster if the random service is initiated in F.Space.

7.2 Evaluations

The first scenario is about Security. Without our work interactions continue, though inappropriate inhabitants join in. With our work, interactions will automatically pause or continue based on the privacy level resulted by users in the space. The second scenario is about how to deal with paused interactions when inhabitants return to space where interactions pause, which is about Comfort. Without our work, interactions are automatically engaged (though inhabitants may not wish to interact). With our work, system prompts inhabitants first, and then waits for them to decide what to do. The evaluation results in Table 3 verify that our work can upgrade the appropriateness of system behaviors.

Table 3. Appropriateness of system behaviors before and after applying our work

	Scenario1(Before)	Scenario1(After)	Scenario2(Before)	Scenario2(After)
Appropriate	7.4%	88.8%	48.1%	85.2%
No Comments	7.4%	5.6%	14.9%	11.1%
Inappropriate	85.2%	5.6%	37%	3.7%

8 Conclusion

In this paper, we proposed a framework to model the interaction between a smart home and its inhabitants, thus to fulfill "comfort+convenience+security" when a smart home performs services to interact with its inhabitants. Our framework mainly focuses on the relationship between services, spaces, and users, and related analysis is provided. A corresponding algorithm is also proposed to appropriately configure services, so that a smart home can behave both smart and as a home.

References

1. Wu, C.-L., Liao, C.-F., Fu, L.-C.: Service-oriented smart home architecture based on OSGi and mobile agent technology. IEEE Transactions on Systems, Man, and Cybernetics C 37(2), 193–205 (2007)

2. Sousa, J.P., Poladian, V., Garlan, D., et al.: Task-based adaptation for ubiquitous computing. IEEE Transactions on Systems, Man, and Cybernetics C 36(3), 328–340 (2006)
3. Ramakrishnan, N., Capra III, R.G., Perez-Quinones, M.A.: Mixed-initiative interaction = mixed computation. In: Proc. ACM SIGPLAN Workshop, pp. 119–130 (2002)
4. Sawhney, N., Schmandt, C.: Nomadic radio: speech and audio interaction for contextual messaging in nomadic environments. ACM Transactions on Computer-Human Interaction 7(3), 353–383 (2000)
5. McCrickard, D.S., Chewar, C.M., Somervell, J.P., et al.: A model for notification systems evaluation - assessing user goals for multitasking activity. ACM Transactions on Computer-Human Interaction 10(4), 312–318 (2003)
6. Hinckley, K., Pierce, J., Horvitz, E., et al.: Foreground and background interaction with sensor-enhanced mobile devices. ACM Transactions on Computer-Human Interaction 12(1), 239–246 (2005)
7. Buxton, W.: Integrating the periphery and context: a new model of telematics. In: Proc. Graphics Interface, pp. 239–246 (1995)
8. Boyle, M., Greenberg, S.: The language of privacy: learning from video media space analysis and design. ACM Transactions on Computer-Human Interaction 12(2), 328–370 (2005)
9. Mani, A., Sundaram, H., Birchfield, D., et al.: The networked home as a user-centric multimedia system. In: Proc. ACM Workshop Next-Generation Residential Broadband Challenges, pp. 19–30 (2004)
10. Schmidt, A., Kranz, M., Holleis, P.: Interacting with the ubiquitous computer: towards embedding interaction. In: Proc. Conference on Smart Objects and Ambient Intelligence: Innovative Context-Aware Services: Usages and Technologies, pp. 147–152 (2005)
11. Elliot, K., Watson, M., Neustaedter, C., et al.: Location- dependent information appliances for the home. In: Proc. Graphics Interface, pp. 151–158 (2007)
12. Nakajima, T., Satoh, I.: A software infrastructure for supporting spontaneous and personalized interaction in home computing environments. Personal and Ubiquitous Computing 10(6), 379–391 (2006)

Biologically Reasoned Point-of-Interest Image Compression for Mobile Robots

M. Podpora and J. Sadecki

Department of Electrical and Computer Engineering,
Opole University of Technology, Opole, Poland
michal.podpora@gmail.com,j.sadecki@po.opole.pl

Abstract. In this paper authors describe image compression based on the idea of biological "yellow spot" in which the quality/resolution is variable, depending on the distance from the point-of-interest. Reducing the amount of data in a robot's vision system enables to use a computer cluster for non-time-critical "mental" processing tasks like "memories" or "associations". This approach can be particularly useful in HTM-based data processing of robot's vision system data.

1 Introduction

Nowadays, mobile robots are mostly controlled by embedded systems. A lightweight but efficient processing solution, powerful enough to acquire input data and to control the mechanics can be purchased for a fair price. On the other hand, intelligent and behavioral algorithms for a robot are much more demanding tasks. Using a vision system can also be a reason for replacing the processor with a faster one. However, powerful computers need more energy, which causes that the "mobility" of a robot is hindered by heavy batteries.

Few years ago, some groups of researchers have tried to make use of a computer cluster for handling algorithms not demanding real-time processing. The efficiency of computer clusters was strongly affected by communication time because video data sent between nodes was causing serious performance problems. Since then, processors have evolved, but wireless networks, which could connect a mobile robot with the cluster, have not seemed to keep up with the processors' evolution.

Designing a mobile robot's human interaction algorithm is always a great challenge and a computer cluster would be a perfect solution for "mental" non-real-time processing tasks like "memories" or "associations". Unfortunately, using a computer cluster for image processing and object recognition (i.e. the usual way of making the robot understand the image) is not possible due to the communication time. Reducing the amount of the transferred data usually causes loss of information in images. Object recognition gives poor results on noisy images.

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 371–382.
springrlink.com

Biologically Reasoned
Unlike “artificial” vision systems, a human eye does not acquire the whole image using the same detail level or resolution for every part of the image. In fact, a greater part of “visual data” comes from a yellow spot, which is a part of the retina containing fovea (i.e. small pit responsible for central vision with more closely packed cones) providing the sharpest, and the most detailed image [1]. The nature has found its way to reduce the amount of data transferred to the “supercomputer”.

This approach to visual data compression (presented in Fig.1) causes a fundamental change in object recognition.

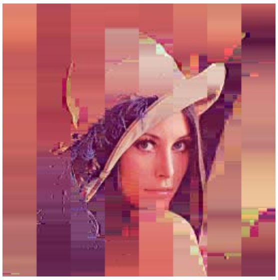


Fig. 1. Lena, vision system’s yellow spot is fixed on her eye

2 Object Recognition

An object recognition algorithm based on a sequence of images compressed as in Fig. 1 can be implemented using HTM networks (Hierarchical Temporal Memory). In HTMs it is absolutely natural to use spatial fragments of input data and to send it to the input of the network in a time sequence [2]. The idea of HTM network’s understanding of the input data/signal [3] is based upon human brain’s neocortex. A human eye makes very fast movements called saccades. These moves change the fixation point of retina’s yellow spot causing a change in the observed detail. By generating sequence of details, it is possible to build up a virtual map of corresponding image.

A robot can also be programmed to use its vision system this way. A video camera or a stereovision system acquires video data and, subsequently, images are compressed with the use of yellow spot’s coordinates for defining the point-of-interest (point of maximum quality of the image after compression). The information is passed to the HTM network on a remote computer and processed in a usual manner, just like any other information within HTM networks, making inference possible.

An abstract notion of a dog can be described as a sum of features like specific sounds or spatial patterns acquired by the retina (e.g. tail, nose, ear,paw) [4].

3 Yellow Spot

Human Eye

In a biological eye, the quality/resolution of image is decreasing with the distance from the fixation point. The difference is not equal in every direction from the fixation point, although it seems to be so. If we look at the middle of a circle, the circle seems to be round (well, it is) but it is not transferred through the optic nerve as a circle (cones are not arranged uniformly on the retina) although the brain interprets it as a circle. This example shows that our "vision system" also makes some conversions and simplifications to the acquired data before processing.

"Robot's Yellow Spot"

An "artificial" vision system acquires visual data as a predefined pattern of pixels organized in columns and rows. This is very problematic for people willing to experiment with an HTM network, because the network expects temporal sequences of spatial patterns on its input. The most frequent solution is to scan the whole image with the input sensor matrix.

Having the yellow spot in mind, "temporal sequence" can be understood not as a horizontal shift of the sensor matrix but as a sequence of patterns found in the neighborhood of the yellow spot. Fig. 1 shows an idea of reducing the amount of data in the image, while not loosing the quality or resolution of the "virtual yellow spot's" neighborhood.

The image reduction cannot be made as it is shown in Fig. 2a because the information outside the central area is sometimes also useful. For instance, HTM cannot decide to change yellow spot's coordinates to the nose because there is no nose in Fig. 2a. Of course, well-trained HTM could "suppose" where the nose should be found, but at first, it should have good opportunities to learn.

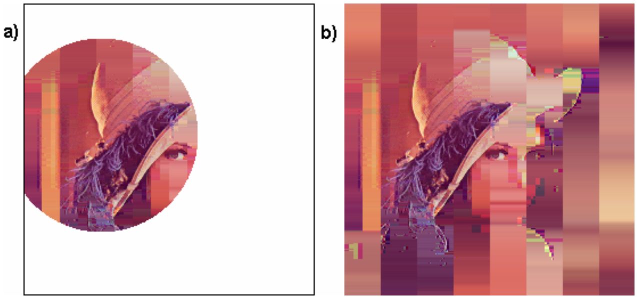


Fig. 2. Lena, vision system's yellow spot is fixed on a detail. (a) part of the image with the highest compression level (i.e. the lowest quality) is cut-off. (b) the same image without cutting.

4 Compression

It does not matter what kind of compression algorithm is to be used for compressing the images exactly but the algorithm should give the following abilities:

- the quality/resolution is changing smoothly with the distance from yellow spot's coordinates,
- it would be very helpful if the high-level compressed area could show, despite compression, the existence of details.

5 Practical Realization

In the practical realization, the simplest possible DWT (discrete wavelet transform), that is the Haar's wavelet transform [5], was used, analyzing horizontal details only. Afterwards, all calculated differential matrices are compressed with a simplified RLE (run-length encoding) procedure, with only zeros compressed. The quality of image was defined by the threshold level determined not as a constant value but as a modified Gaussian function (Fig. 3).

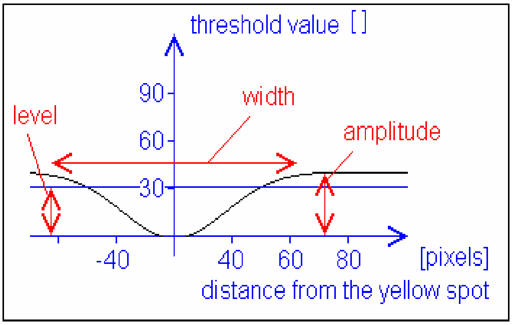


Fig. 3. Threshold function

By modifying the parameters presented in Fig. 3, it is possible to control parameters of the threshold function (see Fig.4).

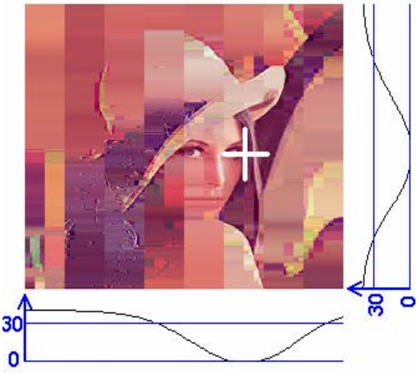


Fig. 4. Lena and threshold curves used within the DWT transformation

Parameters of the threshold function, i.e. the quality of the compressed image, can be chosen automatically by comparing original image to the compressed one, or manually. In the latter case, the application developer has the opportunity to watch the influence of a specific parameter on the compressed image. Fig. 5 shows contour lines of the threshold function.

The image in Fig. 5 c) seems to be useless for image processing and object recognition algorithms, but the neighborhood of the yellow spot is compressed with the best quality. Therefore, in some cases, e.g. for HTM networks, such image is equally as useful as uncompressed one. It is good enough for analysis of the observed detail and for perceiving the existence of other details in its neighborhood.

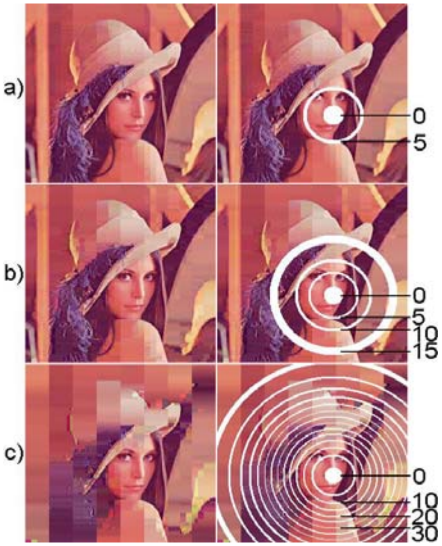


Fig. 5. Contour lines of the threshold function. White lines denote lines/areas with (threshold modulo 5) = 0. Numbers describe threshold value in specified area.

Table 1. Data size of an example transformed image

Image quality	Example data size
Best	e.g. 212 327 bytes
Very good (Fig. 5 a)	e.g. 105 261 bytes
Good (Fig. 5 b)	e.g. 72 070 bytes
Poor (Fig. 5 c)	e.g. 33 285 bytes

Values in Table 1 correspond to images in Fig. 5. Although the original file contained 196662 bytes (196608 and header), the first row of the table shows the weaknesses of the RLE encoding.

The compressed image data size depends on threshold function parameters, source image complexity, exact coordinates of the yellow spot and on some other factors, therefore it is difficult to estimate it.

Entropy coding can produce better results [6] than the proposed compression, it is not the author's intention to find an ultimate compression algorithm, but rather to show usefulness of compressing images with variable quality/resolution.

6 Active Vision vs. Yellow Spot

Choosing a smaller area of the image to speed-up the processing is well-known idea, called "active vision". However, active vision usually ignores the rest of the image, while the proposed algorithm still gives the possibility to "notice" the existence of a new point-of-interest beyond the analyzed area. A DWT-transformed image reduces the information about small input signal changes and stores the information about big changes. Terms "big" and "small" are polarized by the threshold parameter (or function). This allows detecting contrasting objects (differencing in color or lighting) located anywhere within the field of view (or detecting movement) while offering significant data size reduction. The developed application's algorithm is designed to transform the image using threshold function with the predefined maximum value of 32. It offers much better image quality than in Fig. 5c and more efficient data compression than in Fig. 5b. In Figs. 5c and 5b, the maximum threshold value is 18 and 68, respectively. The main goal of using the threshold function instead of a fixed threshold value is to compress the image not loosing the quality in the neighborhood of yellow spot, and not loosing the information about contrasting or moving objects.

Compression of images is essential for reducing communication time, which is an important parameter of a computer cluster. Using a computer cluster with raw data is useless.

Although lossy compression often makes further image processing very difficult or even impossible, the proposed algorithm is transparent for an HTM network because HTM networks require only a specified spatial fragment of the image. Even the image shown in Fig. 5c should be sufficient for an HTM network because the yellow spot's neighborhood is compressed with maximum quality.

7 Inference

HTM networks consist of nodes connected together in a tree topology. The lowest layer of nodes uses the sensory data, whereas all other layers usually use preprocessed knowledge from a previous layer. The hierarchical structure of HTMs combined with nodes' algorithm (based on the autoassociative memory) gives the ability to analyze the input data as a spatial and/or temporal sequence of input patterns, just like it happens in neocortex [3].

HTMs are capable of discovering causes in the analyzed input data, and inferring causes of novel input. Optionally, they can be also used for predicting the input data or controlling the application's and/or robot's behavior [2].

Discovering causes is possible because objects are not stored in the memory as spatial patterns. Although the input data is acquired as a spatio-temporal pattern of pixels, this representation can be found only in the first, sensory-level layer. All other layers build an abstract pattern of active nodes to represent a specified object. An object in the real world is called "cause" and its representation in HTM (a probability of each of the learned causes) is called "belief". As an example, a spoken language can be used. Each layer of an HTM contains some active and some inactive nodes, while every layer represents more and more abstract and complex "causes". The sensory data carries the information about current "active" sound frequencies, but higher layers can indicate phonemes, words, phrases, sentences and ideas. This process is similar for any possible input data, so implementing HTMs in a machine vision approach should not induce any change in the inferring algorithm.

HTMs can be easily used for image recognition [7], if the input node is modified to use spatial, 2-dimensional input data. The temporal aspect of sensory data can be generated artificially by scanning the whole image line-by-line with the input nodes' input arrays. It can also be derived from biology, just like human eye moves from one fixation point to another with every saccade [6].

HTM's layers are built of nodes, where each node learns "causes" and forms "beliefs" [2]. Nodes in the first layer take arrays of pixels and form "beliefs" about e.g. basic shapes (see Fig. 6, first column). Next layer nodes analyze spatial and/or temporal coexistence of active input layer nodes and form further "beliefs" (Fig. 6, columns 2-5).

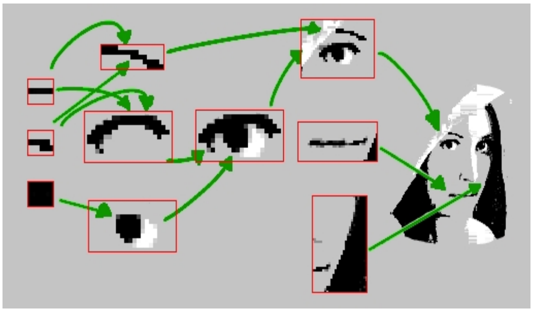


Fig. 6. Hierarchy of patterns on different abstraction levels. The face on the right side of this figure should not be understood as an assembled bitmap, but rather as an abstract state of coexistence of "beliefs" from the previous layer of nodes (e.g. "an eye" [90% of a certainty]+"a mouth" [40% of a certainty] +"a nose" [40% of a certainty] = "a face" [50% of a certainty]).

Various parts of the image can be acquired separately by moving the yellow spot from one part to another. The machine vision system must be cognizant of analyzing one particular object. A highly probable belief of seeing an eye followed (after a saccade) by a highly probable belief of seeing a nose might mean that the robot is currently looking at a human being. This might be checked e.g. by an intentional saccade to another fixation point.

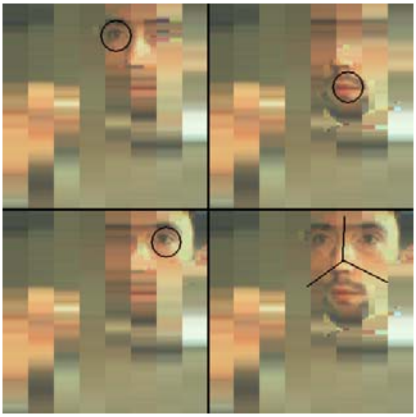


Fig. 7. Three fixation points of the acquired image seem to be enough for a human eye to recognize a face. Although human brain has already recognized the object, saccades still occur to send more details about it.

As shown in Fig. 7, use of saccades and fixation points together with the described DWT+RLE algorithm is a very powerful idea.

The most important feature of this solution is that it uses HTMs – pixels located in the neighborhood of every fixation point are used for generating “beliefs” about fragments of an image (e.g. “an eye [80% of a certainty]”, “a mouth [60% of a certainty]”). These “beliefs” are crucial for inferring what exact “causes” are reasons of a final “belief”.

While HTMs are based on autoassociative memory, input data can be incomplete or slightly differ from the learned example, and the algorithm would still be able to recognize it.

A house with a blue roof will be recognized as a house even if the robot have never “seen” a blue roof before because “beliefs” of some of nodes indicate that this object has windows, door, etc. and the most probable “cause” of the final “belief”, among all known “causes”, is “a house”.

Indicating if the analyzed object is already known or new also seems to be easy task. For all known objects, the belief distribution is peaked, and for unknown objects, it is flat [2].

8 Election of a Next Fixation Point

The developed application has two modes of defining a new fixation point. The first one, with lower priority, defines new fixation point in the middle of a moving object and starts analyzing it. The second operating mode is active only when an object is being analyzed. The movement of other objects is ignored, and further fixation points are defined depending on decision of the analyzing algorithm.

Current version of the application is not using HTMs’ prediction feature, which seems to be the most promising direction of research.

While HTMs store the information about spatio-temporal sequence of patterns, it is possible to make some predictions. If the analyzed image contains “an eye” and “a mouth”, some of the nodes could already indicate, that it is probable that the common “cause” is “a face” (Fig. 8). However, there are more details describing “a face”, for instance “a nose” or “an eyebrow”. HTM might use this information for checking if there is “a nose” (see Fig. 8). Choosing next fixation point with the use of HTMs is a complex task, but it seems to be the most resolute and effective way.

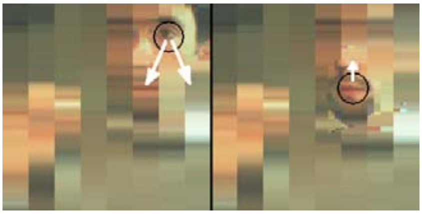


Fig. 8. Election of another fixation point, basing on HTM’s prediction feature (predicting spatio-temporal patterns of partially recognized objects)

9 Mobile Robots and Computer Clusters

The most problematic task in using computer cluster as an enhancement of robot’s memory and/or association ability was the communication time. Input data of robot’s vision system, i.e. images from its camera, produced extremely high volume of the traffic. No parallel algorithm could manage it.

The presented idea of yellow spot enables the possibility of using the cluster. The parallel algorithm should be designed to use task parallelism rather then data parallelism [8]. Analysis of half of an image is pointless, although some data parallelism can also be considered, depending on the case of use.

A computer cluster is, in general, a group of loosely coupled computers that work together closely so that they can be viewed as though they are a single computer [9]. Efficiency of parallel (cluster) implementation of computation depends, in general, on the processor performance and communication bandwidth of the cluster. The value of communication bandwidth is especially important for the problems for which exchange of data between each pair of the cluster nodes is required, especially for systems containing very large number of processors. Highly parallel computers contain thousands of processors. If, for a given problem size, communication will eventually dominate computation as the number of processors is increased, then the speedup cannot scale with large numbers of processors without introducing additional levels of parallelism.

Computational burdens encountered when solving complex problems of control of multidimensional processes (robots), can be essentially reduced by decomposing the problem into a number of subproblems and by solving a set of subtasks associated with a certain coordination task.

Assuming a set of all operations required to be carried out for solving one local task as the least portion of a task which can be performed by a processor, the discussed two-level algorithm can in a simple manner be implemented parallel in multiprocessor system. Notice that the Master-Slave structure, being natural for methods of this type, is less effective for more complex cluster systems. Furthermore, in the distributed memory systems, each of the tasks specified in a given algorithm, should be solved in a parallel way, i.e. in case of the considered methods the coordination algorithm ought also to be solved parallel. The vertical communication corresponds in principle to data exchange between two algorithms implemented by the same processor, namely, between the algorithm implementing a local task and a fragment of, allocated to this processor, the coordination algorithm implemented parallel. On the other hand, the horizontal communication corresponds to that one between particular processors, including the exchange of data necessary for the correct implementation of the coordination algorithm.

Decreasing of the computation time arising as the consequence of parallel realization of two-level optimization problem is rather evident. But decomposition can lead to decrease of the communication requirements too, especially, for the each-to-each communication problem (bi-directional data transmission between each pair of processors). For example, a number of required communication tasks L_s for the latter problem is determined as:

$$L_s = P(P-1) \qquad (1)$$

where P denotes the number of processors.

If we divide all processors into L groups (nodes), each of them containing approximately P/L processors, the total number of communication tasks required to perform the same exchange of data, is determined as follows:

$$L_d = \frac{P}{L}\left(\frac{P}{L}-1\right)L + L(L-1) + \left(\frac{P}{L}-1\right)L \qquad (2)$$

where $(P/L)(P/L\text{-}1)L$ denotes the number of required communications task in all groups of processors (nodes), $L(L\text{-}1)$ denotes the number of communication tasks between all groups and $(P/L\text{-}1)L$ denotes total number of communication tasks allowing to send data obtained from other groups into all processors in each group.

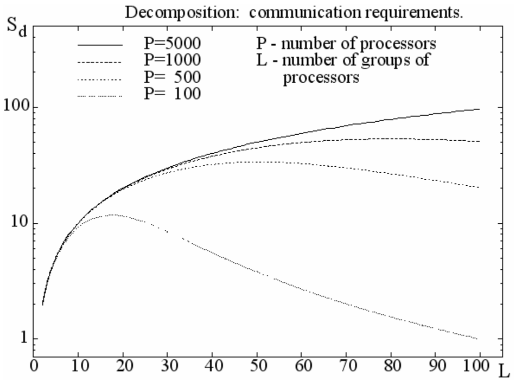


Fig. 9. Communication requirements for the two-level parallel each-to-each communication

Fig. 9 presents values of the factor $S_d=L_s/L_d$ as a function of a number of processor groups L for given value of P. This picture shows that the communication requirements can be significantly decreased as a consequence of the realized decomposition.

10 Conclusion and Future Work

Analysis of high-resolution image sequences demands computation power and operating memory. Using a computer cluster not only proper data representation on system input is an important aspect of fast and efficient processing, but also defining proper “neural” architecture [10] and problem decomposition.

HTM networks seem to be the most natural conception in vision understanding, and a new hope for active vision adherents. While HTM networks are relatively new idea of implementing memory and vision understanding, it is very hard to predict efficiency, elasticity and robustness of developed system before it is ready.

HTM network is known to manage with image recognition [7], but it is not fully implemented in the application (Fig.10) yet. The next step is to implement an HTM network in an example vision system consisting of mobile robot with video capturing devices, and a computer cluster for advanced analysis of image, object recognition and inferring – HTM-based vision understanding.

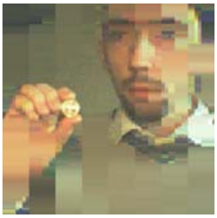


Fig. 10. A single frame with yellow spot fixed on the coin is firstly DWT-transformed. Secondly, the frame is RLE-compressed, then transferred from the input node to the computer cluster and finally IDWT-transformed for further processing. After processing, a new fixation point coordinates are sent back to the input node.

Acknowledgement

Work co-financed by European Social Fund.

References

1. Hecht, E.: Optics, 4th edn. Addison-Wesley, San Francisco (2002)
2. Hawkins, J., Dileep, G.: Hierarchical temporal memory- concepts, theory and terminology. Numenta Inc. (2007), `http://www.numenta.com/Numenta_HTM_Concepts.pdf` (accessed March 2, 2009)

3. Hawkins, J., Blakeslee, S.: On intelligence. Times Books, New York (2004)
4. Hawkins, J.: Learn like a human. IEEE Spectrum 44(7) (2007), `http://www.spectrum.ie-ee.org/apr07/4982` (accessed March 2, 2009)
5. Mallat, S.: A wavelet tour of signal processing. Academic Press, San Diego (1999)
6. Podpora, M.: Biologically reasoned machine vision: RLE vs. entropy-coding compression of DWT-transformed images. In: Proc. EEICT Conference, Brno (2008)
7. Numenta Inc. Numenta Pictures Demonstration Program (2007), `http://www.numenta.com/about-numenta/technology/pictures-demo.php` (accessed March 2, 2009)
8. Sadecki, J.: Parallel optimization algorithms and investigation of their efficiency: parallel distributed memory systems. Internal Report Opole University of Technology, Opole (2001) (in Polish)
9. Baker, M.: Cluster Computing White Paper (2001), `http://arxiv.org/abs/cs/0004014` (accessed March 2, 2009)
10. Nałęcz, M., Duch, W., Korbicz, J., Rutkowski, L., Tadeusiewicz, R.: Biocybernetics and biomedical engineering neural networks. Akademicka Oficyna Wydawnicza Exit, Warsaw 6 (2000) (in Polish)

Surgical Training System with a Novel Approach to Human-Computer Interaction

A. Wytyczak-Partyka[1], J. Nikodem[1], R. Klempous[1], and J. Rozenblit[2]

[1] Wrocław University of Technology, Wrocław
ryszard.klempous@pwr.wroc.pl
[2] University of Arizona, Tucson
jr@ece.arizona.edu

Abstract. In this chapter we present a surgical training system that employs some novel ideas on interaction between the trainee and the system.

1 Introduction

Laparoscopic surgery brings significant benefits into the healing process and therefore an increasing interest in it is observed. Great benefits over traditional surgery include: limited scarring, reduction in pain and recovery time, leading to a smaller risk of complications. Study conducted by Hansen et al. [1] shows that patients who have undergone laparoscopic appendectomy had five times fewer wound infections, two times shorter discharge time and fewer of them required narcotic analgesia. On the other hand, there is a number of downsides, for instance the surgeon's perception – both haptic and visual - is very limited, which extends the procedure time (in the open appendectomy case 40 as opposed to 63 minutes in laparoscopic appendectomy [1]) and the likelihood of human error. Also investment in expensive instruments and a very long training period are required.

Surgical training should be modular, where each module should focus on the development of certain behaviors - knowledge-based, rule-based and skill-based. This chapter will focus on a system designed particularly for the skill-based behavior training.

The primary skills that have to be developed during the training involve - depth perception, dexterity and hand-eye coordination. Traditionally, according to the Surgical Papyrus [2] since the times of ancient Egypt, surgical training has always followed the model of master and apprentice and involved mentorship in real-life clinical cases, where the apprentice would gain skills and experience from his teacher. That approach to surgical education hasn't changed significantly since ancient times. It is also worth noting that the training was strictly dependent on the availability of a patient and tied with the course of patient care [3-5].

Since some of the surgical skills do not strictly require practicing them in a real-life clinical situation - it is desirable to enhance them in a safe environment, without any risk to patients. It is also important to note, that the operating room is not the best learning environment because of factors such as stress, time constraints and costs that have a negative impact on the learning process.

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 383–393.
springrlink.com

On the other hand, reference [6] shows, that surgical simulation has positive impact on improvement of psychomotoric skills, i.e. the gallbladder dissection procedure was 29% faster for residents trained on simulators, also accidental injuries during the procedure were 5 times less frequent in that group. Similar results in a cholecystectomy procedure are shown in [7]. Therefore a number of surgical simulators, both physical-model-based and software-based, have been developed, all allowing the trainees to safely master the basics skillset.

One of the first simulation devices for laparoscopic training was the Pelvitrainer, designed by Karl Semm in 1986 [8]. The original concept consisted of a transparent box, where organs were put. The box contained several holes to introduce the instruments and the camera, Figure 1. The training method proposed by Semm was gradual: at first the trainee would only learn to use the instruments, without the use of the endoscope, secondly the endoscope would be introduced, while the trainee would be still allowed to occasionally look at the organ through the transparent walls of the box, and finally the box would be covered with a cloth to obscure vision.

Since 1986 the concept of a pelvi-trainer has influenced many simulators and is currently used in many training programs.

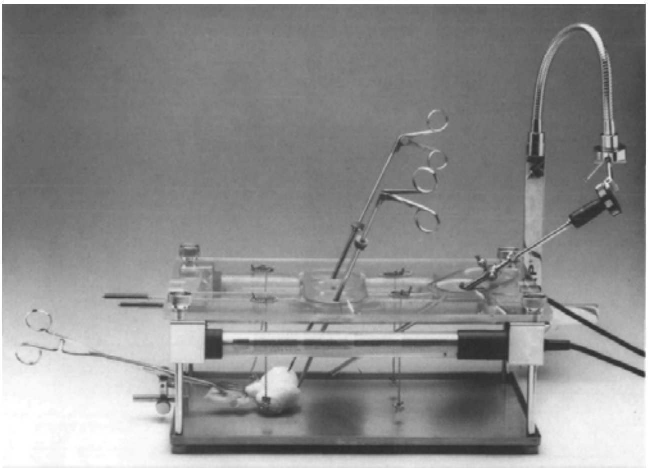


Fig. 1. Karl Semm’s pelvi-trainer [8]

Semm's pelvi-trainer is a typical example of a skill-based training device that focuses on development of dexterity, depth perception and hand-eye coordination. There are several simulators that serve for knowledge- and/or behavior-based training as well as skill-based.

One of the first systems built for that purpose is the Karslruhe Endoscopic Surgery Trainer and the KISMET software [9, 10]. It is a complete training system with built in scenarios for practicing procedures like laparoscopic cholecystectomy. There are several products similar to the Karslruhe Trainer, for instance the MISTELS system [11], or the LapMentor [12]. A lot of effort has been put in those systems to reproduce details such as graphics, organ deformations and haptic feedback. Needless to say those systems are very expensive.

One skill-based approach that combines a simple pelvi-trainer with a high-tech system is the Virtual Assistant Surgical Training (VAST) system, developed at the University of Arizona [13]. It is comprised of a pelvi-trainer and a computer system. The computer, through a magnetic position sensor, collects data about the instrument's tip position, which is used to rate the trainee's performance in a certain exercise, based on time, path length and accuracy. The individual's progress can therefore be precisely measured and monitored.

The approach represented by VAST is especially interesting and will serve as a basis for the further described training system which incrementally adds to the VAST trainer.

2 Description of the System

The purpose of the proposed system is to aid a trainee in the development of basic skills, as dexterity, hand-eye coordination and depth perception. The novel way of interaction proposed in this chapter – allows the trainee to develop the skill of avoiding certain regions where the appearance of an instrument might inflict damage to the patient.

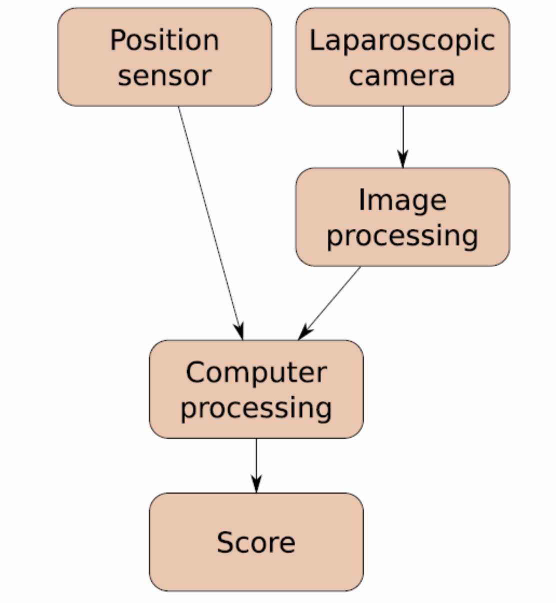


Fig. 2. System outline

2.1 System Outline

The system resembles the VAST prototype trainer [13] and is comprised of a standard pelvi-trainer setting – an endoscopic camera and two instruments - and a computer. The instruments have an embedded position sensor. The video output of the camera, as well as the position sensors, are connected to the computer, where additional processing occurs, as in Figure 2. There are two results of the processing. First - each exercise performed by a trainee is scored, and thus the performance can be analyzed, and secondly – the trainee is optionally informed by an auditory signal coming from a speaker, whenever he approaches the hazardous area with the instruments.

The score achieved in the exercises is calculated by the computer and based on the information collected from position sensors and the video camera.

Initially, before the exercise can start, and after a new model is introduced to the pelvi-trainer – a 3D representation of the model is built, from images collected with the endoscopic camera. Secondly – a hazardous region can be selected, the trainee has the opportunity to see and understand where the region is on the model, so it can later be avoided, during the course of the exercise. This can be done individually for each exercise to introduce more difficulty into the exercises.

Later the coordinates of the instruments obtained from the position sensors can be used to determine the location of the instrument in the reconstructed environment.

The 3D structure of the viewed scene, recovered in the process of the system's work, is at this point not intended strictly for visualization - the trainee observes the image as it comes from the endoscopic camera, but rather for the purpose of determining the hazardous regions within the 3D model. Through fusion of the 3D reconstructed model and sensor data the coordinates collected from the position sensor located on the tip of the instrument can be used to determine it's location and proximity to the hazardous area. If the position sensor indicates appearance of the instrument in a hazardous region an auditory signal is produced and a penalty score applied.

2.2 Training Model

The training is based on simple tasks that trainees have to complete within a specified time. Several types of exercises have been proposed by others [14, 15], the common goal is to practice dexterity, coordination and depth perception.

Example exercises are:

- knot tying,
- cutting and suturing,
- picking up objects,
- touching different points on a model.

Each exercise is associated with a different physical model which is placed in the pelvi-trainer's box.

The trainee's score is calculated with respect to:

- elapsed time,
- length of the path of the instrument tip,
- accuracy.

It is proposed that another factor is introduced and that the score is significantly decreased upon hitting a hazardous region, defined at the beginning of each exercise, therefore the scoring function changes from

$$S = f\left(\frac{k_t}{t}, \frac{k_s}{s}, k_A \cdot A\right), \tag{1}$$

to

$$S = f\left(\frac{k_t}{t} + \frac{k_s}{s} + k_A \cdot A + k_H \cdot H\right), \tag{2}$$

with H, the hazard measure, defined as

$$H = -\frac{2R_H}{|d(H_c, T)| + R_H} + 1 \tag{3}$$

The terms H_c, T and R_H in equation 3 denote respectively - the coordinates of the center of the hazard sphere, the coordinates of the instrument's tip and the radius of the hazard sphere.

Such a way of scoring the trainee allows considering the events of breaching the no-fly zones and is an important part of the training part of the system.

2.3 Processing

The following section describes the method for recovering the 3D structure of the scene applied in the presented system, which is the initial step in the work of the proposed system.

During laparoscopic procedures it is natural that after introducing the camera into the body, the surgeon performs a series of movements with the camera to find the best viewing point for performing the procedure, and also to discover any potential abnormalities in the viewed organs. It is therefore natural that a video sequence containing images of the operating field can be obtained in that manner also during the training procedure and serve as a source of images for a structure from motion 3D geometry recovery algorithm.

Structure from Motion

Since the early 1980's there has been a lot of research in the field of recovering 3D structure from camera image sequences [16, 17]. Current state of the art algorithms [18, 19] perform the task without previous camera calibration and with the only constraints on the image sequence that a certain number of corresponding feature points between frames can be established and that a sufficient baseline (distance between the camera locations) exists. It has been proven [18] that the point-correspondences, along with several constraints on the camera, can lead to a metric reconstruction of the imaged scene.

In general the reconstruction process consists of: (a) selecting potential point matches between the frames of the sequence, (b) estimating the 3D geometry, (c) refinement. Since the whole process is based on point correspondences it is very important that the first step is performed carefully. Therefore potential matches, (m,m'), are selected from a set of points, populated through a Harris' corner detection procedure [20], that contains only interest points that significantly differentiate from their neighborhood and can be matched with their corresponding points in other views using a similarity measure, i.e. cross-correlation. It is important that the usage of a video sequence limits the search for matches to a certain subwindow of the image, because the camera movement between the frames (baseline) is small and therefore feature points stay in a several-pixel range.

After the feature matching has concluded two initial views are selected that will serve as a base for further sequential structure refinement. The criteria of selecting such views are: a) maximization of the number of features matched between the views, and b) sufficient baseline between the views, so the initial 3D estimate can be properly performed.

While the first criterion is easy to fulfill, based on the results of feature matching, it is significantly harder to determine if a view meets the second one, which is especially important in images from a video sequence, where the baselines are usually very small.

The simplest approach would be to limit the frame sequence and select only every *k*-th frame, where *k* depends on the frame rate of the video. In fact, basing on the constraint that the camera moves are rather smooth and slow and the imaged object itself remains still, it is suggested to perform such a reduction of the input video sequence, by reducing the frame rate to 2 frames per second. Such a reduction essentially decreases the computational effort involved with the feature point selection.

However a robust criterion for selecting key-frames is available [21] and has been employed in the proposed system.

The GRIC, Geometric Robust Information Criterion [21], criterion can be used to determine if the geometry between 2 views is better explained by a homography transformation (true for small baselines) or by the fundamental matrix (larger baselines). Therefore GRIC is suitable to select views which are interesting for the 2-view reconstruction process which only deals with the fundamental matrix model. It is especially important to evaluate GRIC for image sequences obtained with a video camera, where baselines can be really small for consecutive frames, it is less significant in case of image sequences from a still-camera, where most of the baselines are wide enough and the fundamental matrix model scores better anyway - as seen in Figure 4.3(a).

GRIC is a scoring function that takes into account several factors - number *n* of inliers plus outliers, the residuals ei, standard deviation of measurement error, the dimension of the data *r* (4 for two views), number *k* of motion parameters ($k = 7$ for F, $k = 8$ for H) and the dimension d of the structure ($d = 3$ for F, $d = 2$ for H).

$$GRIC = \sum \rho(e_i^2) + (nd\ln(r) + k\ln(rn)). \tag{4}$$

After the initial two views have been selected the images may be further processed for the retrieval of the 3D geometry, next additional views are added, selected in a similar manner as in the initial step. The recovery of the 3D geometry is based on the normalized 8-point algorithm and a triangulation method described in [19]. The 8-point algorithm is used to compute the fundamental matrix, F, and the epipoles (e,e'), from a set of point correspondences (m,m') between 2 images by solving (5).

$$m'^{T}\ Fm = 0. \tag{5}$$

The F matrix can be then used to calculate the camera matrices, (P,P'), which are needed for the triangulation step.

$$\begin{aligned} P &= \left[I_{3x3} \middle| 0 \right], \\ P' &= \left[[e']_x F \middle| e' \right], \end{aligned} \tag{6}$$

where [e]x is a skew symmetric matrix of the form:

$$[a]_x = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix}, \tag{7}$$

and linked with vector product in the following manner:

$$a \times b = [a]_x b. \tag{8}$$

The F matrix is calculated using the RANSAC algorithm, proposed by Fischler and Bolles [22], which is less sensitive to outliers in the matched point set than the standard 8-point algorithms, however it has to be fed with a greater number of potential matches.

The information obtained so far is sufficient to calculate rectifying transformations for the image pairs, which further serve as input for a stereo depth matching algorithm that produces a dense depth map of the scene.

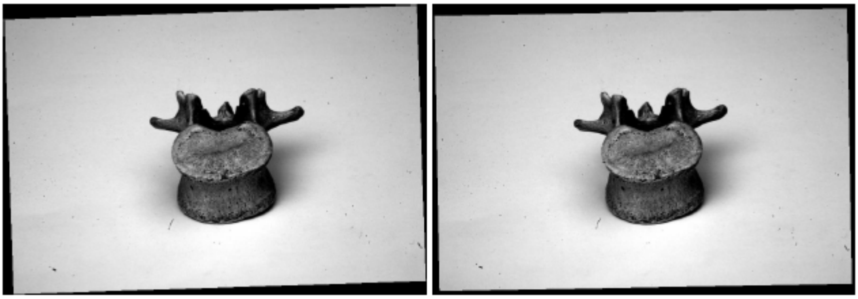


Fig. 3. Rectified image pair

A triangular mesh of the scene is constructed using the Delaunay [23] triangulation method. The mesh is used for visualization of the 3D model of the operating field where at this point no-fly zones can be oriented.

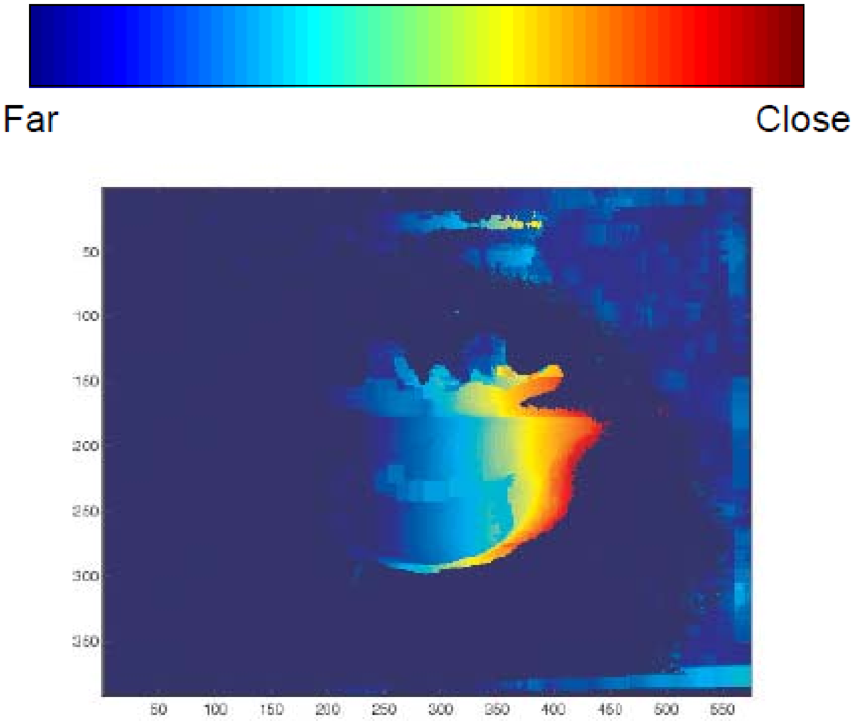


Fig. 4. Non-interpolated depth map of the vertebra, obtained with a segmentation-based depth matching algorithm

Interaction
The interactions are based on the information about the trainee's behavior with respect to the location of no-fly zones. By using fuzzy classification of movements [24] the system can guide the trainee to obtain better results. The advice offered by the system is of the form:

- your movements are too fast,
- your movements are too imprecise,
- your moving too close to a no-fly zone,

also the system can advise the user to go to a certain exercise which is designed to emphasize the particular skill the system considers to be weak.

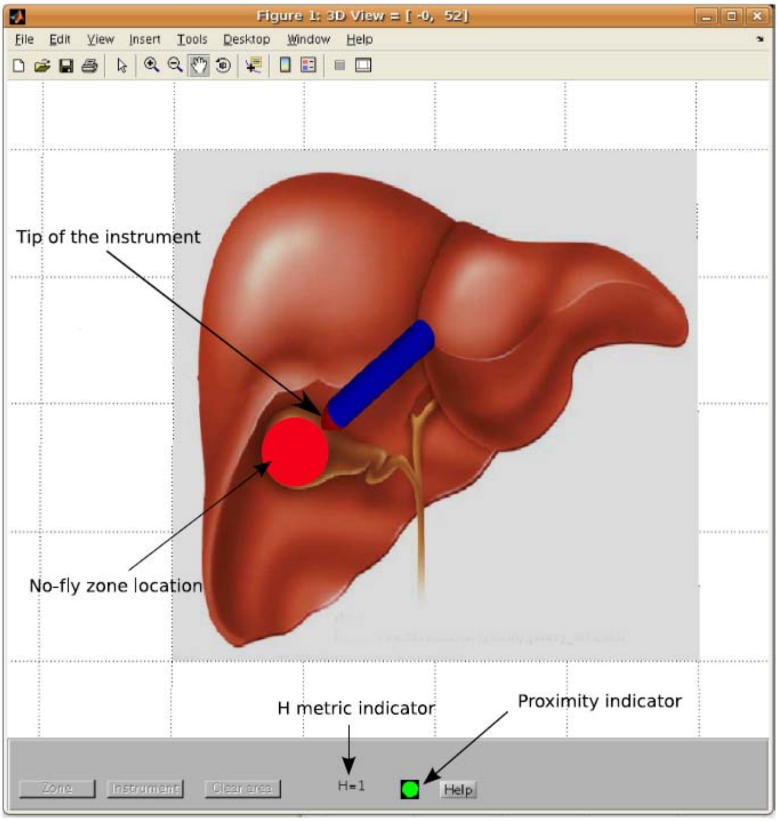


Fig. 5. Illustration of the concept of the simulator of the system, showing example of no-fly zone location and the interaction features like the no-fly zone proximity indicator

3 Conclusions

This chapter describes a training system which utilizes a new interaction scheme for the purpose of aiding in the process of training laparoscopic surgeons.

The video sequence obtained from the laparoscopic camera is used to construct a 3D model of the operating field in which information about no-fly zones is embedded. The locations of no-fly zones and the position of the instrument tip are then used to provide the interaction and guidance during training.

Knowledge of the 3D model and geometry of the scene can be further used to visually augment additional information into video from the endoscopic camera. It is interesting to apply the augmented-reality approach to laparoscopic training and examine it's usability in the Operating Rooms. The presented system can serve as a basis for such augmented-reality trainer.

References

1. Hansen, J.: Laparoscopic versus open appendectomy: prospective randomized trial. World J. Surg. 20(1), 17–21 (1999)
2. Breasted, J.: The Edwin Smith surgical papyrus. University of Chicago Press, Chicago (1991)
3. Gorman, P., Meier, A., Krummel, T.: Computer-assisted training and learning in surgery. Comput Aided Surg. 5, 120–130 (2000)
4. Cosman, P., Cregan, P., Martin, C., Cartmill, J.: Virtual reality simulators: current status in acquisition and assessment of surgical skills. ANZ J. Surg. 72, 30–34 (2002)
5. Kneebone, R.: Simulation in surgical training: educational issues and practical implications. Med. Educ. 37, 267–277 (2003)
6. Seymour, N., Gallagher, A., Roman, S., OBrien, M., Bansal, V., Andersen, D., Satava, R.: Virtual reality training improves operating room performance: results of a randomized, double-blinded study. Ann. Surg. 236, 458–463 (2002)
7. Grantcharov, T., Kristiansen, V., Bendix, J., Bardram, L., Rosenberg, J., Funch-Jensen, P.: Randomized clinical trial of virtual reality simulation for laparoscopic skills training. Brit. J. Surg. 91, 146–150 (2004)
8. Semm, K.: Pelvi-trainer, a training device in operative pelviscopy for teaching endoscopic ligation and suture technics. Geburtshilfe Frauenheilkund 46, 60–62 (1986)
9. Kuhnapfel, U., Akmak, H., Maaß, H.: 3D Modeling for endoscopic surgery. Proc. IEEE SIMSYS, 22–32 (1999)
10. Kuhnapfel, U., Krumm, H., Kuhn, C., Hubner, M., Neisius, B.: Endosurgery simulations with KISMET: a flexible tool for surgical instrument design, operation room planning and VR technology based abdominal surgery training. Proc. IEEE VR 95, 165–171 (1995)
11. Fraser, S., Klassen, D., Feldman, L., Ghitulescu, G., Stanbridge, D., Fried, G.: Evaluating laparoscopic skills. Surg. Laparosc. Endosc. 17, 964–967 (2003)
12. Simbionix Web Site Laparoscopy Simulators, `http://www.simbionix.com` (accessed February 22, 2007)
13. Feng, C., Rozenblit, J., Hamilton, A.: A hybrid view in a laparoscopic surgery training system. Proc. IEEE ECBS, 339–348 (2007)
14. Derossis, A., Fried, G., Abrahamowicz, M., Sigman, H., Barkun, J., Meakins, J.: Development of a model for training and evaluation of laparoscopic skills. Am. J. Surg. 175, 482–489 (1998)
15. Fried, G., Feldman, L., Vassiliou, V., Fraser, S., Stanbridge, D., Ghitulescu, G., Andrew, C.: Proving the value of simulation in laparoscopic surgery. Ann. Surg. 240, 518–528 (2004)
16. Horn, B., Schunck, B.: Determining Optical Flow. Artificial Intelligence 17, 185–203 (1981)
17. Ullman, S., Hildreth, E.: The measurement of visual motion. In: Braddick, O., Sleigh, A. (eds.) Physical and biological processing of images. Springer, Berlin (1983)
18. Pollefeys, M., Van Gool, L., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J., Koch, R.: Visual modeling with a hand-held camera. Intern. J. Computer Vision 59, 207–232 (2004)
19. Hartley, R., Zisserman, A.: Multiple view geometry in computer vision. Cambridge University Press, Cambridge (2003)

20. Harris, C., Stephens, M.: A combined corner and edge detector. In: Proc. Alvey Vision Conference, pp. 189–192 (1988)
21. Torr, P., Fitzgibbon, A., Zisserman, A.: The problem of degeneracy in structure and motion recovery from uncalibrated Image Sequences. Intern. J. Comput. Vision 32, 27–44 (1999)
22. Fischler, M., Bolles, R.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM 24, 381–395 (1981)
23. Delaunay, B.: Sur la sphere vide. Izv AN SSSR. Otd Mat. Estest Nauk 7, 793–800 (1934)
24. Wytyczak-Partyka, A., Nikodem, J., Klempous, R., Rozenblit, J., Feng, C.: Computer-guided laparoscopic training with application of a fuzzy expert system. In: Gelbukh, A., Morales, E.F. (eds.) MICAI 2008. LNCS (LNAI), vol. 5317, pp. 965–972. Springer, Heidelberg (2008)

Kinematics Analysis and Design of 3 DOF Medical Parallel Robots

S.D. Stan[1], M. Manic[2], V. Maties[1], R. Balan[1], and D. Verdes[1]

[1] Technical University of Cluj-Napoca, B-dul Muncii, Romania
sergiustan@ieee.org,{matiesvistrian,radubalan}@yahoo.com,
verdes.dan@gmail.com
[2] University of Idahol Idaho Falls, Dept of CS, Center for Higher Education, Idaho Falls
misko@uidaho.com

Abstract. Robots are one of the most representative fields in mechatronics, this article presents the development process of a mechatronic system, taking into account 3 DOF parallel robots, pointing out the integration of the elements and showing the challenges needed to be faced for the interfaces between the robot modules and the GUI (Graphical User Interface). The interface uses virtual reality to provide the user with an interactive 3D graphical representation of the parallel robot.

1 Introduction

The integration of this disciplines leads to more competitive products, new technical solutions can be achieved from a mechatronic approach solutions like: new technical functions, increased flexibility, extension of the range of parameters used for machine control, reduce size and manufacturing costs due to physical integration.

A mechatronic product is a very complex system with several components, technologies and function all interrelated and interdependent. The technical synergy of a mechatronics system creates critical dependencies between involved engineering disciplines, these dependences are demonstrated in many ways, mechanical properties may for example be strongly linked to the control system characteristic that in turn are strongly linked to software properties and the vice versa.

In the design process of the robot one important aspect that was kept in mind was the modularity of the final product, building a modular robot bring several advantages like reconfiguration, the possibility to improve just one of the module without interceding in the entire structure, the possibility to use robot modules in building other robots or other structures. But beside this advantage a modular product brings new challenge in building interfaces between all the modules that are used for the robot.

Parallel robots have a number of advantages over the traditional serial robots due to their particular architecture [1]. There are several examples of parallel robots, especially in the fields of assembly and medical applications.

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 395–406.
springrlink.com

The paper is organized as follows. The conceptual designs of the medical robots system are proposed in Section II. The kinematics and analysis is carried out in Section III, where the reachable workspace of the robots is generated. Section IV is focused on the virtual reality model of the robot architectures using the Virtual Reality toolbox from MATLAB/Simulink. The design of the 3 DOF parallel robots is presented in Section IV. Finally, this paper concludes with a discussion of future research considerations in Section V.

2 Kinematics Analysis for 3 DOF Parallel Robots

Robot kinematics deals with the study of the robot motion as constrained by the geometry of the links. Typically, the study of the robot kinematics is divided into two parts, inverse kinematics and forward (or direct) kinematics.

2.1 Mathematical Model

To analyze the kinematic model of the parallel robots, two relative coordinate frames are assigned, as shown in Fig. 1.

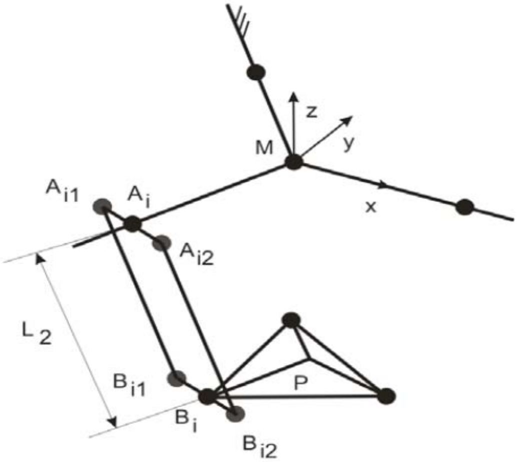


Fig. 1. Schematic diagram of TRIGLIDE parallel robot

A static Cartesian coordinate frame XYZ is fixed at the center of the base, while a mobile Cartesian coordinate frame $X_PY_PZ_P$ is assigned to the center of the mobile platform.

A*i, i* = 1, 2, 3, and B*i, i* = 1, 2, 3, are: the joints located at the center of the base, as presented in Fig. 2, and the platform passive joints, respectively. A middle link L_2 is installed between the mobile and fixed platform.

Let L_1, L_2, L_3 be the link's lengths as expressed in (1):

$$\begin{gathered} L_1 = A_{i1}A_{i2} = B_{i1}B_{i2} \\ L_2 = A_{i1}B_{i1} = A_iB_i \\ L_3 = B_iP \end{gathered} \tag{1}$$

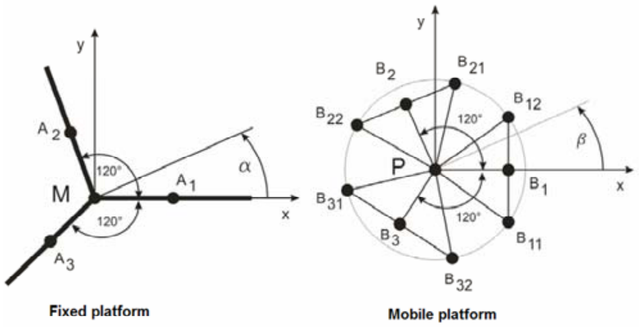


Fig. 2. Schematic diagram of mobile and fixed platform for TRIGLIDE parallel robot

In order to compute $r_{A_iB_i} = r_{B_i} - r_{A_i}$ for i=1, 2, 3, r_{B_i} and r_{A_i} need to be computed first. First, r_{B_i} is defined as:

$$r_{B_i} = r_P + r_{PB_i} = \begin{pmatrix} x_P + L_3 \cdot \cos\,\beta_i \\ y_P + L_3 \cdot \sin\,\beta_i \\ z_P \end{pmatrix} \tag{2}$$

where β_i is computed as $\beta_i = (i-1)\cdot 120\,°$.

Then, r_{A_i} is calculated as:

$$r_{A_i} = \begin{pmatrix} q_i \cdot \cos\,\alpha_i \\ q_i \cdot \sin\,\alpha_i \\ 0 \end{pmatrix} \tag{3}$$

where α_i is computes as $\alpha_i = (i-1)\cdot 120°$. From (3) yields f_i:

$$f_i = \begin{pmatrix} x_p + L_3 \cdot \cos\,\beta_i - q_i \cdot \cos\,\alpha_i \\ y_p + L_3 \cdot \cos\,\beta_i - q_i \cdot \cos\,\alpha_i \\ z_p \end{pmatrix}^2 - L_2^2 = 0; \tag{4}$$

$$L_2 = \left| r_{A_iB_i} \right|$$

$$\begin{aligned}
&(x_p + L_3 \cdot \cos \beta_i)^2 + q_i^2 \cdot \cos^2 \alpha_i \\
&+ 2 \cdot (x_p + L_3 \cos \beta_i) \cdot q_i \cdot \cos \alpha_i + (y_p + L_3 \cdot \sin \beta_i)^2 \\
&+ q_i^2 \cdot \sin^2 \cdot \alpha_i + 2 \cdot (y_p + L_3 \sin \beta_i) \cdot q_i \cdot \sin \alpha_i \\
&+ z_p^2 - L_2^2 =)
\end{aligned} \tag{5}$$

From (4) we obtain (5) and by reformulating (5), (6) is obtained:

$$\begin{aligned}
&q_i^2 + q_i \cdot 2 \cdot [(x_p + L_3 \cdot \cos \beta_i) \cdot \cos \alpha_i + \\
&(y_p + L_3 \cdot \sin \beta_i) \cdot \sin \alpha_i + (x_p + L_3 \cdot \cos \beta_i)^2 \\
&+ (x_p + L_3 \cdot \sin \beta_i)^2 + z_p^2 - L_2^2 = 0
\end{aligned} \tag{6}$$

By substituting (7):

$$\begin{aligned}
u_i &= (x_p + L_3 \cdot \cos \beta_i) \cdot \cos \alpha_i + (y_p + L_3 \cdot \sin \beta_i) \cdot \sin \alpha_i \\
v_i &= (x_p + L_3 \cdot \cos \beta_i)^2 + (y_p + L_3 \cdot \sin \beta_i)^2 + z_p^2 - L_2^2
\end{aligned} \tag{7}$$

in (6), we obtain the inverse kinematics problem of the TRIGLIDE parallel robot from Fig. 1:

$$q_i = u_i \pm \sqrt{u_i^2 - v_i} \tag{8}$$

For the implementation and resolution of forward and inverse kinematic problems of a parallel robot, a MATLAB environment was chosen. This is where a user friendly graphical user interface was developed, as well.

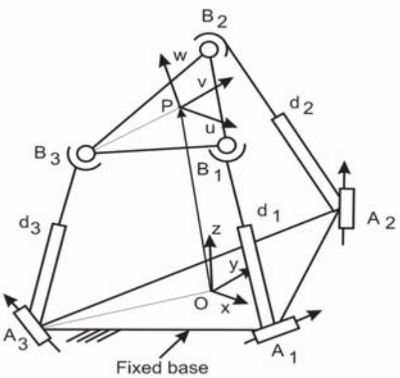


Fig. 3. 3-R$\underline{\text{P}}$S parallel robot with linear actuators

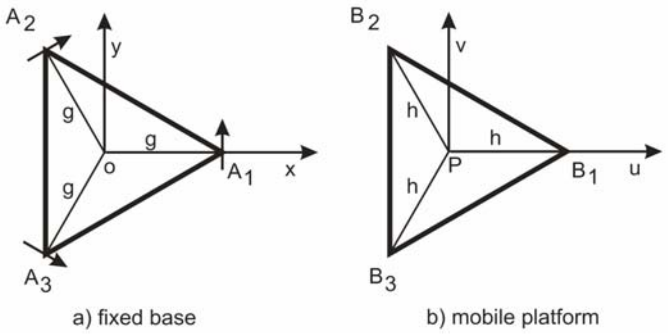
Fig. 3 shows a spatial parallel robot, 3-DOF, 3-R$\underline{P}$S type of parallel robot. It consists of three identical links that connect the moving platform at points B_i by spherical joints to the fixed base at points A_i, by revolute joints.

Each link consists of an upper and a lower member connected by a prismatic joint.

These three prismatic joints are used as inputs for the parallel robot. Overall, there are eight links, three revolute joints, three prismatic joints and three spherical joints. Thus, the degree of freedom of the parallel robot can be computed with:

$$F = \lambda(n-j-1)+\sum_i f_i = 6(8-9-1)+(3+3+9)=3 \tag{9}$$

For the kinematic analysis, two Cartesian coordinate systems A(x, y, z) and B(u, v, w) are attached to the fixed base and moving platform, respectively, as shown in Fig. 4.



a) fixed base b) mobile platform

Fig. 4. Top views of the 3-R$\underline{P}$S parallel robot

The following assumptions are made. Points A_1, A_2, A_3 lie on the xy-plane and B_1, B_2 and B_3 lie on the uv-plane.

As shown in Fig. 3, the origin O of the fixed coordinate system is located at the centroid of $\Delta A_1A_2A_3$ and the x-axis points in the direction of $\overline{OA_1}$.

Similarly, the origin P of the moving coordinate system is located at the centroid of $\Delta B_1B_2B_3$ and the u-axis in the direction of $\overline{PB_1}$.

Both $\Delta A_1A_2A_3$ and $\Delta B_1B_2B_3$ are equilateral triangles, having the following feature $|OA_1| = |OA_2| = |OA_3| = g$ and $|PB_1| = |PB_2| = |PB_3| = h$.

Furthermore, the axis of each revolute joint, J_i, lies on the x-y plane and is perpendicular to the vector $\overline{OA_1}$.

The transformation from the moving platform to the fixed base can be described by a position vector $p = \overline{OP}$ and a 3 × 3 rotation matrix ${}^A R_B$.

Let u, v, and w be three unit vectors defined along u, v, and w axes of the moving coordinate system B, respectively; then the rotation matrix can be expressed in terms of the direction cosines of u, v and w as:

$$^{A}R_{B}=\begin{bmatrix} u_x & v_x & w_x \\ u_y & v_y & w_y \\ u_z & v_z & w_z \end{bmatrix}. \tag{10}$$

We note that the elements of $^{A}R_{B}$ must satisfy the following orthogonal conditions:

$$\begin{gathered} u_x^2+u_y^2+u_z^2=1 \\ v_x^2+v_y^2+v_z^2=1 \\ w_x^2+w_y^2+w_z^2=1 \\ u_xv_x+u_yv_y+u_zv_z=0 \\ u_xw_x+u_yw_y+u_zw_z=0 \\ v_xw_x+v_yw_y+v_zw_z=0 \end{gathered} \tag{11}$$

Let a_i and $^{B}b_i$ be the position vectors of points A_i and B_i, respectively. Then the coordinates of A_i and B_i are given by:

$$\begin{gathered} a_1=[g,\ 0,\ 0]^T \\ a_2=\left[-\frac{1}{2}g,\ \frac{\sqrt{3}}{2}g,\ 0\right]^T \\ a_3=\left[-\frac{1}{2}g,\ -\frac{\sqrt{3}}{2}g,\ 0\right]^T \\ {}^{B}b_1=[h,\ 0,\ 0]^T \\ {}^{B}b_2=\left[-\frac{1}{2}h,\ \frac{\sqrt{3}}{2}h,\ 0\right]^T \\ {}^{B}b_3=\left[-\frac{1}{2}h,\ -\frac{\sqrt{3}}{2}h,\ 0\right]^T \end{gathered} \tag{12}$$

The position vector q_i and B_i with respect to the fixed coordinate system is obtained by the following transformation:

$$q_i=p+{}^{A}R_{B}\,{}^{B}b_i \tag{13}$$

and

$$q_1=\begin{bmatrix} px+hu_x \\ py+hu_y \\ pz+hu_z \end{bmatrix} \tag{14}$$

$$q_2 = \begin{bmatrix} px - \frac{1}{2}hu_x + \frac{\sqrt{3}}{2}hv_x \\ py - \frac{1}{2}hu_y + \frac{\sqrt{3}}{2}hv_y \\ pz - \frac{1}{2}hu_z + \frac{\sqrt{3}}{2}hv_z \end{bmatrix} \tag{15}$$

$$q_3 = \begin{bmatrix} px - \frac{1}{2}hu_x - \frac{\sqrt{3}}{2}hv_x \\ py - \frac{1}{2}hu_y - \frac{\sqrt{3}}{2}hv_y \\ pz - \frac{1}{2}hu_z - \frac{\sqrt{3}}{2}hv_z \end{bmatrix} \tag{16}$$

For the implementation and resolution of forward and inverse kinematic problems of a parallel robot, a MATLAB environment was chosen. This is where a user friendly graphical user interface was developed, as well.

3 Virtual Reality Model

The user simply describes the geometric properties of the robot first. Then, in order to move any part of the robot through 3D input devices, the inverse kinematics is automatically calculated in real time. The interface was also designed to provide the user decision capabilities when problems such are singularities are encountered. More realistic renderings of bodies are possible, if the Virtual Reality Toolbox for Matlab is installed. Arbitrary virtual worlds can be designed with the Virtual Reality Modeling Language (VRML) and interfaced to the SimMechanics model.

The user simply describes the geometric properties of the robot and then inverse kinematics is automatically calculated in real time, in order to move any part of the robot through 3D input devices. The interface is also designed to provide the user decision capabilities when problems, like singularities, are encountered.

Inside this VR interface, the user can interact with the robot in an intuitive way. This means that the operator can pick any part of the robot and move it (in the general sense: translations and rotations), using 3D sensors, wherever he wants, as easily as a “drag and drop” program. Thus, trajectories can be defined, optimized and stored easily. The virtual world can be accessed also from Internet. We have found that the use of a VR interface to simulate robots drastically improves the “feeling” of the robot.

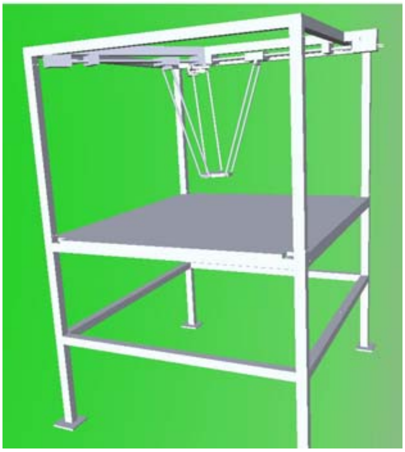


Fig. 5. TRIGLIDE Virtual Reality model made in Matlab/Simulink

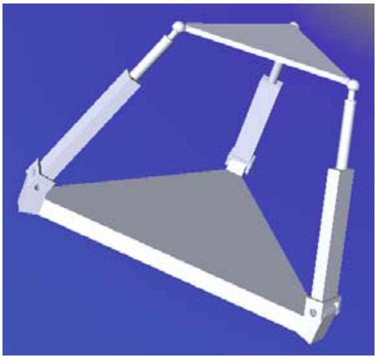


Fig. 6. 3-R$\underline{P}$S Virtual Reality model made in Matlab/Simulink

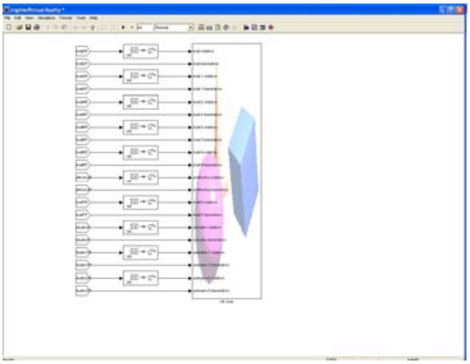


Fig. 7. Triglide Virtual Reality model made in Matlab/Simulink

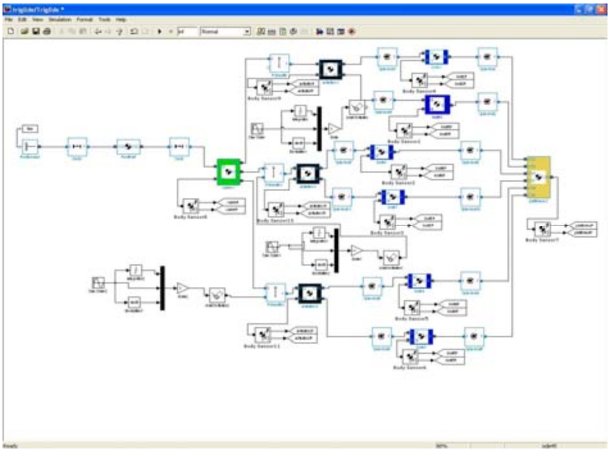
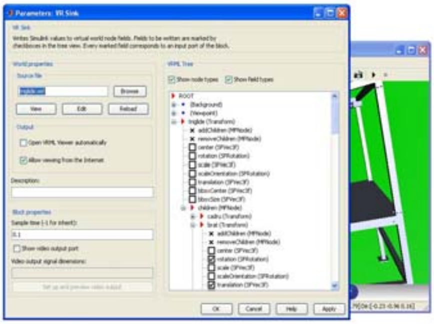


Fig. 8. Triglide Virtual Reality model made in Matlab/Simulink – dynamic model



Parameters: VR Sink

VR Sink

Writes Simulink values to virtual world node fields. Fields to be written are marked by checkboxes in the tree view. Every marked field corresponds to an input port of the block.

World properties

Source file: triglide.wrl

Output

☐ Open VRML Viewer automatically

☑ Allow viewing from the Internet

Description:

Block properties

Sample time (-1 for inherit): 0.1

☐ Show video output port

Video output signal dimensions:

VRML Tree

☑ Show node types ☑ Show field types

```
ROOT
  (Background)
  (Viewpoint)
  triglide (Transform)
    ☒ addChildren (MFNode)
    ☒ removeChildren (MFNode)
    ☐ center (SFVec3f)
    ☐ rotation (SFRotation)
    ☐ scale (SFVec3f)
    ☐ scaleOrientation (SFRotation)
    ☐ translation (SFVec3f)
    ☐ bboxCenter (SFVec3f)
    ☐ bboxSize (SFVec3f)
    children (MFNode)
      cadru (Transform)
      brat (Transform)
        ☒ addChildren (MFNode)
        ☒ removeChildren (MFNode)
        ☐ center (SFVec3f)
        ☑ rotation (SFRotation)
        ☐ scale (SFVec3f)
        ☐ scaleOrientation (SFRotation)
        ☑ translation (SFVec3f)
```

Fig. 9. Triglide Virtual Reality model made in Matlab/Simulink

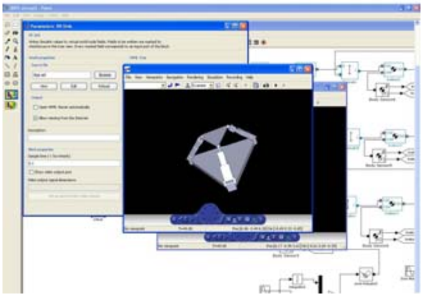


Fig. 10. 3-RPS Virtual Reality model made in Matlab/Simulink – dynamic model

Standard Simulink/SimMechanics blocks have been used to create the Isoglide3 robot modules, whose internal structure can be easily modified by the user, and

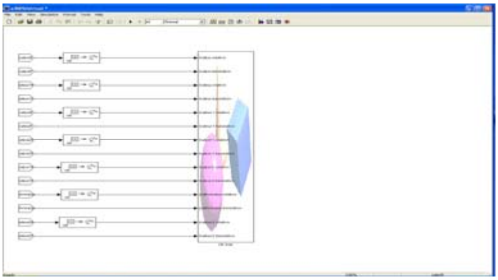


Fig. 11. 3-RPS Virtual Reality model made in Matlab/Simulink

tailored to the requirements of particular applications, should such a need arise. Furthermore, all the available MATLAB/Simulink add-ins (toolboxes, coding options, etc.) can be used within this framework to implement additional features. SimMechanics is capable of modelling a large number of DOF and CAD models exported from, for example, SolidWorks can be imported into SimMechanics providing a relatively straightforward solution to simulate complex 3-D multi-body rover designs.

In particular, the interface allows user to understand the behavior of an existing robot, and to investigate the performance of a newly designed structures without the need and the cost associated with the hardware implementation.

4 Design of TRIGLIDE Parallel Robot

In the followings is presented the design of the 3 DOF parallel robot of type Triglide (Fig. 12). The robot was realized at Department of Mechatronics, Technical University of Cluj-Napoca.

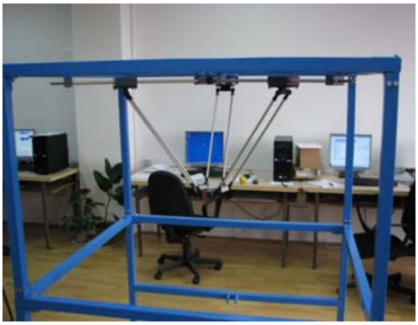


Fig. 12. 3-RPS Virtual reality model made in Matlab/Simulink

5 Conclusion

Parallel robots such in human-systems interaction such as medical, industry depend on accuracy, robustness, precision and dynamic workspace. In the paper was presented the control in Virtual Reality environment of the 3 DOF TRIGLIDE & 3-R$\underline{P}$S parallel robots. For the simulation, we used an evaluation model from the Matlab/SimMechanics. It was proposed an interactive tool for dynamics system modelling and analysis, exemplified on the control in Virtual Reality environment for Isoglide3 parallel robot. The main advantages of this parallel manipulator are that all of the actuators can be attached directly to the base, closed-form solutions are available for the forward and inverse kinematics, and the moving platform maintains the same orientation throughout the entire workspace. By means of SimMechanics we considered robotic system as a block of functional diagrams. Besides, such software packages allow visualizing the motion of mechanical system in 3D virtual space. Especially non-experts will benefit from the proposed visualization tools, as they facilitate the modelling and the interpretation of results.

Acknowledgment

This work was financial supported by CNMP through grants no. 3280 (PARTENERIATE type), title of the project ‘Complex mechatronics systems for medical applications’.

References

1. Gosselin, C.: Determination of the workspace of 6-d.o.f. parallel manipulators. ASME J. Mech. Design 112, 331–336 (1990)
2. Merlet, J.P.: Determination of the orientation workspace of parallel manipulators. J. Intelligent and Robotic Systems 13, 143–160 (1995)
3. Kumar, A., Waldron, K.J.: The workspace of mechanical manipulators. ASME J. Mech. Design 103, 665–672 (1981)
4. Tsai, Y.C., Soni, A.H.: Accessible region and synthesis of robot arm. ASME J. Mech. Design 103, 803–811 (1981)
5. Agrawal, S.K.: Workspace boundaries of in-parallel manipulator systems. International J. Robotics and Automation 6(3), 281–290 (1990)
6. Gosselin, C., Angeles, J.: Singularities analysis of closed loop kinematic chains. IEEE Transactions on Robotics and Automation 6(3), 281–290 (1990)
7. Cecarelli, M.: A synthesis algorithm for three-revolute manipulators by using an algebraic formulation of workspace boundary. ASME J. Mech. Design 117(2(A), 298–302 (1995)
8. Pernkopf, F., Husty, M.: Reachable workspace and manufacturing errors of Stewart-Gough manipulators. In: Proc. Intern. Symp. on Multibody Systems and Mechatronics, Brazil, pp. 293–304 (2005)
9. Stan, S., Diplomarbeit: Analyse und ptimierung der strukturellen abmessungen von werkzeugmaschinen mit parallelstruktur. IWF-TU Braunschweig, Germany (2003)
10. Cleary, K., Arai, T.: A prototype parallel manipulator: Kinematics, construction, software, workspace results, and singularity analysis. In: Proc. International Conference on Robotics and Automation, Sacramento, California, pp. 566–571 (1991)

11. Du Plessis, L.J., Snyman, J.A.: A numerical method for the determination of dextrous workspaces of Gough-Stewart platforms. Methods in Engineering 52, 345–369 (2001)
12. Liu, X.J.: Optimal kinematic design of a three translational DOF parallel manipulator. Robotica 24(2), 239–250 (2006)
13. Stan, S.D., Manic, M., Mătieş, M., Bălan, R.: Evolutionary approach to optimal design of 3 DOF translation exoskeleton and medical parallel robots. In: Proc. IEEE Conference on Human System Interaction, Krakow, Poland, pp. 720–725 (2008)
14. Stan, S.D., Manic, M., Mătieş, M., Bălan, R.: Kinematics analysis, design, and control of an Isoglide3 Parallel Robot (IG3PR). In: Proc. 34th Annual Conference of the IEEE Industrial Electronics Society, Orlando, USA, pp. 1265–1275 (2008)
15. Haupt, R., Haupt, S.E.: Practical genetic algorithms. Willey-IEEE, New Jersey (2004)

Graphical Human-Machine Interface for QB Systems

M. Jasiński and A. Nawrat

Silesian University of Technology,
Department of Automatic Control and Robotics, Gliwice
anawrat@polsl.pl

Abstract. Graphical human-machine interface project is designed to provide algorithms and graphical interface to navigate autonomous objects due to simply creating its trajectory from list of waypoints defined in a "drag-and-drop" technology. Trajectory is generated dynamically and it contains of basic geometry – straight lines and arcs – which keeps this solution easy and fast for calculation. As position of objects is presented in Cartesian coordinate system, presented solution can be easy implemented when it is difficult to obtain GPS data or this data are not accurate. To enable autonomous control in the real object PID algorithm were implemented. Data for PID algorithm were get from Crossbow Stargate™, which were also used as a servo-controller installed on the testing platform.

1 Introduction

Nowadays many solutions on both military and civilian ground are based on UAV (Unmanned Aerial Vehicle), which makes this technology became more and more popular.

This solution enable exploitation of flying machines when it is too risky to use machines controlled directly by human (war patrols or firefighting) or when remote control are too expensive (for example as a moving signal-transmitting stations). This solution provides opportunity to examine wheatear conditions (by sending UAV to the center of the thunder to collect data) or polar glacial.

As UAV objects became more popular problem of controlling this objects became very essential. In the literature two main attitudes could be found:

- Control based on data collected from on-board sensors [5,6];
- Control based on flight program defined by operator before start [9];

Common technique for building trajectory for UAV object is to describe it with some points (so called "waypoints"). Each waypoint is defined as a point placed in the objects move range and it should lay on the objects move path.

Main task of Trajectory Generator project is to develop algorithms and human interface for calculating and generating trajectory for autonomous unmanned vehicle.

Trajectory, in generally, could be defined as the path a moving object follows thought space. In this work trajectory is defined as path of autonomous object, which cross each waypoint defined by operator.

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 407–417.
springrlink.com

2 Trajectory Planer and Trajectory Generator

UAV Controller is an application written in Java language [1]. Application can be utilized to generate and simulate trajectory for various objects using specified algorithms. Choosing Java makes this solution platform independent and enables exploitation of large number of free API and libraries available for this language.

JFlightSimulator enables generation of trajectory for UAV object using coordinate system connected with start position of object. So there is no need for checking object position all the time while its moving – only knowing start position in geographic coordinate system is required. Positions of next points are calculated dynamically during movement.

Trajectory Generator Project provides interface to translate list of waypoints in geographic coordinate system to the list of commands that can be exploited to generate object trajectory dynamically in semi-intelligent way based on simple model [7, 8, 12].

2.1 Parts of the Application

UAV Controller application consist of several parts, which can be used for:

- Viewing and editing map of specified terrain;
- Build a trajectory for an objects. Trajectory is described by series of waypoints placed on map;
- Trajectory generation of specified flight path;
- Read the GPS data from GPS receiver mounted on object;

2.1.1 Waypoint Editor

A Waypoint Editor is part of program which enables a user to define the path by creating it from waypoints.

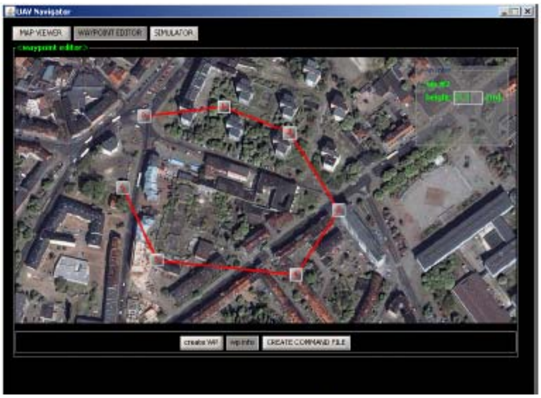


Fig. 1. Trajectory planer witch example fly path and a map

To generate trajectory, map of region of our interest must be loaded (all maps comes from Google Earth program [2]). When we know accurate coordinates from lower-left and upper-right corners of the map we can easily calculate coordinates of each waypoint on the map (utilizing for example UTM coordinate system [3]).

Results of Trajectory Generator program can be changed in text-table mode and saved as *.dat or *.flight file. These files can be later exploited in the Flight Simulator for path generation.

Saving data in *.dat file enables making a text file with waypoints coordinates, then *.flight file options allows to generate command file for the Flight Simulator program which is independent on the coordinate system exploited for the map.

Here is an example of *.dat file:

```
1,140.0,249.0,0.0,
2,165.0,189.0,0.0,
3,202.0,172.0,0.0,
4,262.0,195.0,0.0,
5,277.0,224.0,0.0,
6,273.0,269.0,0.0,
```

First number is PointID, then x, y and z coordinate.

Below is example of *.flight file generated from previous coordinates:

```
start 22
flyto 25 60
flyto 62 77
flyto 122 54
flyto 137 25
land 229
```

Files saved in *.flight format exploit Cartesian coordinate system instead of geographic coordinate system. Start point is located at (0,0) and all waypoints coordinates are translated to it.

2.1.2 Available Trajectory Presentation

During simulation of non-flying object 2D coordinate system is good enough but when we want to simulate flying object we need represent it in its natural coordinate system - 3D.

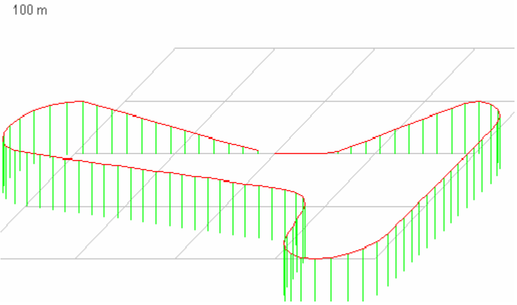


Fig. 2. Perspective view example

Flight Simulator allows us to switch between 3D [Fig. 3] and 3D [Fig. 2] coordinate system during simulation so we can compare generated trajectory from a flat view or perspective view.

At this point Flight Simulator doesn't support a load map option in 3D view.

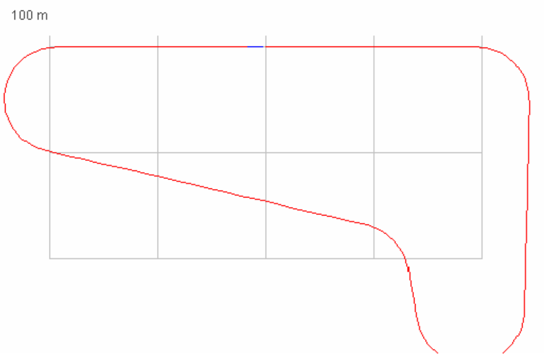
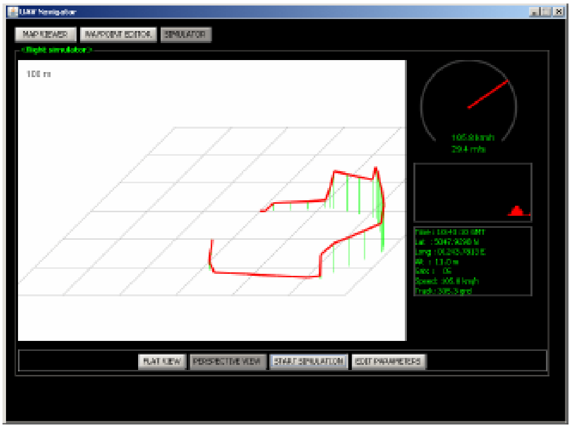


Fig. 3. Flat view example

2.1.3 Flight Simulator

Flight Simulator [Fig. 4] is a part of program where previously created *.flight file is exploited in order to generate trajectory of an object dynamically. Additional information about flight parameters are indicated on the right side of map window.



105.8 km/h
29.4 m/s

Time : 10:43:33 GMT
Lat : 5047.9290 N
Lng : 01243.7913 E
Alt : 11.0 m
Sats : 05
Speed : 105.8 km/h
Track : 305.3 grd

Fig. 4. Trajectory generator with example fly path generated

All constant parameters required for simulation can be freely modified in the configuration panel [Access to these values gives opportunity to change object dynamics and, consequently, type of simulated object. For example setting Ascent and Descent Angle to values about 20 and Curve Radius to 50 gives us simulation for an aircraft. But when we change Curve Radius to values about 5 and Ascent and Descent Angle to values about 80 we receive a simulation for helicopter. As we can set latitude to negative values we can simulate even submarines.

3 Trajectory Generation

Generation of specified fly path is calculated dynamically due to control commands generated from a series of waypoints. Each command is used to calculate specified destination point (represented by next waypoint from the waypoint list) and line (or curve) connecting actual waypoint with a following one (and change altitude if needed).

In this project we decided not to exploit all available commands (described in "Command list" section) but "start", "stop", "ascent", "descent" and "flyto" only. Main reason for this solution is to keep simulation and command file simple for testing purpose. While "start", "stop", "ascent" and "descent" commands will stay the same in final solution, the "flyto" command could be changed and split out to "heading" (on given direction) and "straight" (for given distance) depending on further analyses of simulation procedure.

When procedure of trajectory generations [Fig. 5] starts, engine checks the command file for the flight commands and creates its representation in the memory. Then it is necessary to get any command separately and call action connected with specified command. When action connected with actual command is completed the control is set back to the simulation engine which checks command stack for the next command. Algorithm continues until the command stack is empty.

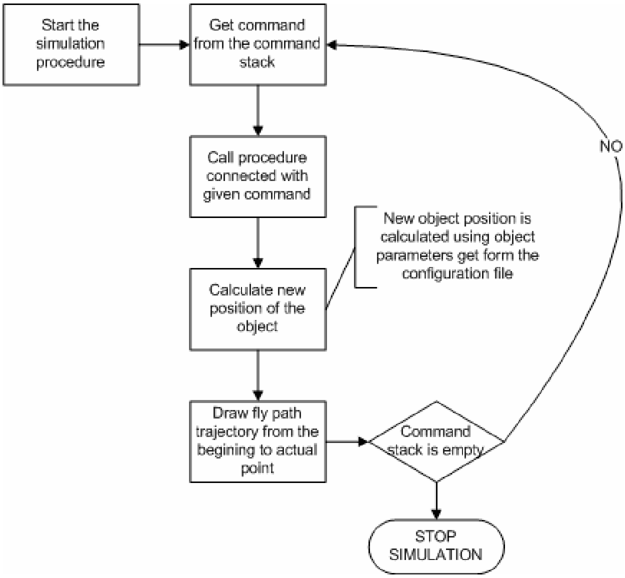

Start the simulation procedure
Get command from the command stack
Call procedure connected with given command
New object position is calculated using object parameters get form the configuration file
Calculate new position of the object
Draw fly path trajectory from the begining to actual point
Command stack is empty
NO
STOP SIMULATION

Fig. 5. Scheme of simulation procedure

3.1 Procedure of Trajectory Generation

Every flight consists of the start procedure, then there comes a set of commands related to main flight procedure like turning or changing altitude and the final point is connected with a landing procedure. Position of object is calculated separately, step by step, by every command procedure.

3.2 Command List

Here is a full list of commands (with short description) which can be used in Flight Simulator.

- Start – start in the given direction, then accelerating until start speed is reached.
- Heading – at first this command detects turning direction and angle, then calculates centre of curve.
- Flyto - calculates the curve and straight to reach a given location.
- Straight – head straight (go ahead) until a given distance is reached.
- Ascent – ascent to the altitude given in the parameter.
- Descent – descent to the altitude given in the parameter.
- Land – informs simulator that this is end of simulation.
- Approach – this command generates sequence of actions prior to landing (it is assumed that the touchdown point is the same as the start point).

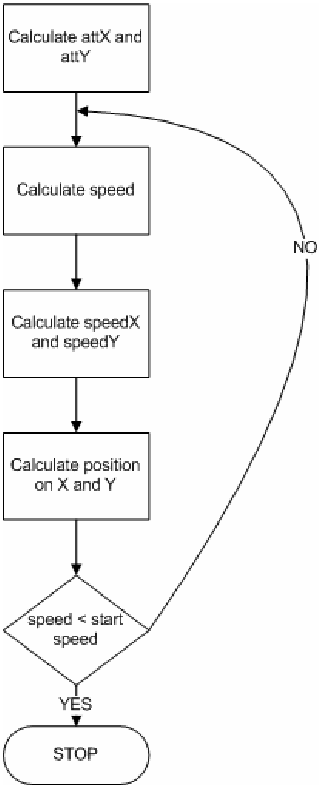


Fig. 6. Schematic of start algorithm

3.2.1 Start Command

During the start procedure [Fig. 6] speed of the object is growing until it reaches start speed specified in the object parameters.

Acceleration of simulated object and its start speed is taken from the object parameters.

At first, altitude on x and y should be calculated (X_{att} and Y_{att}). This parameters are needed to split speed vector into its x and y coordinates. Let vector z will be the speed vector. Then we can calculate X_{att} and Y_{att} as:

$$X_{att} = \frac{dx}{z} = \cos\alpha \tag{1}$$

$$Y_{att} = \frac{dy}{z} = \sin\alpha \tag{2}$$

Now speed and position of object can be calculated as:

$$v(t) = at \tag{3}$$

$$x(t) = x_0 + v_0 t + \frac{at^2}{2} \tag{4}$$

where:

- a – acceleration [m/s^2]
- t – time [s]
- x_0 – previous position
- v_0 – previous speed [m/s]

Knowing X_{att} and Y_{att} [and] it is possible to calculate speed on x and y as:

$$v_x = vX_{att} \tag{5}$$

$$v_y = vY_{att} \tag{6}$$

Actual position on x and y can be calculated as:

$$x_x = X_{att}\frac{at^2}{2} \tag{7}$$

$$x_y = Y_{att}\frac{at^2}{2} \tag{8}$$

3.2.2 FlyTo Command

FlyTo command [Fig. 7] is utilized to move from point (x_i, y_i) to the point (x_{i+1}, y_{i+1}). It calculates curve and straight line from start point to the end point utilizing curve radius get from the object parameters.

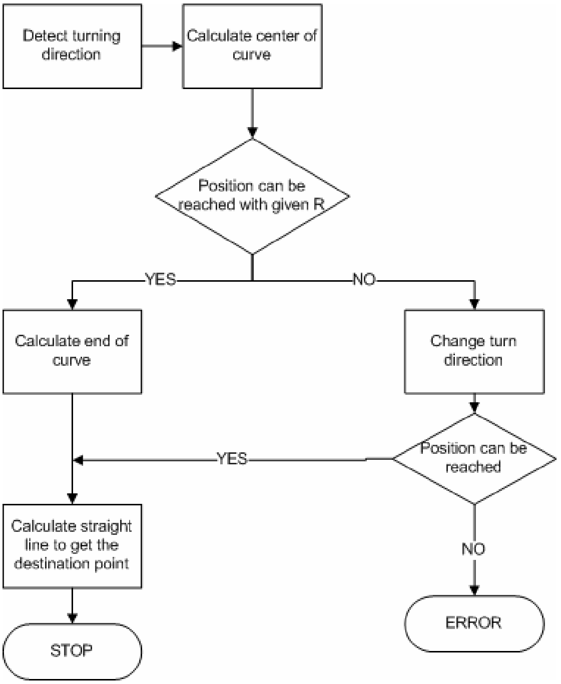

Detect turning direction
Calculate center of curve
Position can be reached with given R
YES
NO
Calculate end of curve
Change turn direction
YES
Position can be reached
NO
Calculate straight line to get the destination point
STOP
ERROR

Fig. 7. Schematic of FlyTo algorithm

First step of the algorithm is selection of turn direction (it calculates additional help parameter which says that turn should be to the left or to the right).

$$T_r = X_{att}\left(c_y - Y_{pos}\right) - T_{att}\left(c_x - X_{pos}\right) = \\ = \begin{cases} 1.0, & T_r > 0 \\ -1.0, & T_r \leq 0 \end{cases} \quad (9)$$

where:

- c_x, c_y – destination point (on x and y);
- X_{pos}, Y_{pos} – start position (on x and y);

Then center of the curve is calculated as:

$$m_x = X_{pos} - T_r R Y_{att} \quad (10)$$

$$m_y = Y_{pos} - T_r R X_{att} \quad (11)$$

where R is curve radius of the simulated object.

Now algorithm checks if a destination position can be reached by the object with specified curve radius. Additional help parameter q_r is calculated. If this parameter is negative it means that destination point is located too close to the start point and can't be reached with specified turn direction and curve radius – thus, turn direction (T$_r$ is changed to opposite).

If q$_r$ parameter is equal or greater than 0 algorithm calculates position of point B where curve ends and straight line that connects point B with specified destination point as following:

$$b_x = \frac{\left(c_x + qc_y + q_q m_x - qm_y\right)}{1 + q_q} \tag{12}$$

$$b_y = \left(c_y q b_x + qm_x\right) \tag{13}$$

where additional parameters are:

$$q_q = \frac{\left(c_x - m_x\right)^2 + \left(c_y - m_y\right)^2 - R^2}{R^2} \tag{14}$$

Simulation of turning process is just a simulation of moving on the edge of circle with a constant speed:

$$X_{pos} = m_x + R\cos\alpha \tag{15}$$

$$Y_{pos} = m_y + R\sin\alpha \tag{16}$$

where:

$$\alpha = \alpha_0 + T_r \omega t \tag{17}$$

ω – angular speed [1/s];

Formulas to simulate of flight directly forward are the same like in case of the start simulation so it is simulation of moving on the straight line with constant acceleration.

4 Test Results

Designed algorithm were tested on specially prepared platform. Test assumes that object is hanging over certain point and direction is controlled by human-navigator by changing destination point.

Test shows that designed algorithm is good enough to control UAV objects [Fig. 8]. Moreover both start and land maneuvers were successfully processed.

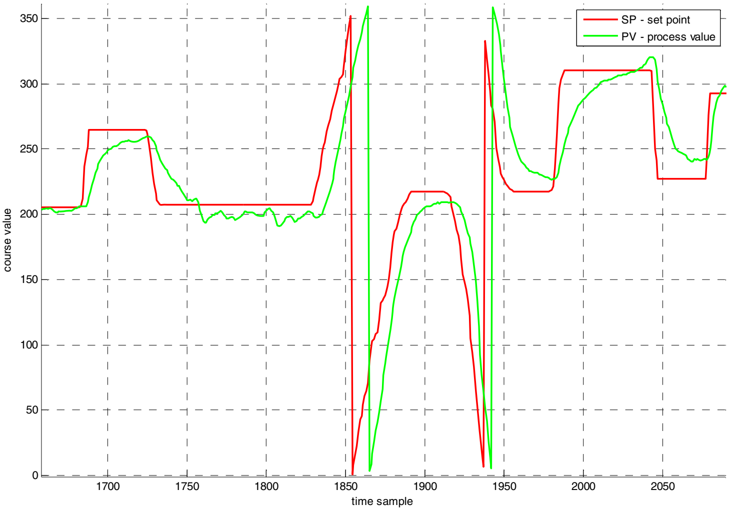


Fig. 8. Test results

5 Conclusions

Trajectory Generator Project provides algorithms to navigate UAV object by making trajectory from a set of waypoints. To navigate the object we only have to know its start position (in geographic coordinate system). All other positions are calculated in the Cartesian coordinate system connected with start point which makes this project easier to implement when position data collected form GPS is not accurate or difficult to obtain [10, 11].

Results of tests show that simulated trajectory is good enough to take further research to fully adopt project to control flying autonomous object [7, 8].

Full support of 3D view is very convenient to present trajectory of flying object - when we need information about x and y coordinates as well as z coordinate (altitude).

Future works over the project assume providing possibility of to switching between automatic and manual control in any time. It is also planned to implement autonomous control for main rotor moves, witch allows controlled object to follow by more complicated trajectory.

References

1. Java: platform independent, object oriented modern programming language developed by Sun Microsystems, `http://java.sun.com` (accessed April 27, 2009)
2. Google Maps is a free web mapping service and technology provided by Google that powers many map-based services including the Google Maps website, Google Ride Finder and embedded maps on third-hand websites via Google Maps API, `http://maps.google.com` (accessed April 27, 2009)

3. UTM: Universal Transverse Mercator is a grid-base method of specifying location on the surface of Earth
4. Proctor, A., Johnson, E.: Vision-only aircraft flight control methods and test results. In: Proc. AIAA Conference on Guidance, Navigation, and Control, Providence, Rhode Island (2004)
5. Al-Hasan, S., Vachtsevanos, G.: Intelligent route planning for fast autonomous vehicles operating in a large natural terrain. Journal of Robotics and Autonomous Systems 40, 1–24 (2002)
6. Dittrich, J., Johnson, E.: Multi-sensor navigation system for an autonomous helicopter. In: Digital Avionics Systems Conference (2002)
7. Johnson, E., Kannan, S.: Adaptive flight control for an autonomous unmanned helicopter. In: Proc. AIAA Conference on Guidance, Navigation, and Control (2002)
8. Kannan, S., Johnson, E.: Adaptive Trajectory-based flight control for autonomous helicopters. In: Proc. 21st Conference on Digital Avionics Systems (2002)
9. Nikolos, I.K., Zografos, E.S., Brintaki, A.N.: UAV path planning using evolutionary algorithms. Studies in computational intelligence, vol. 70, pp. 77–111. Springer, Heidelberg (2007)
10. Bevilacqua, R., Yakimenko, O., Romano, M.: On-line generation of quasi-optimal docking trajectories. In: Proc. 7th International Conference on Dynamics and Control of Systems and Structures in Space, Greenwich, London, England, pp. 203–218 (2006)
11. Shim, D.H., Chung, H., Sastry, S.: Conflict-free navigation in unknown urban environments. IEEE Robotics and Automation Magazine 13, 27–33 (2006)
12. Harbick, K., Montgomery, J.F., Sukhatme, G.S.: Planar spline trajectory following for an autonomous helicopter. Department of Computer Science University of Southern California (2001)

Visualization of Two Parameters in a Three-Dimensional Environment

L. Deligiannidis[1] and A. Sheth[2]

[1] Department of Computer Science, Wentworth Institute of Technology, Boston, MA
deligiannidisl@wit.edu
[2] Kno.e.sis Center, CSE Department, Wright State University, Dayton, OH
amit.sheth@wright.edu

Abstract. Visualization techniques and tools allow a user to make sense of enormous amount of data. Querying capabilities and direct manipulation techniques enable a user to filter out irrelevant data and focus only on information that could yield to a conclusion. Effective visualization techniques should enable a user or an analyst to get to the conclusion in a short time and with minimal training. We illustrate, via three different research projects, how to visualize two parameters where a user can get to a conclusion in a very short amount of time with minimal or no training. The two parameters could be the relation of documents and their importance, or spatial events and their timing, or even the ration between carbon dioxide emission levels and number of trees per country. To accomplish this, we visualize the data in a three dimensional environment on a regular computer display. For data manipulation, we use techniques familiar to novice computer users.

1 Introduction

Nowadays, with the enormous amount of data available online and on other data repositories, analysts face a challenging problem in trying to make sense out of all these data [1]. Analysts need tools to guide them and interact with the system to enrich the analysis with details and insight Such tools need to assist an analyst in visualizing complex relationships and identify new patterns of significance. After all, visual analytics through visualizations require extensive functionality in both the system and the interactive interface [2]. This can be characterized as an *art work* that consists of two major components: (i) the data transformation into a visual form that humans can easily understand and (ii) an interactive environment where the data can be explored and analyzed in order to carry out a better reasoning [3]. The data presentation (i.e. rendering component) should provide the user with the capabilities to explore her dataset, highlight important data features such as relationships, similarities and anomalies, query the dataset to narrow down the result, organize the result in her logical groups so that he can comprehend the data [4]. This is the result of transforming data into information and consequently into knowledge.

Static images have limited capacity to supply enough information for the user to conceive appropriate reasoning. Therefore, a visualization analytics environment

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 419–428.
springrlink.com

should provide an interface where the user can interact dynamically and in real time with the system. Furthermore, we believe that the interface should be easy to master because powerful tools that are difficult to master are normally abandoned by the user community. Human visual perception is fundamentally three dimensional and a support for natural human activities such as grabbing an object can make man's interactions with machine much more natural and hence effective.

In this paper we present our results of a technique that enables visualization of two parameters in a short amount of time. The two parameters are different of the three presented projects. An analyst can analyze the data to discover interesting relationships; connect the dots and derive knowledge. This is accomplished by presenting the data in a three dimensional environment where the user can manipulate the view-port to visualize the data from a different prospective.

2 Related Work

Linking tables with maps and images provides analysts with summary of information in the context of space and time. As an example, elegant interaction techniques and devices (the TouchTable) for manipulating, visualizing, and studying maps have been implemented by Applied Minds Inc (http://www.touch-table-.com/). Their latest tools can also build physically 3D terrains of maps. MapPoint and ESRI ArcView can display position of events on a 2D map.

By using maps we can visualize the position of events. Maps have been used for years to illustrate the space around us and provide a geographic understanding. With today's growth of geospatial information, detailed maps can be constructed on demand that can be enhanced with digital terrains and satellite imagery such as Google Maps (www.google.com/maphp) and Google Earth (earth.google.com).

Lifelines [5] and Microsoft (MS) Project display attributes of events over the single time dimension. Netmap (www.netmapanalytics.com), Visual Analytics (www.visualanalytics.com), and Analyst Notebook (www.i2inc.com) display events as a graph of objects with connections between them.

While research in GIS for event visualization is in the early stages, results seem promising for the 3D environments [6-8]. First results show how it is possible to visualize paths or track activities but there is no animation or support for visual analytics and most of the early work is concentrated in the "overview" aspect of visualization rather than the "detail" needed in many cases for visual analytics [9, 10]. A pioneering visualization system that displays information on a highly interactive 3D environment that consists of a 3D terrain is GeoTime [11-13]. In GeoTime, events are displayed onto a 3D terrain to assist an analyst in correlating events and geographical locations of activities. GeoTime demonstrates that a combined spatial and temporal display is possible, and can be an effective technique when applied to analysis of complex event sequences within a geographic context.

Effective presentation of data plays a crucial role in understanding the data and its relationships with other data because it helps the end-user analyze and comprehend the data. As a result, data is transformed into information and then into knowledge [14]. Efficient understanding of semantic information leads to more actionable and timely decision making. Thus, without an effective visualization tool, analysis and understanding of the results of semantic analytics techniques is difficult, ineffective, and at times, impossible.

The fundamental goal of visualization is to present, transform and convert data into a visual representation. As a result, humans, with their great visual pattern recognition skills, can comprehend data tremendously faster and more effectively through visualization than by reading the numerical or textual representation of the data [15]. Interfaces in 2D have been designed for visualization results of queries in dynamic and interactive environments (e.g. InfoCrystal [16]). Even though the textual representation of data is easily implemented, it fails to capture conceptual relationships. Three-dimensional (3D) interactive graphical interfaces are capable of presenting multiple views to the user to examine local detail while maintaining a global representation of the data (e.g. SemNet [17]).

3 Effective Visualization of Two Parameters

An effective visualization technique should not only transform data into a visual representation but also provide the mechanics to either query and filter out irrelevant data or interaction techniques so that the user can see the data from different prospective. If the visual representation is difficult to understand, the technique used is not effective. For a visualization technique to be effective, the end result, the visual component, should be easy to understand and require minimal training to use.

We present an effective technique where two parameters (relation of documents and their importance, spatial events and their timing, and the ration between carbon dioxide emission levels and number of trees per US state) can be visualized at the same time in a three dimensional environment. The technique requires minimal training and the visual component is easy to understand. Three independent research projects illustrate the effectiveness of our technique.

3.1 Semantic Analytics Visualization

Semantic Analytics Visualization (SAV) [6] is a tool for semantic analytics in 3D visualizations. SAV is capable of visualizing ontologies and meta-data including annotated web documents, images, and digital media such as audio and video clips in a synthetic three-dimensional semi-immersive environment. More importantly, SAV supports visual semantic analytics, whereby an analyst can interactively investigate complex relationships between heterogeneous information. The backend of SAV consists of a Semantic Analytics system that supports query processing and semantic association discovery. The user can select nodes in the scene and either play digital media, display images, or load annotated web documents. SAV can also display the ranking of web documents related documents. SAV supports dynamic specification of sub-queries of a given graph and displays the results based on ranking information, which enables the users to find, analyze and comprehend the information presented quickly and accurately. A snapshot of SAV is shown in Figure 1 as seen from the front.

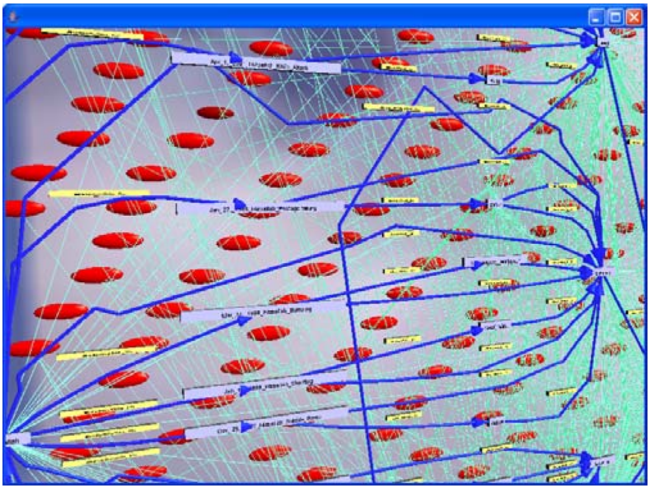


Fig. 1. Layout of instances and their relationships in the foreground. In the background, the document nodes are shown as red 3D ovals. The intensity of the red color applied on the document nodes, the documents' position and width depends upon the document ranking.

SAV partitions the 3D space into two volumes, the foreground and background. In the foreground it visualizes the ontology entities and their relationships, and in the background it visualizes the documents associated with each entity, as shown in Fig. 2.

The documents are represented as red spheres in the background. The position of a document changes depending on its ranking. The higher the ranking, or else

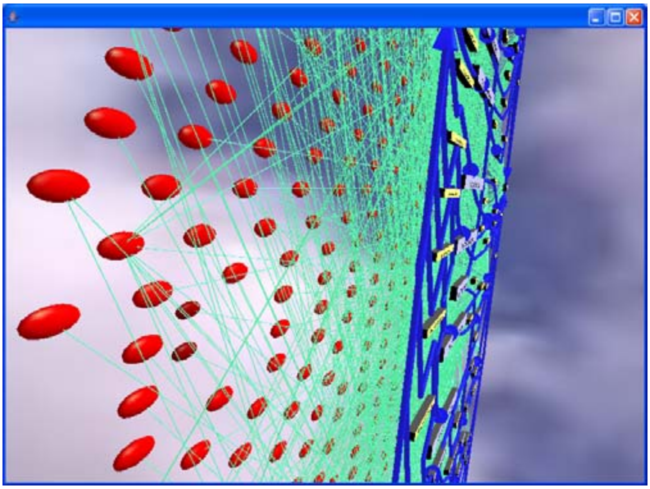


Fig. 2. Volume partitioning: visualization of entities and their relationships in the foreground. Documents associated with entities are visualized in the background.

the more important a document is, the closer to the foreground it is placed. The intensity of the red color also becomes higher when the ranking is higher (redder spheres indicate higher ranking). Additionally, depending on a document's ranking, the width of the sphere representing the document becomes bigger, in which case a sphere becomes a 3D oval shape (wider 3D ovals indicate higher ranking).

Since the most relevant documents are closer to the foreground, the user can select them more easily because they are closer to the user and the objects (the spheres) are bigger in size. The number of documents relative to a search could range from one hundred to several thousand. As a result, we visualized the entities and their relationships in the foreground while the document information stays in the background without cluttering the user's view. However, by rotating (using the mouse to drag the scene) the environment left or right, the user can see very easily which documents are important, less important or not so important; documents that are closer to the user are more important.

Redgraph, a virtual reality based tool, allows a use to interact with Semantic Web data in an immersive environment. Users can extrude nodes of interest from a 2D domain into the 3D space. This enables the users to quickly identify clusters and other interesting patterns in two dimensional structures [7]. User studies showed that the utilization of the third dimension improved the users' ability to answer fine-grained questions about the data-set, but no significant difference was found for answering broad questions concerning the overall structure of the data [18].

3.2 Semantic Event Tracker

Semantic Event Tracker (SET) [8] is an interactive visualization tool for analyzing events (activities) in a three-dimensional environment. In SET an event is modeled as an object that describes an action, its location, time, and relations to other objects. Real world event information is extracted from Internet sources, then stored and processed using Semantic Web technologies that enable us to discover semantic associations between events. SET is capable of visualizing as well as navigating through the event data in all three aspects of space, time and theme. Temporal data is illustrated as a 3D multi-line in the 3D environment that connects consecutive events. It provides access to multi-source, heterogeneous, multimedia data, and is capable of visualizing events that contain geographic and time information.

The visualization environment consists of a 2D geo-referenced map textured onto the surface of a 3D object. The users use the mouse to change the orientation of the map. We use the third dimension to visualize the time aspect of the events. The events are connected in space via lines to reveal the sequence of the events in time. There is an event at each end of a line segment. These events are visualized as small spheres and are selectable by the user. A user can select one of these events and then by issuing voice commands the metadata extracted from the ontologies can be presented. Examples include playing movie clips, showing digital images, playing audio clips, audio presentation of the date, time, casualties and other information associated with the selected event. Figure 3 shows a visualization of the events and how the sequence of events is visualized in space and time.

Three experiments were performed. The first was based on a web-browser interface using Google maps to visualize terrorist events. The second interface was a

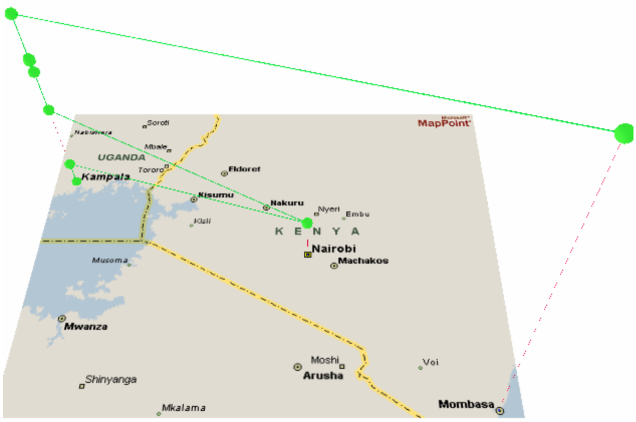


Fig. 3. Visualization of events on top of a map and their relationship with each other in the time domain

3D interface. The events were visualized on a 2D map but the third dimension was used to show the time domain (how far apart the events occurred). The third interface was a VR based interface. A user, using a data glove, could manipulate the data as well as the 3D scene. For the user studies, to compare the three interfaces, we used a between-subject experimental design using one-way ANOVA. Our independent variable was "interface type" (2D-browser based, 3D, and VR) and our dependent variable was performance (time to answer each question). The two questions the subjects were asked are:

- **Question 1 (Q1).** What specific pattern do you see in the geographic distribution of the locations of the events? More specifically, where most of the events occur?
- **Question 2 (Q2).** What specific pattern do you see in the temporal distribution of the events? More specifically, when most of the events occur?

ANOVA showed that the "Interface Type" is indeed a significant factor (F(2,18)=34.763, p<0.001). However, we did not find a significant difference for the first (temporal) question (F(2,18)=0.3654, p=0.699). Then we performed Tukey's *Honest Significant Difference* (HSD) method to create the confidence intervals. From this analysis we found that there is a significant difference between the 2D and VR interfaces (p<0.001, TukeyHSD) as well as between the 2D and 3D interfaces (p<0.001, TukeyHSD) at 95% confidence level. There is no significant difference between the 3D and VR interfaces (p=0.93, TukeyHSD). This increased performance in the 3D and VR interfaces, compared to the 2D interface, is based on the user's ability to rotate quickly the map and see the location, Figure 4 (left), and the distance in time, Figure 4 (right), between the events.

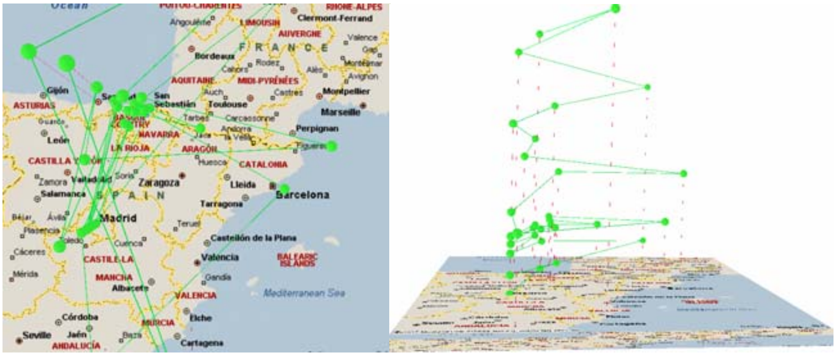


Fig. 4. Visualization of event location (left). Visualization of when the events occurred

In the 2D interface, the users had to click, explore, and remember the metadata information of the events so that they can mentally sort them based on time and figure out their concentration in a time period (when most of the events occurred). The 2D interface could be enhanced by adding a timeline that plots the sorted events in a separate window. This may enhance the users' ability to answer easily temporal questions such as the question Q2. On the other hand, the 3D view makes it easier to comprehend spatiotemporal relationships because there is no disconnect between the two views as opposed with the 2D map plus the timeline plot.

3.3 Health of US States

Recently, we received a sponsorship from The Climate Project (www.theclimate-project.org). For this project we wanted to visualize how each US state "behaves". The definition of this is that a well behaved state has low carbon dioxide emission levels as well as many trees per square mile to clean our environment. Since we wanted to visualize two parameters, a) carbon dioxide emission levels and b) number of trees per square mile, and the information we wanted to visualize was geographic in nature, we utilized Google Earth as the visualization platform. We encoded levels of carbon dioxide with colors ranging from green to red, for well behaved to not-well behaved respectively. For the second parameter, we encoded the number of trees per square mile as the height of 3D models of trees that were placed at the center of each state. The result of this visualization is shown in Figures 5 and 6.

Well behaved states are shown to have tall and green trees; tall for having many trees per square mile, and green for low carbon dioxide emission levels. On the other hand, badly behaved states are shown to have short trees with their leaves being red. This created visualizations with trees of different heights and a variation in their leaves' color. When looked from the top, one can visualize easily the color of the trees and can compare the different states to see which ones emit less carbon dioxide in the atmosphere. When the earth is tilted, one can visualize the heights of the trees and can compare the different states to see which ones have many trees per square mile. Going back and forth to visualize both parameters, is a simple operation, simply hold the earth in Google Earth and move the mouse up and down.

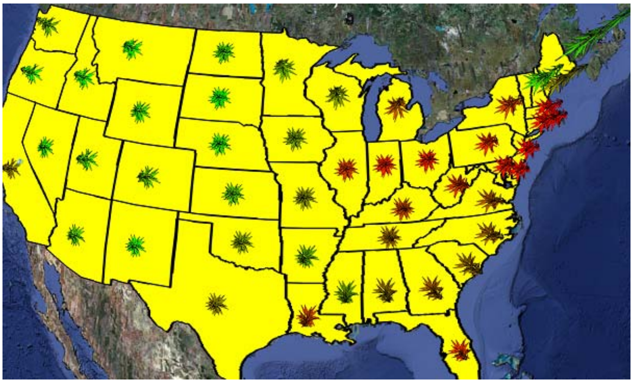


Fig. 5. Visualization of carbon dioxide emission levels per state

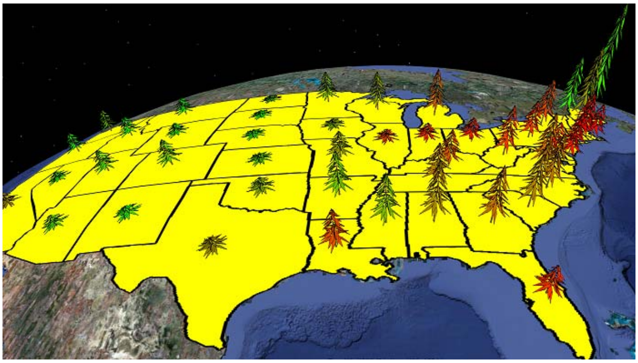


Fig. 6. Visualization of number of trees per square mile per state

4 Conclusions

We illustrated via three different research projects how we can use the third dimension effectively to visualize two parameters. The end result is easy to understand and requires no user training in using such a technique. Empirical results as well as measurable research results were presented that show the effectiveness of our technique. We showed that minimal manipulation is required on the visual display to easily "connect the dots" of the presentation.

Acknowledgments

This research is funded in part by the National Science Foundation (NSF) under Award No. 071441 titled *SemDis: Discovering Complex Relationships in the Semantic Web*. Special thanks to the *Tides Foundation* on recommendation of the Google.org Fund. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF, Google, and Tides Foundation.

References

1. Tomaszewski, B.M., Robinson, A.C., Weaver, C., et al.: Geovisual analytics and crisis management. In: Proc. 4th Intern. ISCRAM Conference, Delft, the Netherlands, pp. 173–179 (2007)
2. Slocum, T.A., McMaster, R.B., Kessler, F.C., et al.: Thematic cartography and geographic visualization, 2nd edn. Prentice-Hall, Upper Saddle River (2004)
3. Card, S.K., Mackinlay, J., Shneiderman, B.: Readings in information visualization using vision to think. Morgan Kaufmann, San Francisco (1999)
4. Shneiderman, B.: The eyes have it: A task by data type taxonomy for information visualizations. In: Proc. IEEE Symposium on Visual Languages, pp. 336–343. IEEE Computer Society, Washington (1996)
5. Plaisant, C., Milash, B., Rose, A., et al.: LifeLines: Visualizing personal histories (1996), http://portal.acm.org/citation.cfm?doid=238386.238493#citedby (accessed April 6, 2009)
6. Deligiannidis, L., Sheth, A.P., Aleman-Meza, B.: Semantic analytics visualization. In: Mehrotra, S., Zeng, D.D., Chen, H., Thuraisingham, B., Wang, F.-Y. (eds.) ISI 2006. LNCS, vol. 3975, pp. 48–59. Springer, Heidelberg (2006)
7. Halpin, H., Zielinski, D.J., Brady, R., Kelly, G.: Redgraph: Navigating semantic web networks using virtual reality. In: Proc. IEEE Conference on Virtual Reality, Reno, NV, pp. 257–258 (2008)
8. Deligiannidis, L., Hakimpour, F., Sheth, A.P.: Event visualization in a 3D environment. In: Proc. IEEE Conference on Human System Interaction, Cracow, Poland, pp. 158–164 (2008)
9. Mei-Po Kwan, M.P., Lee, J.: Geovisualization of human activity patterns using 3D GIS: A time-geographic approach. In: Goodchild, M.F., Janelle, D.G. (eds.) Spatially integrated social science, pp. 48–66. Oxford University Press, New York (2004)
10. Huisman, O., Forer, P.: The complexities of everyday life: Balancing practical and realistic approaches to modeling probable presence in space-time. In: Whigham, P.A. (ed.) The 17th Annual Colloquium of the Spatial Information Research Centre, New Zealand, pp. 155–167 (2005)
11. Shuping, D., Wright, W.: GeoTime visualization of RFID providing global visibility of the DoD supply chain (2005), http://oculusinfo.com/papers/GeoTime_RFID_Final_05_June.pdf (accessed April 6, 2009)
12. Kapler, T., Harper, R., Wright, W.: Correlating events with tracked movements in time and space: A GeoTime case study. In: Proc. Intelligence Analysis Conference, McLean, VA (2005), http://oculusinfo.com/papers/Oculus_GeoTime_TaxiCaseStudy_FinalDistrib.pdf (accessed March 20, 2006)
13. Kapler, T., Wright, W.: GeoTime information visualization. In: Proc. IEEE InfoVis Conference, pp. 25–32 (2004)
14. Sheth, A.P., Aleman-Meza, B., Arpinar, I.B., et al.: Semantic association identification and knowledge discovery for national security applications. J. Database Management 16(1), 33–53 (2005)

15. DeFanti, T.A., Brown, M.D., McCormick, B.H.: Visualization: expanding scientific and engineering research opportunities. Computer 22(8), 12–16, 22–25 (1989)
16. Anselm, S.: InfoCrystal: A visual tool for information retrieval. In: Proc. IEEE Visualization Conference, San Jose, CA, pp. 150–157 (1993)
17. Fairchild, K.M., Poltrock, S.E., Furnas, G.W.: SemNet: Three-dimensional graphic representation of large knowledge bases. In: Proc. Conference on Cognitive Science and its Application for Human-Computer Interface, pp. 201–233. Erlbaum, Hillsdale (1988)
18. Halpin, H., Zielinski, D., Brady, R., Kelly, G.: Exploring semantic social networks using virtual reality. In: Sheth, A.P., Staab, S., Dean, M., Paolucci, M., Maynard, D., Finin, T., Thirunarayan, K. (eds.) ISWC 2008. LNCS, vol. 5318, pp. 599–614. Springer, Heidelberg (2008)

Part V
Various H-CSI Applications

An Evaluation Tool for End-User Computing Competency in an Organizational Computing Environment

C.Y. Yoon[1], J.N. Jeon[1], and S.K. Hong[2]

[1] School of Electrical, Electronic & Computer Engineering, Chungbuk National University,
Cheongju city, Chungbuk, South Korea
{yoon0109,joongnam}@chungbuk.ac.kr
[2] Department of Industrial and Management Engineering, Chungju National University,
Chungju city, Chungbuk, South Korea
skhong@chungju.ac.kr

Abstract. The development and management of end-user computing capability is needed to efficiently do the given tasks in an organizational computing environment. This study presents a 16-item tool that can effectively evaluate end-user computing capability with an evaluation system, procedure and method. The utilization and application of the developed tool is confirmed by applying it to a case study and presenting its results.

1 Introduction

In order to perform end-users' given tasks, they use their computing systems in an enterprise. In this environment, their computing capability influences on the efficiency and performance of their tasks. Hence, we need a measure to efficiently assess the end-user computing ability and improve it for raising his or her business performance and the competitiveness of an enterprise. But studies on the evaluation of the end-user computing ability have not actively executed, and these focus on specific software skills, professional skills, and operational skills [1-3]. For the end-users effectively accomplish their tasks in a computing environment, they have to be qualified with just not fragmentary computing skills, but with the total computing capability including the understanding, knowledge, and skills of an end-user computing.

Therefore, this study presents an efficient tool to evaluate end-user computing competency that can efficiently execute the given tasks in an organizational computing environment.

2 Computing Competency

In previous literature, most studies defined an end-user as an individual who directly interacts with his or her computer [4-6]. End-user computing refers to direct

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 431–442.
springrlink.com

interaction with computing application software by managerial, professional, and operating level personnel in user departments [7-8]. Hence, the end-user computing is defined as an end-user directly interacts with computer application software and computing systems in his or her task departments.

Competency is a total set of knowledge, skills, and attitudes as the action characteristics of an organizational member that can do his or her tasks outstandingly in an organizational environment [9]. The competency is a set of observable performance dimensions, including individual knowledge, skills, attitudes, and behaviors, as well as collective team, process, and organization capabilities that is linked to high performance, and provides the organization with sustainable competitive advantage [10]. The competency is a measurable pattern of knowledge, skill, abilities, behaviors, and other characteristics that an individual needs to perform work roles or occupational functions successfully [11].

With analysis of the literature, we summarized five major components of competency as shown in Table 1: Motives, Traits, Self-concepts, Knowledge, and Cognitive and Behavioral Skills [9-12].

Table 1. Major components and definitions of competency

Division	Definitions and contents
Cognitive/Behavior Skills	The ability to perform specific mental or physical tasks. Mental or cognitive skills include analytical or cognitive thought.
Knowledge	This is information that knows for specific department. It only indicates that what a person can do, but does not predict what a person will actually do.
Self-Concepts	This means attitude, a sense of value, and a self-portrait. A sense of value is an element which reflects on responsible activities in a given situation for a short-period.
Traits	This means a consistent response to physical characteristics, situation or information. An emotional self-control and careful attitude is a consistent response of a more complicated form.
Motives	This is a cause of activity leading an individual to do what he wants to do and what he consistently had in mind to do. This is an action which selects and instructs a trigger for a specific activity or an objective.

In general competency, individual characteristics such as motives, traits, self-concepts and knowledge lead to skills, and the action of a person with skills has an effect on the performance of his or her task [12]. In other words, computing competency can be explained by transforming general competency into a type of competency based on a computing perspective.

Hence, end-user computing competency (EUCC) can be defined as a total set of knowledge, technology, skills and attitudes which function as action characteristics of an organizational member who can do his or her task outstandingly in a computing environment. EUCC indicates an individual's total ability to apply computing knowledge, solutions, and computing systems to his or her tasks. It finally means a total computing capability that an end-user can efficiently perform his or her given tasks in an organizational computing environment.

This study generated the first 32 items to evaluate EUCC based on the 5 components of general competency. These items were developed from studies and discussions by the experts such as postdoctoral researchers, professors and senior developers in IT and computing research centers, and the previous literature on an end-user computing [1,7-12].

3 Research Methods

Many previous studies presented the methods to verify the validity of a model construct. Generally, two methods were presented to verify a model construct validation: (1) correlations between total scores and item scores, and (2) factor analysis [13]. The former used correlation analysis to verify the validity of the model construct [1,16]. The latter utilized factor analysis to verify the validity of the model construct [14,15]. This study used factor analysis and reliability analysis to verify the validity and reliability of the tool construct and to extract adequate items for measuring EUCC. The evaluation questionnaire used a five-point Likert-type scale; where, 1: not at all; 2: a little; 3: moderate; 4: good; 5: very good. The survey was gathered data from various departments to generalize the results.

A sample of 228 usable responses was obtained from a variety of industries and business departments, and from management levels. The industries represented in the sample were manufacturing (10.1%), construction (9.2%), bank and insurance (18.9%), communication and service (29.8%), and information system implementation service (32.0%). The respondent had on average of 8.9 years of experience (S.D. = 1.164) in their field, their average age was 35.8 years old (S.D. = 5.643). The survey method used in this evaluation questionnaire was based on two collection methods: by direct collection and e-mail.

3.1 Analysis and Results

By the criterion of previous literature, Items were excluded when their correlation with the collected item-total was < 0.5 or when their correlation with the criterion scales was < 0.6 [1,7,8]. The correlations with the corrected item-total and the criterion item were significant at $p < 0.01$ and similar to those used by others in previous studies [7,8]. The elimination was considered enough to ensure that the retained items were adequate measures of the EUCC.

Table 2. Factor loadings, corrected item-total correlation and coefficients alpha of extracted items

Variable	Factor Loading				Corrected Item-total Correlation	Coefficients Alpha
	Factor 1	Factor 2	Factor 3	Factor 4		
V1	0.813				0.689	
V2	0.748				0.647	0.807
V3	0.701				0.562	
V4		0.871			0.734	
V5		0.792			0.684	
V6		0.746			0.620	0.867
V7		0.702			0.601	
V8		0.683			0.584	
V9			0.882		0.756	
V10			0.825		0.712	
V11			0.746		0.673	0.894
V12			0.721		0.602	
V13			0.652		0.579	
V14				0.781	0.681	
V15				0.698	0.604	0.801
V16				0.647	0.523	

* Significant P ≤ 0.01

The tool construct was verified by factor analysis and reliability analysis. The inadequate items for the tool were deleted by the analysis results. These deletions resulted in a 16-item scale for evaluating EUCC. In general, the 16 items had factor loadings > 0.647 and the coefficients (Cronbach's alpha) > 0.801. The descriptions and loadings for the 16 items are presented in Table 2, and grouped by their higher factor loading. Each of the 16 items had corrected item-total correlations > 0.523 ($p \leq 0.01$). Hence, the evaluation items with a validity and reliability were extracted as shown in Table 2 and 3.

4 Evaluation Tool

This study extracted 16 items to evaluate EUCC and the extracted items were classified as four factor groups as presented in Table 2. These groups indicate the potential factors for evaluating EUCC, and they become four evaluation factors of the developed tool. With exploring the evaluation items of each factor, we identified the four potential factors as follows: factor 1: computing perception; factor 2: computing knowledge; factor 3: computing utilization; and factor 4: computing potential. Fig. 1 shows the developed evaluation tool for the EUCC based on the 4 potential factors and 16 evaluation items.

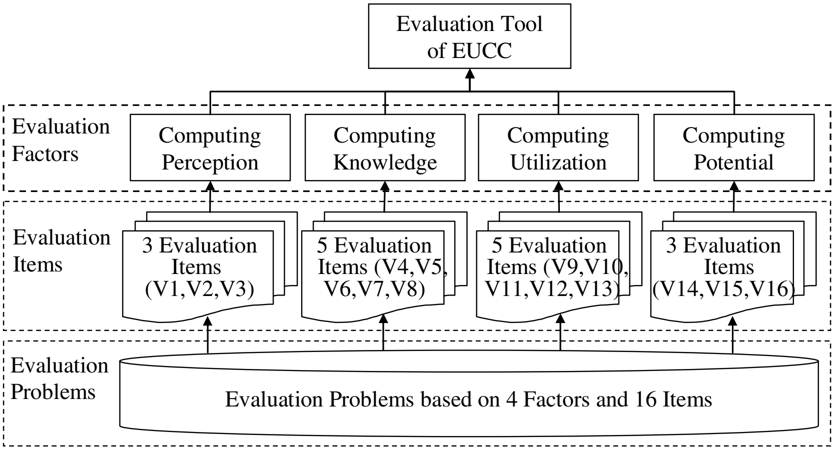

Evaluation Tool of EUCC
Evaluation Factors
Computing Perception
Computing Knowledge
Computing Utilization
Computing Potential
Evaluation Items
3 Evaluation Items (V1,V2,V3)
5 Evaluation Items (V4,V5, V6,V7,V8)
5 Evaluation Items (V9,V10, V11,V12,V13)
3 Evaluation Items (V14,V15,V16)
Evaluation Problems
Evaluation Problems based on 4 Factors and 16 Items

Fig. 1. Structure of evaluation tool

The four potential factors are considered the major components of this tool. This is a measure that can totally evaluate an end-user computing capability, and is different from the previous evaluation tools that can gauge the fragmentary knowledge and capability such as knowledge, technology, and skills of an end-user computing [17].

Table 3. Evaluation factors and items

Factors	Extracted Evaluation Items
Computing Perception	V1: Concepts on future computing paradigm
	V2: Understanding of computing progress trends
	V3: Ethic consciousness in a computing environment
Computing Knowledge	V4: Knowledge related to H/W, S/W, N/W, and DB etc.
	V5: Solution knowledge related to ERP, SCM, KMS, and CRM etc.
	V6: Knowledge related to solutions (B2E, B2C, and B2B)
	V7: Knowledge related to operation systems
	V8: Knowledge related to computing security systems
Computing Utilization	V9: Ability of word processing, presentation, and spreadsheet
	V10: Ability using solutions of ERP, SCM, CRM, and KMS etc.
	V11: Ability using H/W, S/W, N/W, and D/B of computing Systems
	V12: Ability applying computing systems to B2E, B2C, and B2B
	V13: Ability of security establishment and management
Computing Potential	V14: Number of working years in computing department
	V15: Completion of computing education and training courses
	V16: Presentation of ideas and articles on websites or in journals

4.1 Evaluation Factors and Items

In Table 3, the computing perception evaluates acknowledgement, attitude, a sense of value, and adaptability on end-user computing with the evaluation items such as concepts of future computing, understanding of computing progress trends, and ethic consciousness in an organizational computing environment.

The computing knowledge appraises the knowledge of computing solutions and systems with the evaluation items such as the knowledge of H/W, S/W, N/W and DB, solutions of ERP, SCM, KMS, and CRM, e-Business (B2E, B2C, and B2B), and operation of computing systems and security systems.

The computing utilization gauge the capability that applies the computing knowledge, solutions, and systems to his or her tasks with the evaluation items such as OA ability of spreadsheet, presentation and word processing, the ability to use business solutions of ERP, SCM, CRM, and KMS, the ability to use hardware, software, network and database, the ability to apply the computing systems to the end-user's tasks such as e-business of the form B to E, B to C, and B to B, and the skills related to establishment and management of the security system.

The computing potential assesses the potential development probability of end-user computing capability by job experience, participation of educations and trainings, and presentation of ideas and articles on websites or in journals. This is an important factor for the extension of computing competency in terms of breadth and depth of end-user computing.

As shown in Fig. 1 and Table 3, the evaluation tool with 4 factors and 16 items is an important theoretical construct to assess the end-user's total computing ability that can efficiently do his or her tasks in an organizational computing environment.

5 Evaluation Systems

The evaluation system, Fig. 2, comprises the evaluation stages and procedures for gauging EUCC. It has two main processes of the evaluation stage and the presentation stage of the evaluation results. The evaluation stage first extracts the evaluation problems from the problem database based on each factor and item. By the characteristics of each factor, the evaluation problems are identified as three kinds of problem forms such as a questionnaire test, a written test, and a utilization test.

The factors such as the computing perception and the computing potential are tested by a questionnaire form, and the computing knowledge and the computing utilization are appraised by a written and a utilization form. After that, the tool examines the end-users by the extracted evaluation problems.

The evaluation results are analyzed by extracting the evaluation values in each factor. The presentation of the evaluation results provides the interpretations of the evaluation results of each factor. The results are explained by each evaluation index (EI) extracted from each factor. The interpretation of the evaluation results

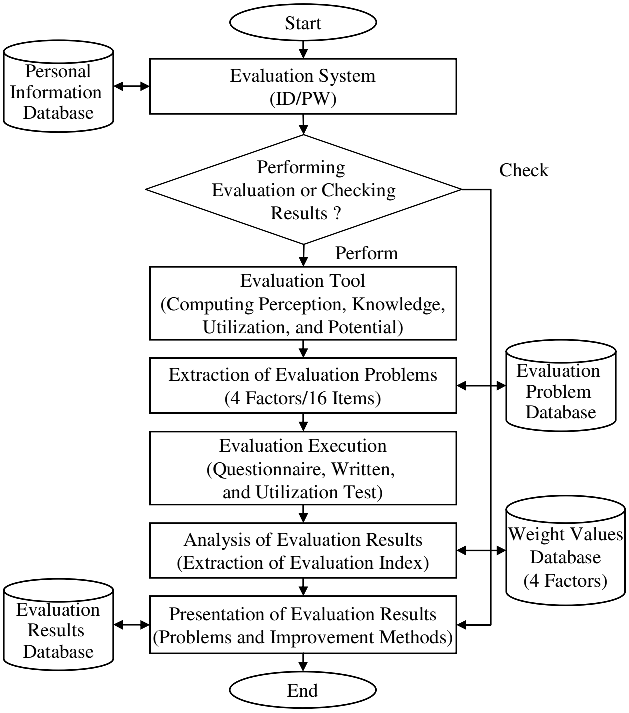

Start
Personal Information Database
Evaluation System (ID/PW)
Performing Evaluation or Checking Results ?
Check
Perform
Evaluation Tool (Computing Perception, Knowledge, Utilization, and Potential)
Extraction of Evaluation Problems (4 Factors/16 Items)
Evaluation Problem Database
Evaluation Execution (Questionnaire, Written, and Utilization Test)
Analysis of Evaluation Results (Extraction of Evaluation Index)
Weight Values Database (4 Factors)
Evaluation Results Database
Presentation of Evaluation Results (Problems and Improvement Methods)
End

Fig. 2. Evaluation system

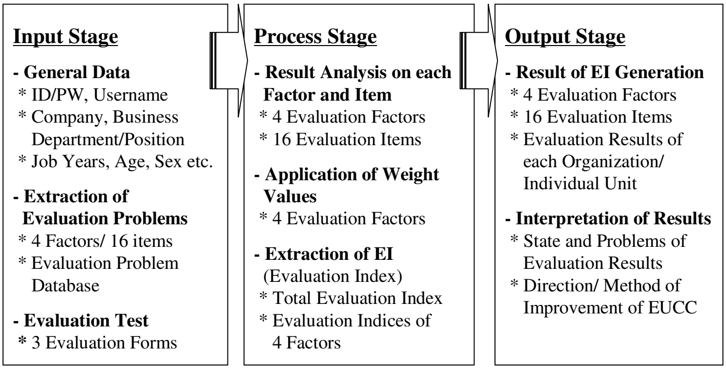

Input Stage
- General Data
* ID/PW, Username
* Company, Business Department/Position
* Job Years, Age, Sex etc.
- Extraction of Evaluation Problems
* 4 Factors/ 16 items
* Evaluation Problem Database
- Evaluation Test
* 3 Evaluation Forms
Process Stage
- Result Analysis on each Factor and Item
* 4 Evaluation Factors
* 16 Evaluation Items
- Application of Weight Values
* 4 Evaluation Factors
- Extraction of EI (Evaluation Index)
* Total Evaluation Index
* Evaluation Indices of 4 Factors
Output Stage
- Result of EI Generation
* 4 Evaluation Factors
* 16 Evaluation Items
* Evaluation Results of each Organization/ Individual Unit
- Interpretation of Results
* State and Problems of Evaluation Results
* Direction/ Method of Improvement of EUCC

Fig. 3. Three stages and contents of the evaluation system

presents the present states and problems of the EUCC, and the directions and methods to efficiently improve the EUCC based on the extracted evaluation indices. Fig. 3 indicates the input, process, and output stage in this evaluation system.

The input stage includes a general data, extraction of evaluation problems, and evaluation test. The process stage analyzes the evaluation results by extracting the EI. The output stage provides interpretations of the analysis results.

5.1 Evaluation Method

This study used the weight values for each factor in order to develop an efficient tool considered the relative importance of each factor in evaluating EUCC. The weight values, Table 4, were extracted from the analysis results of the questionnaire survey (AHP) for about 40 experts working in computing departments.

The evaluation method first calculates the evaluation values of each factor through analyzing the evaluation results that an end-user is measured by the extracted problems. The tool figures out the EI of each factor by multiplying each weight value by the evaluation values of each factor. The sum of evaluation indices of each factor becomes the total EI of an end-user. Namely, the total EI of an EUCC is the sum of evaluation indices of each factor.

Table 4. Weight value of each evaluation factor

Evaluation Factor	Weight Value
Computing Perception	0.24
Computing Knowledge	0.26
Computing Utilization	0.31
Computing Potential	0.19

Hence, the EI of each factor can be presented as Equation (1).

$$\mathrm{EI}_{\mathrm{EFi}} = \mathrm{EV}_{\mathrm{EFi}} \text{ x } \mathrm{WV}_{\mathrm{EFi}} \tag{1}$$

where:

$\mathrm{EI}_{\mathrm{EFi}}$: Evaluation index (EI) of the i-th Evaluation Factor
$\mathrm{EV}_{\mathrm{EFi}}$: Evaluation Value (EV) of the i-th Evaluation Factor
$\mathrm{WV}_{\mathrm{EFi}}$: Weight Value (WV) of the i-th Evaluation Factor

Here, the sum of the weight values of each factor is 1.00 and i = 1, 2, 3 and 4 indicate four evaluation factors.

Therefore, the total EI can be defined as Equation (2) with Equation (1):

$$\text{Total EI} = \sum_{i=1}^{4} \mathrm{EI}_{\mathrm{EFi}} \tag{2}$$

here, i = 1, 2, 3 and 4 mean the four evaluation factors.

In this way, this tool presents the evaluation results of EUCC based on the total EI and the EI of each factor.

6 Case Study and Analysis Results

We applied the developed tool to 164 end-users working in “A” Enterprise, Republic of Korea. The business departments of respondents were identified as follows: strategy plan department: 25.4%; development and maintenance department: 20.6%; business application department: 35.2% and administration support department: 18.8%. The business positions of respondents were classified as follows; top managers: 4.1%; middle managers: 25.7% and workers: 70.2%. The respondents had on average 7.6 years of experience (SD = 0.587).

6.1 Application and Analysis of overall Organization

The total EI of the overall organization was 61.96, and the strategy plan department and the business application department were 62.57 and 64.68 as shown in Fig. 4. The evaluation results of each department showed that the EI of the business application department was higher than those of the other departments.

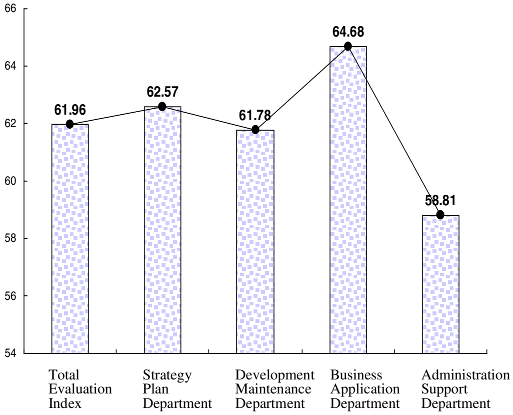


Fig. 4. Evaluation indices of each business department

This is due to their ability that effectively accomplishes their tasks by frequently applying computing knowledge, solutions, and systems to e-Business of the form B to E, B to C and B to B, and by utilizing the various solutions such as ERP, SCM, CRM, and KMS in order to do their business tasks in an organizational computing environment. But the end-users in the administration support department have to make an effort to efficiently raise their total computing capability.

6.2 Application and Analysis of a Business Department

The total EI of the strategy plan department (SPD) was 62.57, and it was quite high. The evaluation indices of the SPD were quite high in two evaluation factors, except for the computing knowledge and potential as shown in Fig. 5.

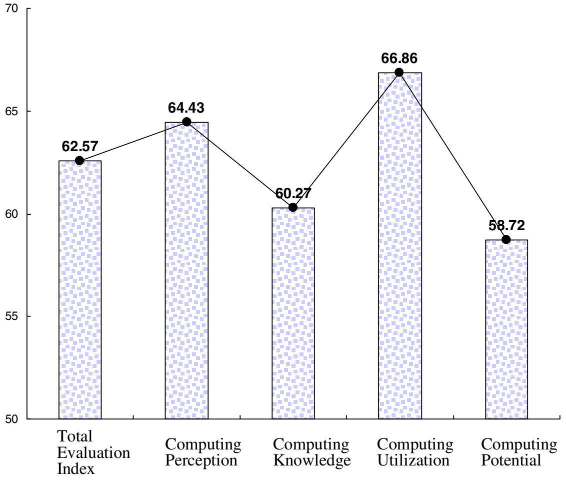


Fig. 5. Evaluation indices of each factor of the SPD

Hence, the end-users of the SPD should endeavor to improve the computing knowledge and potential factors through acquirement of degrees and certificates, completion of educations and trainings, and production of computing knowledge in order to effectively raise the organizational computing competency.

6.3 Application and Analysis of an Individual

An end-user working in the administration support department (ASD) were evaluated as an example. Table 5 shows the extraction process of the total EI for an end-user. The total EI of this EUCC was 63.04 as indicated in Table 5. Especially, the EI of the computing utilization was high. This means the outstanding application ability that efficiently applies the computing knowledge, solutions, and systems to the given tasks in an organizational computing environment.

The evaluation indices of the computing perception, knowledge, and utilization were high. But the EI of the computing potential was very low.

Table 5. Extraction process of total evaluation index of an end-user

Division	Computing Perception	Computing Knowledge	Computing Utilization	Computing Potential	Total Evaluation Index
Evaluation Indices of Each Factor	62.76	63.38	68.56	57.61	
Weight Value of Each Factor	0.24	0.26	0.31	0.19	1.00
Calculation of Total Evaluation Index	15.06	16.48	21.25	10.95	63.74

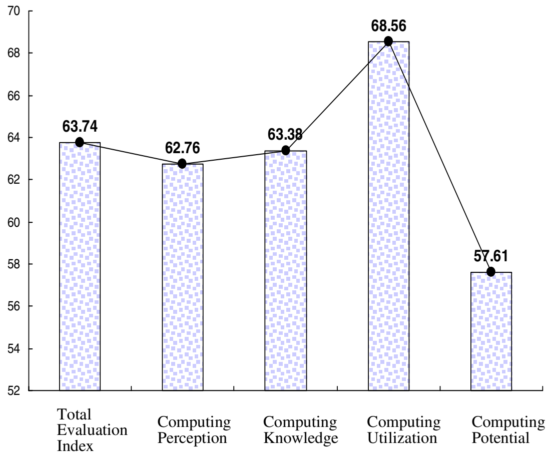


Fig. 6. Evaluation indices of an end-user in the ASD

Therefore, this end-user should make an effort to complete some computing education and training, acquire diplomas, and produce computing knowledge in order to efficiently raise his or her total computing capability.

7 Conclusions

Generally, most evaluation tools have limitations of their applications in a specific perspective. We should endeavor after additional evidence in terms of the tool's validity, internal consistency, and stability. We verified a validity and reliability of the developed tool and made an effort to generalize it. This study presented a feasible tool that can efficiently evaluate the EUCC in an organizational computing environment. This tool provides the concrete evaluation items with evaluation system, process, and method. Hence, this study offers a new groundwork for developing a practical tool that can efficiently evaluate the end-user computing capability required to perform the given tasks in an organizational computing environment.

References

1. Torkzadeh, G., Lee, J.W.: Measures of perceived end-user's computing skills. Information and Management 40(7), 607–615 (2003)
2. Bostrom, R.P., Olfman, L., Sein, M.K.: The importance of learning style in end-user training. MIS Quartery 14(1), 101–119 (1990)
3. Cheney, P.H., Mann, R., Amoroso, D.L.: Organizational factors affecting the success of end-user computing. Journal of Management Information Systems 3(1), 65–80 (1986)
4. Rockart, J., Flannery, L.: The management of end user computing. Communication of the ACM 26(10), 776–784 (1983)

5. Martin, J.: Application development without programmers. Prentice-Hall, Eaglewoods (1982)
6. McLean, E.R.: End-user of application developers. MIS Quarterly 10(4), 37–46 (1979)
7. Doll, W.J., Torkzadeh, G.: The evaluation of end-user computing involvement. Management Science 35(10), 1151–1171 (1989)
8. Brancheau, C., Brown, V.: The management of end-user computing: Status and Directions. ACM Computing Surveys 25(4), 437–482 (2002)
9. Boyatiz, R.E.: The competent manager: A model for effective performance. John Wiley & Sons, New York (1982)
10. Arthey, T.R., Orth, M.S.: Emerging competency methods for the future. Human Resource Management 38(3), 215–226 (1999)
11. Rodriguez, D., Patel, R., Bright, A., Gregory, D., Gowing, M.K.: Developing competency models to promote integrated human resource practices. Human Resource Management 41(3), 309–324 (2002)
12. Spencer, L.M., Spencer, S.M.: Competence at work: Models for superior performance. John Wiley & Son Inc., Chichester (1993)
13. Kerlinger, F.N.: Foundations of behavioral research. McGraw-Hill, New York (1978)
14. Doll, W.J., Torkzadeh, G.: The evaluation of end-user's computing satisfaction. MIS Quarterly 12(2), 982–1003 (1988)
15. Etezadi-Amoli, J., Farhoomand, A.F.: A structural model of end user computing satisfaction and user performance. Information and Management 30(2), 65–73 (1996)
16. Torkzadeh, G., Doll, W.J.: The development of a tool for measuring the perceived impact of information technology on work. Omega, International Journal of Evaluation Science 27(3), 327–339 (1999)
17. Rifkin, K.I., Fineman, M., Ruhnke, C.H.: Developing technical managers - first you need a competency model. Research Technology Management 42(2), 53–57 (1999)

Enterprsise Ontology for Knowledge-Based System

J. Andreasik

Zamość University of Management and Administration, Zamość, Poland
jandreasik@spp.org.pl

Abstract. The paper presents an original enterprise ontology oriented to the diagnosis of an economic situation. Three categories form the ontology: an agent argumentation (A), an expert assessment (E) and explanation acts (AE). An evaluation model defined in the paper consists of definitions of the set of images of enterprise assessments Ω, the assessment φ, agent argumentation for the potential range ARG(P), agent argumentation for the risk range ARG(R), and the generalized score trajectory Tgeneral. This model creates a basis for construction of the A-E-AE ontology; related taxonomy diagrams are presented.

1 Introduction

Development of modern software systems supporting enterprise management, i.e., systems such as MRP (Material Resources Process Planning), ERP (Enterprise Resources Process Planning), CRM (Customer Resources Process Planning), SC (Supply Chain), PM (Project Management), DSS (Decision Support System), EIS (Executive Information System), BI (Business Intelligence), ES (Expert System), KMS (Knowledge Management System), KOS (Knowledge Organization System), CM (Corporate Memories), etc., requires conceptualization and structuralization of the knowledge of the enterprise adequate to its development level. This aspect of software system construction concerns creating the enterprise ontology. Until now a number of enterprise ontologies have been developed, that can be classified according to the enterprise description point of view: a) **Domain approach**, b) **System approach**, c) **Planning approach**, d)**Transactional approach**, e) **Identification approach** and f) **Diagnostic approach**.

The EO ontology developed by the Artificial Intelligence Applications Institute (AIAI), The University of Edinburgh in 1997 [1] is regarded as a representative of the **domain approach**. In this ontology, an enterprise is characterized in four areas: strategy, marketing, organization, and activities. The ontology includes the term dictionary consisting of five sections:

1. Metaontology terms including notions: entity, actors, relationships, actor roles, time.
2. Terms related to planning including: activities, tasks, resources, resources allocation.

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 443–458.
springrlink.com

3. Terms related to the organization structure including: legal entities, organizational units, ownerships, personnel, management functions.
4. Terms related to defining a strategy including: vision, purposes, mission, decisions, critical success factors, risk.
5. Terms related to identification of marketing processes including: goods, service, prices, promotion, customers.

Creation of mechanisms of mapping the structure of the enterprise activity is essential in the **system approach**. The enterprise is treated as a set of subsystems (production, sale, storage, technical preparation of production, quality management, logistic, scheduling and planning, controlling, etc.), The system approach is a characteristic feature of the TOVE (Toronto Virtual Enterprise) ontology elaborated in Department of Mechanical and Industrial Engineering, University of Toronto [2]. This ontology is based on the enterprise ontology defined as a set of the following constraints [3]:

$$\varepsilon_{action} \cup \varepsilon_{resource} \cup \varepsilon_{goals} \cup \varepsilon_{products} \cup \varepsilon_{services} \cup \varepsilon_{occ} \cup \varepsilon_{external} \qquad (1)$$

These constraints are defined according to principles of the predicate calculus of the first order. Such principles concern constraints put on operations and tasks (*action*), on resources used in realization of activities and tasks (*resource*), on organizational roles, on goals put for an enterprise (*goals*), on technological requirements, structural and quality requirements related to products (*products*), on services related to product distribution (*services*), on internal conditions related to activities and tasks (*occ*), on the external environment dealing with customers, suppliers, competitors, etc. (*external*).

Another example of the system approach is an ontology for virtual organization (VO) elaborated within the European 6th Framework Programme (IST-1-506958-IP) in Jozef Stefan Institute, Ljubljana [4]. This ontology consists of three parts: a part which defines the structure and function of the virtual organization (VBE virtual organization breeding environments), a part describing roles of participants of VBE, a part describing competencies, resources, and their availability.

The **planning approach** includes construction of ontologies oriented to concetualization of varied business processes. In this group, the REA ontology elaborated by Geerts and McCarthy [5] is the most representative. This ontology concerns modeling of financial flows [6]. The main classes of this ontology are the following: economic resources, economic events, and economic agents. A class diagram in the UML language has been elaborated for this ontology [7].

The **transactional approach** includes construction of the ontology oriented to business interaction. Review of this type of ontology has been made by Rittgen [8]. The most advanced ontology representing the transactional approach is the enterprise ontology elaborated by Dietz [9]. This ontology is based on Ψ-theory. Four axioms are defined there:

1. Enterprise activities occur thanks to competent and responsible performing roles by actors. These roles are enforced by acts of coordination and acts of production.

2. Acts of coordination are presented as steps in typical behavior patterns. These patterns are called transactions.
3. A business process is the composition of transactions.
4. In activities performed by actors, three abilities are taken into consideration: informational (forma), reasoning (informa), and engagement (pergorma).

For modeling processes according to presented theory, a special model called CRISP has been defined. This model is defined by a tuple:

$$\left\langle C,R,I,S,P\right\rangle \tag{2}$$

where:

C – is a set of coordination facts,
R – a set of rules of actions,
I – a set of intentions,
S – a set of state facts, and
P – a set of production facts.

A technique of building diagrams called Crispienet has been also developed.

The **identification approach** is oriented to creating and archiving the resources of corporate memory. Conceptualization of competences and experience of an enterprise is essential in this approach. The analytical knowledge model has been presented by Huang, Tseng and Kusiak in [1010]. In this model, the structure of information according to the IDEF3 standard has been used. For identification of competences in an enterprise, the language called UECML (Unified Enterprise Competence Modelling Language) [11] has been elaborated in Ecole Politechnique Federalne de Lausanne, Laboratory for Production Management and Processes. The key constructions of this language are the following: activities, processes, organization objects, resources, unit competences, individual competences, collective competences, humans, actors, teams, organization units. At University of Lausanne a concept of ontology [122] has been elaborated, which can be a complete representative of the identification approach. The idea of this ontology is considered on three levels:

1. Core business: represented by business processes and activities.
2. Performance indicators: key performance indicators and intellectual capital indicators,.
3. The corporate memory chunks: represented by procedures.

The main classes of the ontology are the following: Strategy, Process, Activity, Procedure, Actor, Key performance indicators, Intellectual capital indicators. Characteristic of this ontology is the orientation to identification of the enterprise condition thanks to the introduction to the ontology indicator classes taking intellectual capital indicators into consideration.

The **diagnosis approach** is represented until now by ontologies oriented to the data mining methodology. Fan, Ren and Xiong in [13] discuss the ontology model of this approach. It is oriented towards an enterprise assessment from its credit capacity point of view. Components of this ontology are the following: credit data ontology, credit criteria ontology, evaluation method ontology, and credit knowledge ontology. In this approach, there is a lack of the ontology oriented to conceptualization of the cause-effect analysis of an enterprise.

The concept of the author of this paper includes creating the ontology taking into consideration an assessment of the potential and risk of an enterprise as conditions of its existence and functioning. Financial end economic scores are considered in time forming trajectories of scores. Ontology includes conceptualization of characteristic patterns of trajectories of scores. In this conception, the analysis of reasoning processes is expressed according to the CBR (Case-Based Reasoning) methodology [14]. Therefore, the class of explanation procedures is a part of this ontology.

2 Enterprise Model – A Diagnostic Approach

Presented idea of an enterprise assessment assumes that the assessment is carried out by experts from a consulting agency on the basis of argumentation presented by a multiagent system (MAS). The multiagent system is built on a data warehouse. Its task is to extract data for an assessment of the enterprise competence in the specific range of analyzed potential of an enterprise and for an assessment of a competence gap in the specific range of analyzed risk of an enterprise. Each expert can perform assessments in certain time instants (for example, every quarter). An enterprise can invite to the assessment external as well as internal experts. Then the group assessment images are created.

Experts make an assessment of the enterprise potential on the basis of an assessment of competences defined in individual potential ranges. Experts make an assessment of the risk of the enterprise activity on the basis of an assessment of a competence gap defined in individual risk ranges. Characteristic curves of financial and economic scores of an enterprise in a given period of its working are extracted from a data warehouse. These characteristic curves form graphical images called score trajectories.

Definition 1. *An enterprise assessment* Ω *consists of a set of images of enterprise assessments {e_i} in specific time instants {t_j} on the basis of argumentation of artificial agents {a_k} of the multiagent system MAS built on a data warehouse:*

$$\Omega = \{\omega_l\}, l = 1,\ldots,n$$

$$\forall_{\omega_l} \exists_{e_i} \exists_{a_k} \omega_l = \left\langle A(P)_{e_i}^{t_j} \mid ARG(P)_{a_k}^{t_j}, A(R)_{e_i}^{t_j} \mid ARG(R)_{a_k}^{t_j} \right\rangle$$

where:

Ω – is an enterprise assessment,
ω_l – an enterprise assessment image made by the expert e_i on the basis of argumentation of the agent a_k in the time instant t_j,

$A(P)_{e_i}^{t_j}$ – an assessment of the enterprise competences in the range of defined potential made by the expert e_i in the time instant t_j,

$A(R)_{e_i}^{t_j}$ – an assessment of the enterprise competence gap in the range of defined risk made by the expert e_i in the time instant t_j,

$ARG(P)_{a_k}^{t_j}$ – a confirmation related to evaluation of the enterprise competences in the range of defined potential, presented by the agent a_k in the time instant t_j,

$ARG(R)_{a_k}^{t_j}$ – a confirmation related to evaluation of the enterprise competence gap in the range of defined benefit, presented by the agent a_k in the time instant t_j.

The expert makes an assessment of enterprise competences in each determined range of potential according to taxonomy of potential (see below).

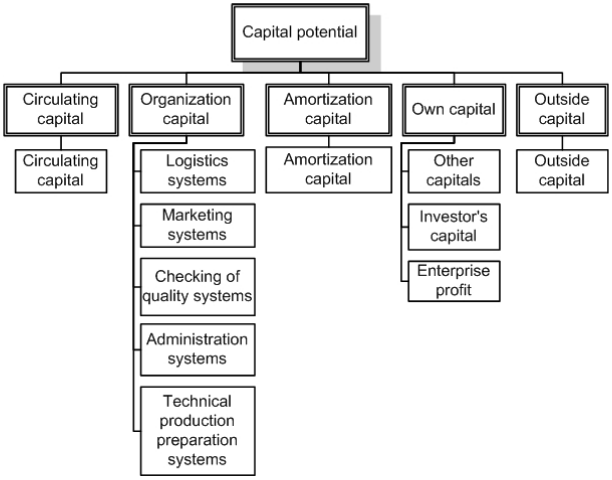


Fig. 1. Taxonomy diagram for capital potential

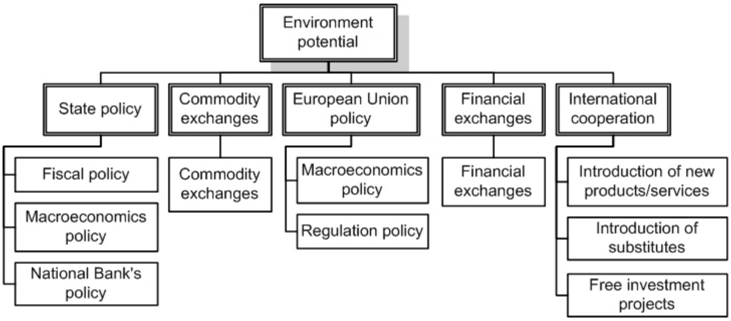


Fig. 2. Taxonomy diagram for environment potential

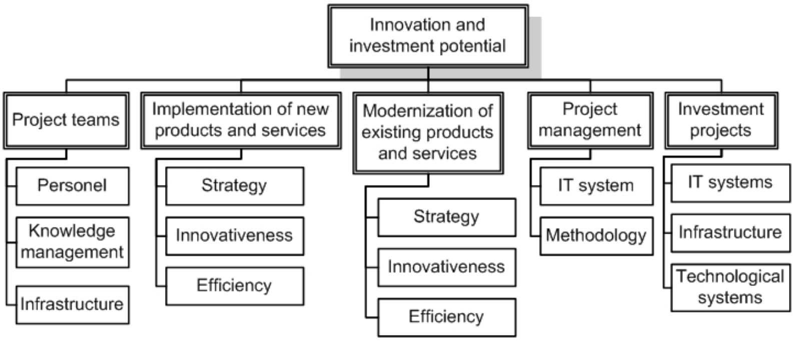


Fig. 3. Taxonomy diagram for innovation and investment potential

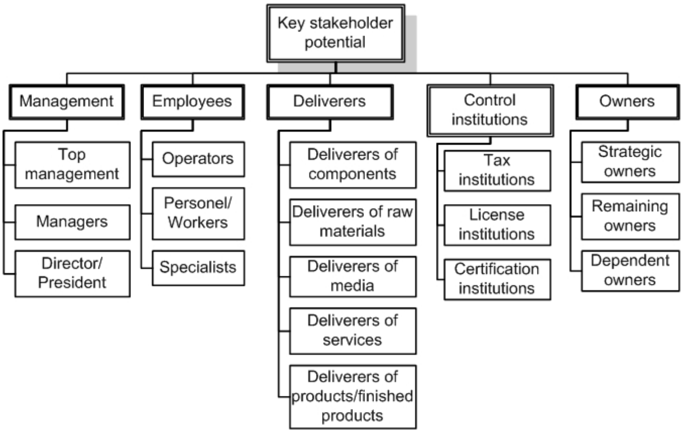


Fig. 4. Taxonomy diagram for key stakeholder potential

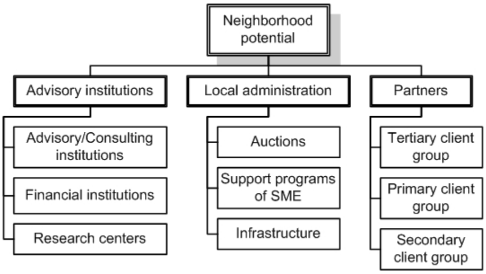


Fig. 5. Taxonomy diagram for neighborhood potential

Fig. 6 shows a hierarchical structure of taxonomy of potential consisting of three levels:

Level I: type of potential,
Level II: kinds of potential,
Level III: ranges of potential for given type and kind.

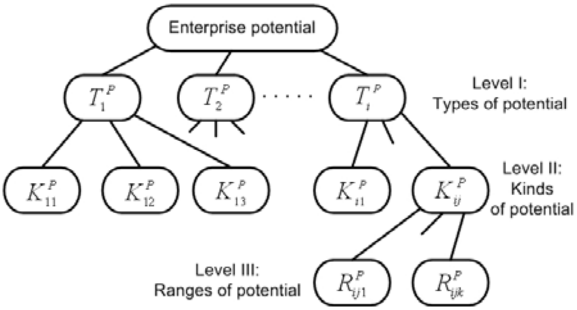


Fig. 6. Diagram of potential taxonomy

The expert makes an assessment of an enterprise competence gap in each determined range of risk according to taxonomy of risk (see below).

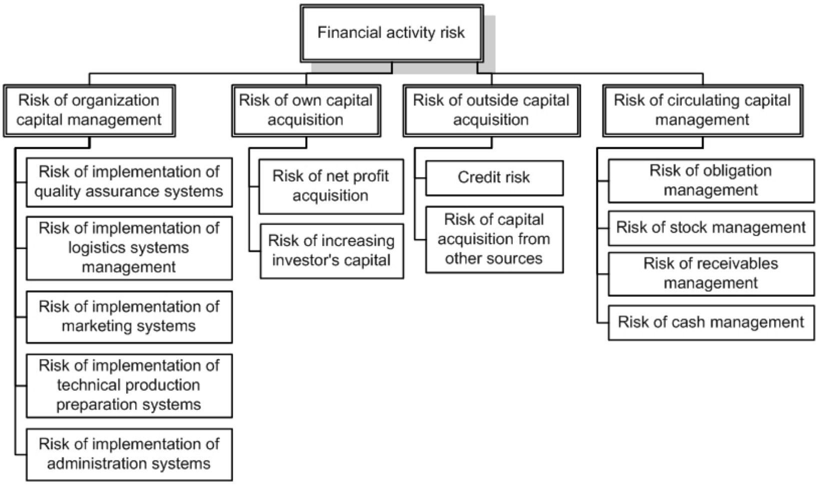

Financial activity risk
Risk of organization capital management
Risk of own capital acquisition
Risk of outside capital acquisition
Risk of circulating capital management
Risk of implementation of quality assurance systems
Risk of implementation of logistics systems management
Risk of implementation of marketing systems
Risk of implementation of technical production preparation systems
Risk of implementation of administration systems
Risk of net profit acquisition
Risk of increasing investor's capital
Credit risk
Risk of capital acquisition from other sources
Risk of obligation management
Risk of stock management
Risk of receivables management
Risk of cash management

Fig. 7. Taxonomy diagram for financial activity risk

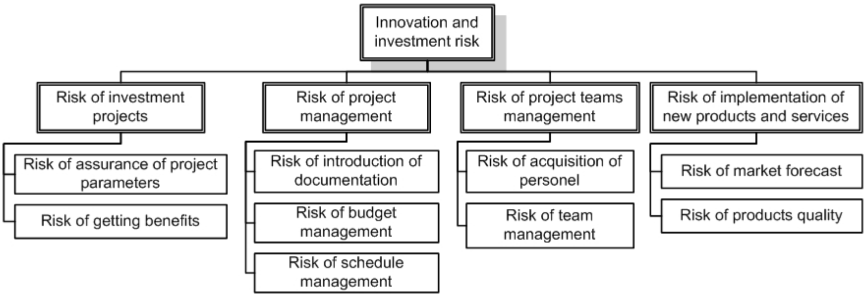

Innovation and investment risk
Risk of investment projects
Risk of project management
Risk of project teams management
Risk of implementation of new products and services
Risk of assurance of project parameters
Risk of getting benefits
Risk of introduction of documentation
Risk of budget management
Risk of schedule management
Risk of acquisition of personel
Risk of team management
Risk of market forecast
Risk of products quality

Fig. 8. Taxonomy diagram for innovation and investment risk

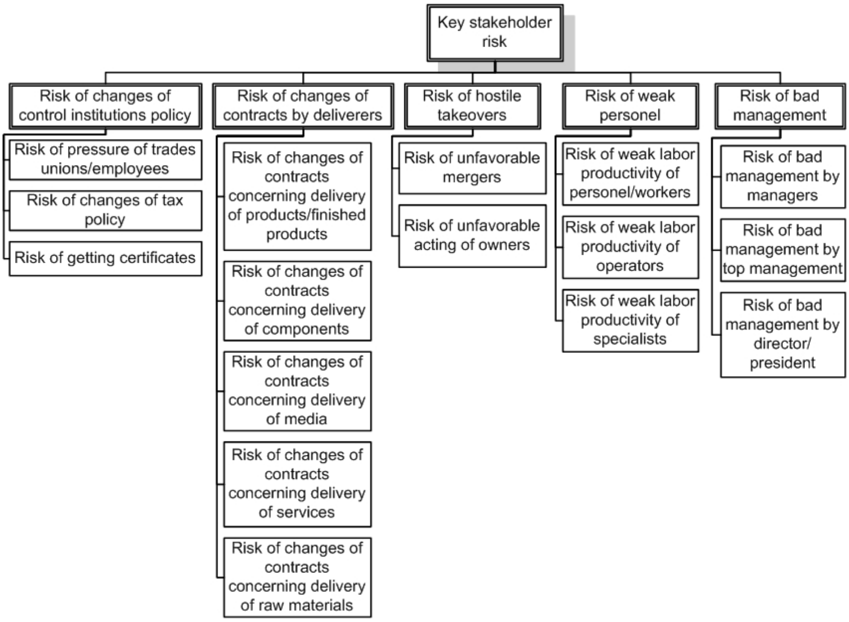

Key stakeholder risk
Risk of changes of control institutions policy
Risk of pressure of trades unions/employees
Risk of changes of tax policy
Risk of getting certificates
Risk of changes of contracts by deliverers
Risk of changes of contracts concerning delivery of products/finished products
Risk of changes of contracts concerning delivery of components
Risk of changes of contracts concerning delivery of media
Risk of changes of contracts concerning delivery of services
Risk of changes of contracts concerning delivery of raw materials
Risk of hostile takeovers
Risk of unfavorable mergers
Risk of unfavorable acting of owners
Risk of weak personel
Risk of weak labor productivity of personel/workers
Risk of weak labor productivity of operators
Risk of weak labor productivity of specialists
Risk of bad management
Risk of bad management by managers
Risk of bad management by top management
Risk of bad management by director/ president

Fig. 9. Taxonomy diagram for key stakeholder risk

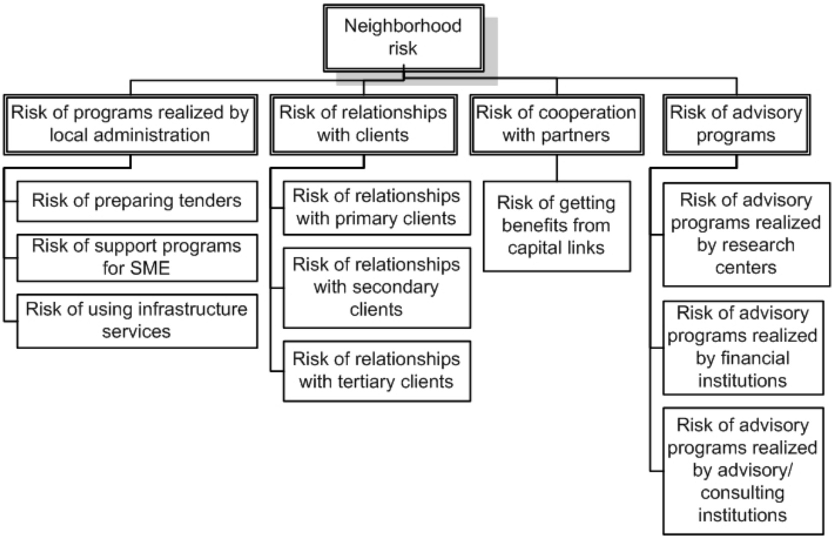

Neighborhood risk
Risk of programs realized by local administration
Risk of preparing tenders
Risk of support programs for SME
Risk of using infrastructure services
Risk of relationships with clients
Risk of relationships with primary clients
Risk of relationships with secondary clients
Risk of relationships with tertiary clients
Risk of cooperation with partners
Risk of getting benefits from capital links
Risk of advisory programs
Risk of advisory programs realized by research centers
Risk of advisory programs realized by financial institutions
Risk of advisory programs realized by advisory/ consulting institutions

Fig. 10. Taxonomy diagram for neighborhood risk

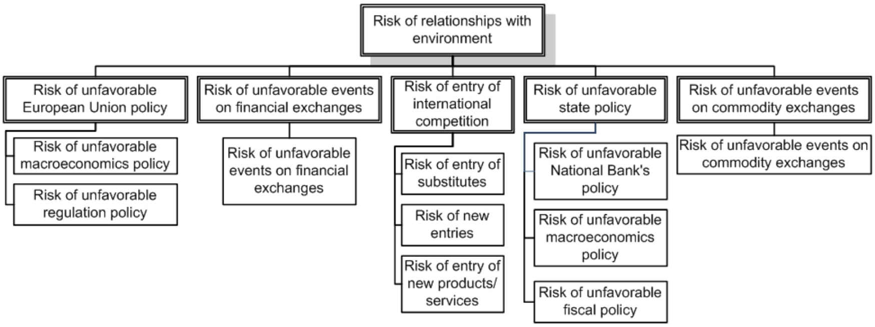

Risk of relationships with environment
Risk of unfavorable European Union policy
Risk of unfavorable macroeconomics policy
Risk of unfavorable regulation policy
Risk of unfavorable events on financial exchanges
Risk of unfavorable events on financial exchanges
Risk of entry of international competition
Risk of entry of substitutes
Risk of new entries
Risk of entry of new products/ services
Risk of unfavorable state policy
Risk of unfavorable National Bank's policy
Risk of unfavorable macroeconomics policy
Risk of unfavorable fiscal policy
Risk of unfavorable events on commodity exchanges
Risk of unfavorable events on commodity exchanges

Fig. 11. Taxonomy diagram for risk of relationships with environment

Fig. 12 shows a hierarchical structure of taxonomy of risk consisting of three levels:

level I: type of risk,
level II: kinds of risk,
level III: ranges of risk for given type and kind.

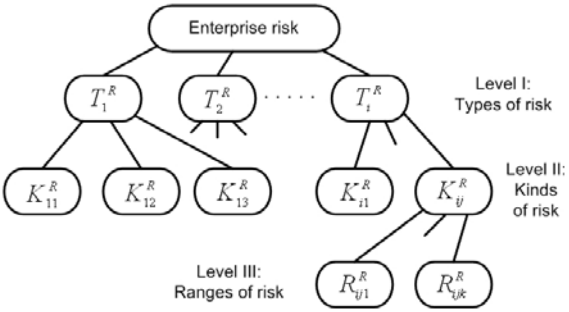


Fig. 12. Diagram of risk taxonomy

The expert performs expertise $\varphi(e_i)$ making an assessment of the enterprise competence in each determined range of potential and making an assessment of the enterprise competence gap in each determined range of risk. Additionally, the expert determines the significance of each element of potential and each element of risk for the assessment on each taxonomy level assigning adequate weights to them.

Definition 2. *Expertise φ of the expert e_i consists of: an assessment of potential A(P), an assessment of risk A(R), a set of weights of the significance of particular assessments according to taxonomy of potential W(P), a set of weights of the significance of particular assessments according to taxonomy of risk W(R):*

$$\varphi(e_i) = \langle A(P), A(R), W(P), W(R) \rangle$$

where:

$$A(P) = \{\forall_{R_{ijk}^{P}} \exists_{c(R_{ijk}^{P})} c(R_{ijk}^{P}) \in \langle 0,1 \rangle\}$$

$$A(R) = \{\forall_{R_{ijk}^{R}} \exists_{c(R_{ijk}^{R})} c(R_{ijk}^{R}) \in \langle 0,1 \rangle\}$$

$$W(P) = \langle W(T_i^P), W(K_{ij}^P), W(R_{ijk}^P) \rangle$$

$$W(R) = \langle W(T_i^R), W(K_{ij}^R), W(R_{ijk}^R) \rangle$$

Each expert assessment *c* is a numerical value from the interval $\langle 0,1 \rangle$. Moreover, each single weight of the significance of an assessment is a numerical value from the interval $\langle 0,1 \rangle$. There is argumentation $ARG(P)_{a_k}^{t_j}$ presented by the agent a_k in the time instant t_j on the basis of data extracted from a data warehouse assigned to each range of potential defined at level III of the potential taxonomy. This argumentation is a basis for an assessment of the enterprise competence in the analyzed range of potential.

Definition 3. *Agent argumentation for each range of potential is a set of competence assessments with reference to quantitative characteristic of the resource availability, quantitative characteristic of resource qualifications, quantitative characteristic of abilities, quantitative characteristic of experience, and quantitative characteristic of available technologies, methods and procedures:*

$$\forall_{R_{ijk}^{P}} \exists_{a_k} ARG(P) = \{comR^P, comQ^P, comA^P, comE^P, comT^P\}$$

here:

$comR^P$ – is the quantitative depiction of competence with respect to the resource availability required for realization of the enterprise strategy,
$comQ^P$ – the quantitative depiction of the resource qualifications required for realization of the enterprise strategy,
$comA^P$ – the quantitative depiction of abilities which characterize resources, and
$comE^P$ – the quantitative depiction of experience characteristic,
$comT^P$ – the quantitative depiction of technologies (methods) required for realization of the enterprise strategy.

Similarly, there exists an agent argumentation for each range of the risk. This argumentation is the basis of an assessment of the competence gap of an enterprise in the analyzed range of the risk.

Definition 4. *Agent argumentation for each range of the risk is a set of the competence gap assessments with reference to quantitative characteristic of the resource availability, quantitative characteristic of resource qualifications, quantitative characteristic of abilities, quantitative characteristic of experience, and quantitative characteristic of available technologies, methods and procedures*:

$$\forall_{R_{ijk}^{R}} \exists_{a_k} ARG(R) = \{comR^{R}, comQ^{R}, comA^{R}, comE^{R}, comT^{R}\}$$

where:

comR^R – is the quantitative depiction of competence with respect to the resource availability required for realization of the enterprise strategy,
comQ^R – the quantitative depiction of the resource qualifications required for realization of the enterprise strategy,
comA^R – the quantitative depiction of abilities which characterize resources,
comE^R – the quantitative depiction of experience characteristic, and
comT^R – the quantitative depiction of technologies (methods) required for realization of the enterprise strategy.

Information on changes in time of different financial indicators (for example, net profit, ROI, ROE, EBIDTA, etc.) and economic indicators (for example, EVA, SVA) and also intellectual capital indicators (for example, VAIC) is extracted from a data warehouse. Quantitative characteristics of these indicators form score trajectories.

Definition 5. *A generalized score trajectory is a set of component trajectories and a procedure of generalization:*

$$T_{general} = \left\langle \{T_i\}, PROC \right\rangle$$

where:

T_i – stands for component score trajectories, and
PROC – is a procedure of generalization (determined on the basis of methods of the multiple-criteria decision making theory).

In a process of research on enterprises of the SME sector, it will be created an album of typical patterns of trajectories. It is approach consistent with the conception of Argenti's score trajectories [15]. This album makes up a basis for creating a procedure of explanation in the CBR system for predicting situation of enterprises. Procedures of the CBR cycle include a process of matching a given score trajectory of the considered enterprise to the cases inserted into the case base of the CBR system.

Definition 6. *An album of typical images of trajectories consists of a set of T(IMAGES) created on the basis of research on enterprises of the SME sector:*

$$T(IMAGES) = \{T(IMAGE_i)\}$$

The presented model of an enterprise disregards the functional structure of an enterprise and varied processes proceeding in an enterprise and its environment. It is oriented to the tripartite analysis concerning:

a) an analysis of enterprise competence with respect to defining its potential,
b) an analysis of a competence gap identified as a result of analysis of the risk both external and internal, and

c) an explanation of score trajectories on the basis of standard images of score trajectories included in the album of the CBR system.

3 A-E-AE Enterprise Ontology – A Diagnostic Approach

The following protocols constitute enterprise ontology according to the diagnostic approach on the basis of the model presented in Section 2:

1. A protocol coming from an assessment made by the expert.
2. A protocol of agent argumentation on the basis of data extracted from a data warehouse.
3. A protocol of explanation of an enterprise position on the basis of indexation algorithms of the CBR system.
4. A protocol of matching score trajectories according to images inserted into an album of typical score trajectories of enterprises of the SME sector.
5. A protocol of predicting being an effect of the whole cycle of reasoning according to the CBR methodology.

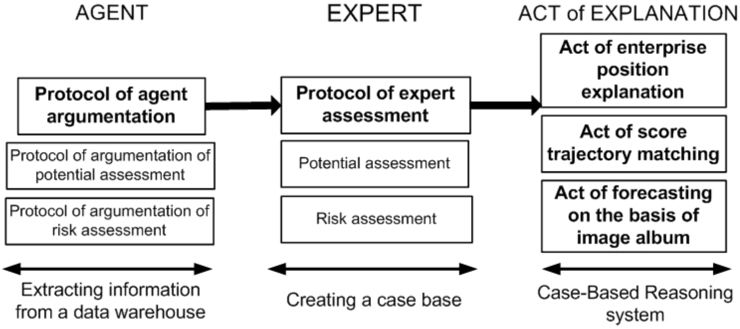

AGENT
EXPERT
ACT of EXPLANATION
Protocol of agent argumentation
Protocol of argumentation of potential assessment
Protocol of argumentation of risk assessment
Protocol of expert assessment
Potential assessment
Risk assessment
Act of enterprise position explanation
Act of score trajectory matching
Act of forecasting on the basis of image album
Extracting information from a data warehouse
Creating a case base
Case-Based Reasoning system

Fig. 13. Diagram of the A-E-AE ontology

4 Knowledge Base Model According to A-E-AE Ontology

According to the enterprise assessment model presented in Section 2, the knowledge model can be determined. Its structure is expressed by A-E-AE ontology. The knowledge about an enterprise is included in three acts of explanation:

- the act of position explanation (APE),
- the act of trajectory explanation (ATE),
- the act of forecast explanation (AFE).

Fig. 14 shows a structure of the knowledge about an enterprise.

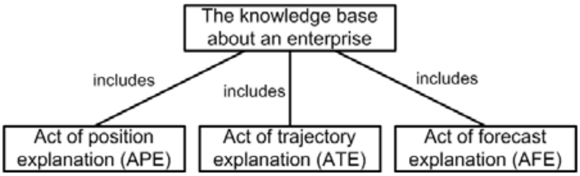

The knowledge base about an enterprise
includes
includes
includes
Act of position explanation (APE)
Act of trajectory explanation (ATE)
Act of forecast explanation (AFE)

Fig. 14. A structure of the knowledge base about an enterprise according to A-E-AE ontology

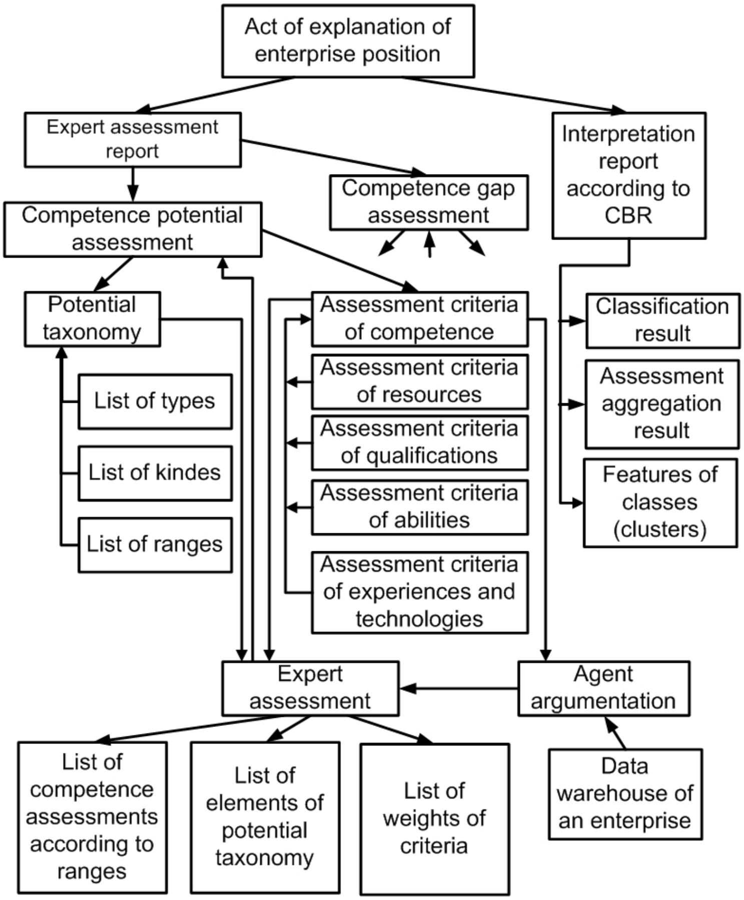

Act of explanation of enterprise position
Expert assessment report
Interpretation report according to CBR
Competence gap assessment
Competence potential assessment
Potential taxonomy
List of types
List of kindes
List of ranges
Assessment criteria of competence
Assessment criteria of resources
Assessment criteria of qualifications
Assessment criteria of abilities
Assessment criteria of experiences and technologies
Classification result
Assessment aggregation result
Features of classes (clusters)
Expert assessment
Agent argumentation
List of competence assessments according to ranges
List of elements of potential taxonomy
List of weights of criteria
Data warehouse of an enterprise

Fig. 15. A structure of the class APE (the act of position explanation) – the subclass <competence gap assessment> has the same structure as the subclass <competence potential assessment>

Each of three acts of explanation includes expert assessment made on the basis of argumentation of agents serving a data warehouse of an enterprise. Hence, the main subclasses for the classes APE, ATE, AFE are: <expert assessment>, <agent argumentation>.

A structure of the class APE consisting of two main subclasses <expert assessment report> and <CBR interpretation report> is shown in Fig. 15. A structure of the assessment report corresponds to Definitions 1, 2, 3, and 4 from Section 2. The report of interpretation of enterprise position creates automatically the knowledge base system KBS on the basis of the CBR (Case Based Reasoning) methodology. Exact procedures of the report of interpretation of enterprise position are presented by the author in [18]. These procedures include aggregation of assessments estimated by an expert on the basis of argumentation of an agent serving a data warehouse of an enterprise. The another procedure concerns determining parameters of classification (clustering) on the basis of all cases inserted into the case base The last procedure determines features of classes (clusters) using data mining methods (e.g., decision trees, rule-based systems).

The class <act of explanation of enterprise result trajectories> includes the subclass <tabular data of enterprise results> and the subclasses consisting of procedures of separating primitives according to the specially worked out grammar of shapes of result trajectories. Typical shapes of trajectories were worked out by J. Argenti [15].

The class <act of forecast explanation> includes a set of procedures of matching generalized result trajectory with the most similar trajectory on the basis of searching cases inserted into the case base. These procedures are based on the CBR (Case Based Reasoning) methodology.

The author gives the grammar of result trajectories and procedures for acts of explanation (ATP and AFP) in other papers. The structure of classes of A-E-AE ontology presented in Fig. 14 and 15 is defined in the Protege editor (version 3.3.1) [6].

5 Conclusions

In this paper the original enterprise ontology oriented to an assessment of the enterprise potential, the risk analysis and the benchmark analysis of financial and economic scores has been defined. This ontology creates a basis for building an intelligent system for predicting the economic situation of enterprises in the **S**mall-**M**edium-**E**nterprise (SME) sector; the structure of this system has been presented elsewhere [17]. Three categories constitute the ontology: agent argumentation (A), an expert assessment (E), explanation acts (AE). Agent argumentation is understood as a result of extraction of data from a data warehouse and generation of suggestions for the expert by the multiagent system (MAS). Argumentation makes up an assessment of enterprise competence in ranges of the analyzed potential and a competence gap in ranges of the analyzed risk. Structures of the potential and risk are expressed by suitable taxonomies making up instances for defined classes of the ontology. An expert assessment is a protocol of the assessment process of competence and a competence gap according to the potential and the risk, respectively. Additionally, an expert makes an assessment of the significance of each type, kind and range of the potential and the risk. In this way, a description of a case is created for inserting it into a case base of the CBR system. Explanation acts are created unaided by the CBR system on the basis of suitable procedures of the CBR methodology cycle in the refine, reuse and retrieve processes. The act of prediction is a result of automatic interpretation of a generalized score trajectory of a considered enterprise against a background of suitably matched typical images of score trajectories deposited in the album of trajectories during the research on the SME sector.

References

1. Uschold, M., King, M., Moralee, S., Zorgios, Y.: The enterprise ontology. The Knowledge Engineering Review 13, 47–76 (1998)
2. Fox, M.S., Gruninger, M.: Enterprise modelling. AI Magazine, 109–121 (1998)
3. Gruninger, M., Atefi, K., Fox, M.S.: Ontologies to support process integration in enterprise engineering. Computational & Mathematical Organization Theory 6, 381–394 (2000)
4. Plisson, J., Ljubic, P., Mozetic, I., Lavrac, N.: An ontology for virtual organization breeding environments. IEEE Transactions on Systems, Man, and Cybernetics. Part C: Applications and Reviews 37(6), 1327–1341 (2007)
5. Geerts, G.L., McCarthy, W.E.: The ontological foundation of REA enterprise information systems (2000), `http://www.msu.edu/~mccarth4/Alabama.doc` (accessed April 8, 2009)
6. Dunn, C.L., Cherrington, J.O.: Enterprise infromation systems: a pattern–based approach. McGraw-Hill, Boston (2005)
7. Hruby, P.: Model–driven using business patterns. Springer, Heidelberg (2006)
8. Rittgen, P.: Handbook of ontologies for business interaction. Information Science Reference, Hershey (2008)
9. Dietz, J.L.G.: Enterprise ontology: theory and methodology. Springer, Heidelberg (2006)
10. Huang, C.C., Tseng, T.L., Kusiak, A.: XML-based modeling of corporate memory. IEEE Transactions on Systems, Man, and Cybernetics. Part A: Systems and Humans 35(5), 629–640 (2005)
11. Pepiot, G., Cheikhrouhou, N., Furbringer, J.M., Glardon, R.: UECML: Unified enterprise competence modelling language. Computers in Industry 57, 130–142 (2007)
12. Jussupova-Mariethoz, Y., Probst, A.R.: Business concepts ontology for an enterprise performance and competences monitoring. Computers in Industry 58, 118–129 (2007)
13. Fan, J., Ren, B., Xiong, L.R.: Modeling and management of ontology-based credit evaluation meta-model. In: Proc. IEEE International Conference on Systems, Man and Cybernetics, Hague, Netherlands, pp. 3164–3168 (2004)
14. Schank, R.C., Kass, A., Riesbeck, C.K.: Inside case-based explanation. L. Erlbaum Associates Publishers, Mahwah (1994)
15. Argenti, J.: Corporate collapse. McGraw-Hill Inc., London (1976)

16. Protege (open source ontology editor), `http://protege.stanford.edu/` (accessed April 1, 2009)
17. Andreasik, J.: A case-base reasoning system for predicting the economic situation of enterprises – tacit knowledge capture process (externalization). In: Kurzynski, M., Puchala, E., Wozniak, M., Zolnierek, A. (eds.) Computer Recognition Systems 2, pp. 718–730. Springer, Heidelberg (2007)
18. Andreasik, J.: Decision support system for assessment of enterprise competence. In: Kurzynski, M., Wozniak, M. (eds.) Computer Recognition Systems CORES3. Springer, Heidelberg (2009) (in printing)

Constructing Ensemble-Based Classifiers Based on Feature Transformation: Application in Hand Recognition

H. Mirzaei[1] and M. Jafarzadegan[2]

[1] Intelligence Databases, Data Mining and Bioinformatics Research Laboratory
Department of Electrical and Computer Engineering, Isfahan University of Technology
Isfahan, IRAN
hmirzaei@ec.iut.ac.ir

[2] Islamic Azad university, Mobarake branch,
mdjfrzdgn@miu.ac.ir

Abstract. This paper presents a new method for user identification based on hand images. Because in user identification the processing time is an important issue, we use only the hand boundary as a hand representation. Using this representation, an alignment technique is used to make the index of corresponding features of different hands get the same index in the feature vector. To improve the classification performance a new ensemble-based method is proposed. This method uses feature transformation to create the needed diversity between base classifiers. In other words, first different sets of features are created by transforming the original features into new spaces where the samples are well separated, and then each base classifier is trained on one of these newly created features sets. The proposed method for constructing an ensemble of classifiers is a general method which may be used in any classification problem. The results of experiments performed to assess the presented method and compare its performance with other alternative classification methods are encouraging.

1 Introduction

In recent years, a series of automated human recognition systems have emerged. One can categorize useful metrics for user recognition in three groups.

What user enters into system for introducing himself, such as username, password, PIN code, and so on.

What user offers into system to access the system, such as Card, Token, and so on.

What the user is: that is the user's voice, fingerprint, eye retina, and so on.

Two first groups have the characteristic to be transferred to others (voluntary or involuntary). The third one which consists of physical or behavioral characteristics of humans such as fingerprint, face, DNA, hand image, and signature is called biometrics. Biometrics, due to the lack of portability, is the strongest metric in spite of its high implementation cost.

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 459–469.
springrlink.com

Discussion about user recognition is divided into user verification and user identification. In user verification we try to identify that the entered user is truly the same person as he/she has claimed, so the output of the system would be true or false. In user identification we want to determine which person enters to the system.

Biometric authentication systems based on hand image can be divided to palmprint recognition systems and hand –geometry based authentication systems [1-21]. Palmprints are the pattern of skin on the surface of the palm. Palmprint appear to be closely similar to fingerprint, whose identification is mature [22]. However most of the methods proposed for fingerprints are not suitable for palmprints and many researches have been carried out toward deriving specialized methods for palmprints [13-21].

Another possibility for user recognition based on hand image is to use the geometry of a person's hand to authenticate his identity [1-12]. Features like area/size of palm, length, width, and angle of fingers, palms aspect ratio, and various contour transformations can be used in such a system. Although there may be less significant information in hand geometry than e.g. iris, this is more convenient to use your hand than to stare in a camera. Also there is cheaper hardware in hand geometry solutions than in other solutions. Other advantage of using the hand as a biometric is that a hand is easier and faster to put in front of a device. Additionally, simple measurement of hand geometry features, and the possibility to be integrated with other biometrics make hand geometry an effective biometric.

Various approaches have been proposed for hand geometry based user recognition. Some researchers have tried to model hand image data as a function by interpolation or some other data fitting methods. In [1] different 3-D hand modeling and gesture recognition techniques are studied. There, it is stated that a geometrical hand shape model can be approximated using splines. For example in [2] a 4 B-spline curve have been used to represent fingers. In this system, the fingers (except the thumb) are represented using the curves and used for comparison. Besides the splines parameters a set of 3-D points can also be used to model the hand [1]. The polygon meshes formed by these set of points in 3-D space can be used to approximate the hand shape. Hand model can also be constructed from a set of images of the hand from different views [1]. In [3, 4] implicit polynomials, which are very powerful in object modeling, have been used for recognizing hand shapes.

Another popular approach for hand geometry based user recognition is to extract some features from the hand image and use these features in the matching phase. In [5] 30 different features are obtained from the hand image, then in the 30 dimensional feature space a bounding box is found for each person's training set. The distance of the query image to these bounding boxes is used as the measure of similarity. If this similarity is smaller than a threshold, the user is identified as unknown. In [6] the eigen vectors of the original features of hand are used as a new set of features. Euclidean distance classifier is used for palmprint recognition on these newly created features. In [7] several features are extracted from each feature point of the vein patterns and used these features in biometric user identification. In [8] and [9] new hand shape features are proposed and based on these features a bimodal system, a combination of two biometrics using hand geometry and palmprints, is proposed.

Some of the proposed hand recognition systems use pegs to make all of the captured images aligned with each others [10-11]. This makes the feature extraction and matching process easier. Some of the other works do not use any pegs to cause any inconvenience to the users. In theses systems some kind of image alignments is required prior to feature extraction [3, 4, 12].

The performance of some hand–geometry based authentication systems are summarized in Table 1. In this table, FAR (False Accept Rate) is the rate at which the system accepts a non-authenticated user, FRR (False Reject Rate) is the rate of rejection of a genuine user by the system, and *n* stands for not reported values.

Table 1. Summarized characteristics of some hand–geometry based authentication systems

Reference	Datasets size	Recognition	rate	FRR	FAR
		Identification	Verification		
[2]	120 image of 20 person	97%	95%	n	n
[3]	n	n	99%	n	n
[5]	714 Image of 70 person	n	n	6%	1%
[6]	200person	n	n	1%	0.03%
[7]	-	n	n	1.5%	3.5%
[8]	100person	n	n	0.6%	0.43%
[9]	1000 image of 100 person	n	n	1.41%	0%
[11]	360 image of 50 person	n	n	15%	2%
[12]	353 image of 53 person	n	n	3.5%	2%

Ensemble-based methods have successfully been used in a variety of application. It has been proven that increasing diversity between ensemble members without increasing their individual test error necessarily results in a decrease in ensemble test error. On the other hand, if all ensemble members agree in all of their classifications, then the aggregated classification will be identical to that of any individual ensemble member without any decrease in ensemble test error. In this paper a new ensemble creation method based on feature transformation is proposed which produces a set of different classifiers. To assess the utility of the proposed method, it is applied to the problem of hand recognition.

The rest of this paper is organized as follows. In the next section we provide explanations of methods used for preprocessing and feature extraction. In Section 3 we describe our proposed ensemble method based on feature transformation. As part of this, diversity creation method is given. Experimental setup and the produced results are given in Section 4. Finally, conclusions are presented in Section 5.

2 Preprocessing and Feature Extraction

In order to design a classification system, it is essential to use a proper representation of input patterns (here the shape of person's hand). Because in user identification the processing time is an important issue, users expect the system to be quick and safe and they do not want to wait for the response of the system, we must use simple but sufficient representation of input objects. In this paper, the hand contour features are used as a representation of person's hand. The details of extracting hand contour features are given in the following.

The first step of our preprocessing operation is the removal of background. In this step the difference between the hand image and an image taken with no hand in front of the scanner is calculated. In the resulting image the background will be a constant gray level region. The next step is the determination of image pixels which are probably on the hand's edge. These points can be determined by a variety of edge detection methods. In this paper the Sobel's operator (further called sobel, for short) is used. The resulted image of applying this operator is set with a threshol, and then filtered by a median filter. These operations increase the clearness of image around the hand boundary. Finally using a boundary tracking method, a continuous and smooth contour of hand is extracted. After finding the hand contour some features can be extracted from it. We used k-slope features as a set of hand features. The k-slope feature is extracted at each point by measuring the angle between the vector created from the current point and some of its preceding points and the vector created from the current point and some of its following points.

To reduce the computational complexity, the k-slope feature is calculated every n points (in our experiments we set n=10). It is worth noting that the resulted list of features is a rotation invariant representation. Because the feature vectors produced until this step are time series, to compare them the correspondence problem between their elements must be solved first. The correspondence problem can be solved by the well-known algorithms such as DTW or HMM. Using DTW method the classification time and using the HMM method the training time will be increased significantly. Because we are interested in using an ensemble-based method for classification, the training and classification time are very important. In this paper we proposed to use the geometrical characteristics of hand image to solve the mentioned correspondence problem quickly.

2.1 Aligning the Elements of Feature Vectors

To solve the correspondence problem, different parts of feature vectors are aligned to a common scale. To partition the hand contour, we use the fact that the angle change for fingertips and valleys is bigger compared with angle change in other areas. The fact that most hands have five fingers or less is also exploited. The following algorithm splits the hand contour into nine blocks, corresponding to five fingertips and the four valleys. These blocks will be aligned on a common scale.

1. Compute the angle change between each point and its previous point.
2. Apply a low pass filter α times to smooth the result of previous step.
3. Convert the produced series to a binary valued series by the threshold of θ .
4. Find the start and end of each sequence of running ones.
5. If run length of ones is smaller than λ remove the corresponding block and convert its elements to zero.

6. While the number of blocks is greater than 9
    a. find two nearest blocks
    b. merge these two blocks and update the number of block.

We have set the value of α, θ, and λ to 5, 0.1, 12 respectively. These values are found experimentally. Figures 1 through 5 show the result of applying the above mentioned partitioning algorithm. In Figure 1 the raw features vector (the angle or k-slope values for boundary points) and in Figure 2 the absolute difference between k-slope values are shown. Figure 3 presents the result of applying a mean filter 5 times to the angle difference values. The thresholded results are also indicated on this plot. In this experiment there is no block smaller than λ pixels. The next figure illustrates the merge operations performed to reduce the number of blocks to 9. The final result of splitting algorithm and the original time series are shown together in figure 5.

After finding 9 blocks of hand image boundary, one can place the blocks on a common scale for all images. To do so the blocks elements are mapped to the final feature vector blocks centered at 25, 75, 125, and so on. The remaining values are filled by zero. Doing this procedure all feature vectors will have the same length. Furthermore, the corresponding features of different images will take the same index of final feature vector.

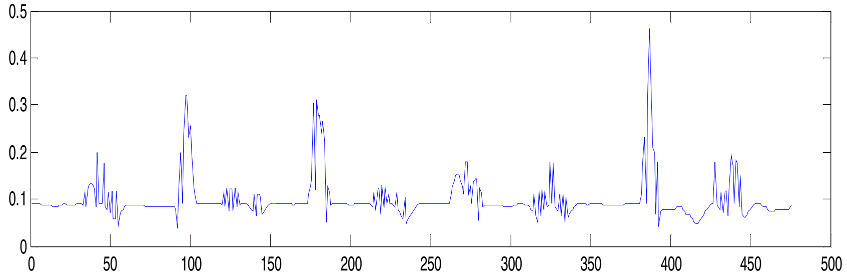


Fig. 1. The k-slope values for boundary points of a hand image

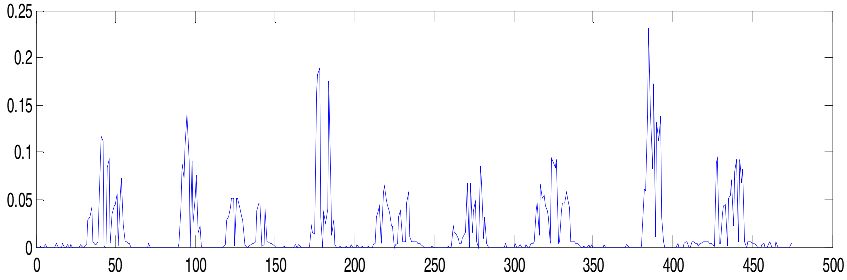


Fig. 2. The absolute difference of k-slope(angles) values along the boundary points of hand image

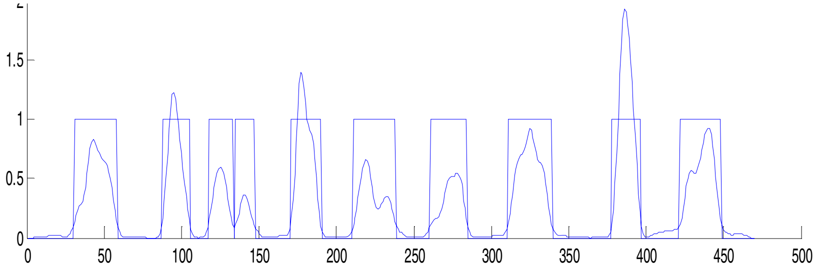


Fig. 3. Smoothed and thresholded angle difference values

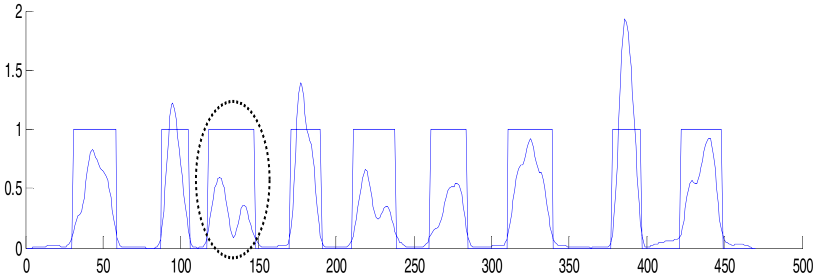


Fig. 4. Merging nearest blocks to reach to 9 blocks

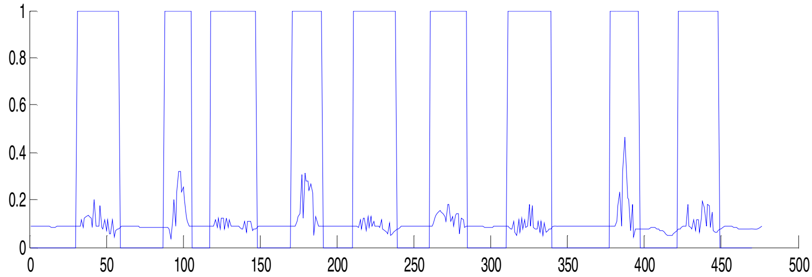


Fig. 5. The blocks generated from splitting algorithm is shown on the k-slope values of boundary points

3 Classifier Ensemble Based on Feature Transformation

To create a classifier ensemble two main issues must be concerned. The first one is that how the base classifiers are created. To use the advantage of ensembles in learning, the base classifiers output on different patterns must be as correct as possible and at the same time, their outputs must be different from each other. In other words the diversity between classifiers is an important factor in classifier ensembles.

The second issue to be concerned about classifier ensembles is that how to aggregate the output of base classifiers into a single output. Various aggregators are presented to combine the outputs of different classifiers (see for example [23] and the references in it). In the next subsections, the diversity creation method is described.

3.1 Diversity Creation Method

A basic classification system consists of three levels. In the first level the raw data of input patterns are collected. After that, in level 2, some features are extracted from these data. Finally a classifier is trained based on extracted features. If one is interested to create a collection of diverse classifiers, he must create classifiers which differ from each others in one of these levels. In some ensemble based methods such as bagging and boosting learning is done on different patterns sets. Some other methods, e.g. Random Subspace Method (RSM), use different feature sets to make base classifiers different. Using different learning method or the same learning method with different parameters can also be used as a diversity creation method.

In this paper, we propose a diversity creation method based on feature transformation. Let C be the number of target classes of input patterns. We train M base classifiers where M is equal to $G*C$ and $G \geq 1$. These classifiers are divided to C groups, where each group corresponds to a specific class. The features given to the classifiers of each group are the ones that discriminate the corresponding class of this group from other classes in a proper manner. It should be noted that although the input features to each classifier are appropriate for discrimination one class from others, all classifiers are asked to learn all classes. The pseudo code of the proposed algorithms is given in the following lines.

Training

```
for each c=1,2,…,C
    for each g=1,2,…G
        $i = i + 1$
        find the feature transformation matrix which discriminate class c from
         other classes
        $T_i$ = Feature_trans_mat($\{(x_1, y_1),\ldots,(x_N, y_N)\}$,c)
        use the returned transformation matrix to project training samples to a
        new domain for each pattern x
                $x' = x * T_i$
        $h_i = L_b(\{(x'_1, y_1),(x'_2, y_2),\ldots,(x'_N, y_N)\})$
```

Classification

Return $h_{fin}(x) = \arg\max_{y \in Y} \sum_{i=1}^{G*C} I(h_i(x * T_i) = y)$

Feature_trans_mat ($\{(x_1, y_1),\ldots,(x_N, y_N)\}$,c)

1. Create an observation matrix with each row indicating a simple observation and each column a feature.

   $X = [x_1, x_2, \ldots x_N]^t$

   where X^t is the transpose of X

2. Divide the input patterns to the patterns with target class c and the remaining patterns (with target class $\overline{c}$). Create observation matrices for each group.
   1. P = observation matrix composed of $\{x_i | y_i = c\}$

      Q = observation matrix composed of $\{x_i | y_i \neq c\}$
   2. Normalize the features of X , P , and Q to zero mean

      $X_m = X_m^{old} - \overline{X}_m$ for m = 1,…, M

      $P_m = P_m^{old} - \overline{P}_m$ for m = 1,…, M

      $Q_m = Q_m^{old} - \overline{Q}_m$ for m = 1,…, M

      where X_m, P_m, and Q_m are the mth normalized feature(column m), X_m^{old} , P_m^{old} , and Q_m^{old} are the mth original feature, and $\overline{X}_m$, $\overline{P}_m$, and $\overline{Q}_m$ are the mean value of mth original feature in X , P , and Q respectively.
   3. Compute Total sum of squares matrix T

      $T = X^t * X$

      where X^t is the transpose of X
   4. Compute Within-groups sum of squares matrix W

      $W = P^t * P + Q^t * Q$
   5. Compute Between-groups sum of squares matrix B

      $B = T - W$
   6. Calculate the eigenvalues and eigenvectors of $W^{-1}B$

      eigenval= eigenvalues of $W^{-1}B$

      eigenvec= eigenvectors of $W^{-1}B$
   7. Sort eigenvectors based on the descending order of eigenvalue and select k of them randomly. In this selection the eigenvectors corresponding to higher eigenvalue have higher probability to be selected.
   8. Put together r selected eigenvectors to form a matrix of M*r dimension and return it as a transformation matrix.

4 Experimental Results

To evaluate the proposed method, we collected data of 40 persons. For each person 10 images were acquired using the Hewlett Packard Scanjet 3P scanner. Scans were taken at a resolution of 640*480 pixels, in gray level format. Each hand takes 30 second to be scanned once. After each hand scanning, the user removes his hand completely from the glass and a new scan is taken which will be used in background removal.

Throughout the experiments k-nearest neighbor algorithm with k equal to 5 was used for classification purpose. This value was determined experimentally. The leaving-one-out technique is used in evaluating the accuracy of the classifiers. In the proposed ensemble based techniques one base classifier is trained foe each class (G=1). So there will be 40 base classifiers. In order to compare the performance of proposed method with non-ensemble based techniques, we build classifiers on raw features, and extracted features using dimensionality reduction techniques. PCA and Fisher transformations are used for dimensionality reduction. The recognition rate of these systems and the proposed ensemble-based method are shown in Table 2. In these experiments the dimensionality of new feature vectors was selected to be 40.

Table 2. Classification accuracy of proposed ensemble method and some single classifier methods

	Single			Ensemble based
Classifier	Raw features	PCA	Fisher	
Recognition Rate	73.67%	79.33%	94.81%	96.50%

As it can be seen from Table 2, using dimensionality reduction techniques leads to superior results with respect to using raw features. In addition Fisher transformation performed better than PCA. This is due to the fact that patterns classes are not considered by the PCA transformation, but Fisher transformation uses this information to find the new space which the patterns are projected to. Another interesting point is that between the compared classifiers, the ensemble based one is the clear winner. In the second experiment the effect of changing the dimensionality of final feature space used by base classifiers is studied. We evaluated the performance of proposed method with base classifiers trained on a space with the number of dimensions as 10, 20, 30, 40, 50, and 60. The result of this experiment is shown in figure 6.

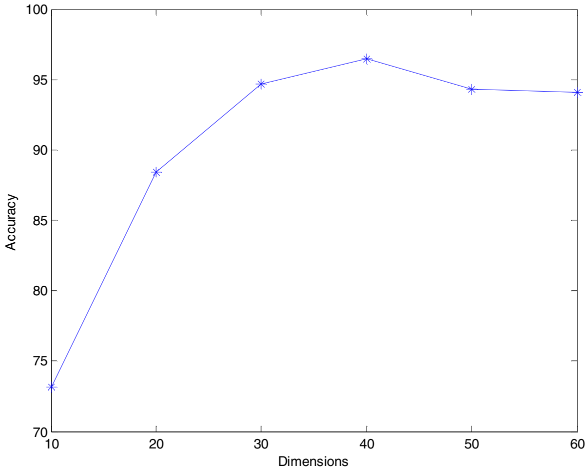


Fig. 6. Accuracy of the proposed method with base classifiers trained on 10, 20, ...,60 features

In order to study the effects of combination method used to aggregate the decision of different base classifiers, we used BKS (Behavior Knowledge Space), Ranked voting, and naive Bayes besides simple voting. The final feature space of base classifiers is chosen to be of dimension 40. The performance of these combination methods, except the BKS method, is observed to be acceptable and similar to each others. The performance of hand recognition system using BKS as a combiner is extremely low. This is because this method needs a lot of training patterns. In this method a lookup table of size C^L is produced, where C is the number of target classes and L is the number of base classifiers. If 10 patterns is needed to learn each table cells then $(40)^{40} \times 10$ patterns is required to fully learn the table. In this experiment the number of training patterns is extremely lower than $(40)^{40} \times 10$, so its low quality is predictable.

Table 3. The effect of aggregator in the performance of ensemble based approach

Combination Method	BKS	Ranked voting	naive Bayes	simple voting
Recognition Rate	82.1%	96.3%	95.9%	96.50%

5 Conclusions

One way to find a good classifier for a particular problem is to integrate several simple classifiers. However researchers have shown that the performance of such an ensemble of classifiers in addition to the base classifier performance is dependent on their diversity. In this paper we proposed a diversity creation method to be used in ensemble based classification. The proposed algorithm is a general technique and can be used in any classification problem. The results of applying this method to the hand recognition show its effectiveness in practical problems.

References

1. Wu, Y., Huang, T.: Analysis and Animation in the Context of HCI. Human hand modeling. In: Proc. International Conference on Image Processing, Japan, pp. 6–10 (1999)
2. Ma, Y.L., Pollick, F., Hewitt, T.W.: Using b-spline curves for hand recognition. In: Proc. International Conference on Pattern Recognition, vol. 3, pp. 274–277 (2004)
3. Oden, C., Yildiz, V., Kirmizitas, H., Buke, B.: Hand recognition using implicit polynomials and geometric features. In: Bigun, J., Smeraldi, F. (eds.) AVBPA 2001. LNCS, vol. 2091, pp. 336–341. Springer, Heidelberg (2001)
4. Oden, C., Ercil, A., Buke, B.: Combining implicit polynomials and geometric features for hand recognition. Pattern Recognition Letters 24(13), 2145–2152 (2003)
5. Bulatov, Y., Jambawalikar, S., Kumar, P., Sethia, S.: Hand recognition using geometric classifiers. In: Zhang, D., Jain, A.K. (eds.) ICBA 2004. LNCS, vol. 3072, pp. 753–759. Springer, Heidelberg (2004)
6. Lu, G., Zhang, D., Wang, K.: Palmprint recognition using eigenpalms features. Pattern Recognition Letters 24(9-10), 1463–1467 (2003)

7. Lin, C.L., Fan, K.C.: Biometric verification using thermal images of palm-dorsa vein patterns. IEEE Transactions on Circuits and Systems for Video Technology 14(2), 199–213 (2004)
8. Kumar, A., Zhang, D.: Integrating shape and texture for hand verification. In: Proc. 3rd International Conference on Image and Graphics, Washington, DC, pp. 222–225 (2004)
9. Kumar, A., Wong, D., Shen, H., Jain, A.: Personal verification using palmprint and hand geometry biometric. In: Kittler, J., Nixon, M.S. (eds.) AVBPA 2003. LNCS, vol. 2688, pp. 668–678. Springer, Heidelberg (2003)
10. Ross, A., Jain, A., Pankati, S.: A prototype hand-geometry based verification system. In: Proc. Intern. Conference on Audio- and Video-based Biometric Person Authentication, Washington, DC, pp. 166–171 (1999)
11. Jain, A., Probhaker, S., Ross, A.: Biometric-based web access. Department of Computer Science and Engineering; Michigan State University, East Lansing (1998)
12. Jain, A., Duta, N.: Deformable matching of hand shapes for verification. In: Proc. International Conference on Image Processing, Kobe, Japan, pp. 857–861 (1999)
13. Zhang, D., Shu, W.: Two novel characteristics in palmprint verification: Datum point invariance and line feature matching. Pattern Recognition 32(4), 691–702 (1999)
14. Duta, N., Jain, A., Mardia, K.V.: Matching of palmprint. Pattern Recognition Letters 23(4), 477–485 (2001)
15. Han, C., Cheng, H., Fan, K., Lin, C.: Personal authentication using palmprint features. Pattern Recognition 36(2), 371–381 (2003)
16. You, J., Kong, W., Zhang, D., Cheung, K.: On hierarchical palmprint coding with multi-features for personal identification in large databases. IEEE Transactions on Circuit Systems for Video Technology 14(2), 234–243 (2004)
17. You, J., Li, W., Zhang, D.: Hierarchical palmprint identification via multiple feature extraction. Pattern Recognition 35, 847–859 (2002)
18. Zhang, L., Zhang, D.: Characterization of palmprints by wavelet signatures via directional context modeling. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 34(3), 1335–1347 (2004)
19. Li, W., Zhang, D., Xu, Z.: Palmprint identification by Fourier transform. International Journal of Pattern Recognition and Artificial Intelligence 16(4), 417–432 (2002)
20. Zhang, D., Kong, W., You, J., Wong, M.: On-line palmprint identification. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(9), 1041–1050 (2003)
21. Joshi, D., Rao, Y., Kar, S., Kumar, V.: Computer vision based approach to personal identification using finger crease pattern. Pattern Recognition 31(1), 15–22 (1998)
22. Jain, A., Pankanti, S., Prabhakar, S., Ross, A.: Recent advances in fingerprint verification. In: Bigun, J., Smeraldi, F. (eds.) AVBPA 2001. LNCS, vol. 2091, pp. 182–191. Springer, Heidelberg (2001)
23. Kuncheva, L.: Combining pattern classifiers: Methods and algorithms. Wiley & Sons Inc., New York (2004)

A New and Improved Skin Detection Method Using Mixed Color Space

M.M. Aznaveh, H. Mirzaei, E. Roshan, and M.H. Saraee

Intelligence Databases, Data Mining and Bioinformatics Research Laboratory,
Department of Electrical and Computer Engineering, Isfahan University of Technology
Isfahan, IRAN
{mohsen.mahmoudi,hmirzaei,roshan,saraee}@ec.iut.ac.ir

Abstract. In this paper a new and robust approach of skin detection is proposed. In the previous proposed system, we introduced a method for skin detection based on RGB vector space. An extended and modified approach based on a mixed color space is presented. The new approach overcomes the shortcoming of the previous one on detecting complex image's background. This has been achieved by using the HSV parameters to obtain accurate skin detection results. Furthermore an iterative technique is significantly useful for obtaining the more accurate and efficient method. This can be done by changing the vectors in two phases. Skin color has proven to be a useful cue for pre-process of face detection, localization and tracking. Image content filtering, content aware video compression and image color balancing applications can also benefit from automatic detection of skin in images. In order to evaluate our proposed approach we present the results of the experimental study. The results obtained are promising and show that our proposed approach is superior to the existing ones in terms of the number of pixels detected.

1 Introduction

An important new application domain of computer vision has emerged over the past few years dealing with the analysis of images involving humans. This domain called skin detection has proven to be a useful cue for preprocess of face recognition, localization and tracking. Because of new trend in this area researchers are turning their attention more and more toward skin detection. Indeed, skin detection plays an important role in a wide range of image processing applications ranging from face detection, face tracking, gesture analysis, content-based image retrieval systems and to various human computer interaction domains.

Many strategies based on heuristic and pattern recognition have been used for skin color segmentation. The most recent skin-detection methods are based on skin color. Skin detection based on color can be processed very fast and is highly robust to geometric variations of the face pattern [1]. Numerous techniques in literature for skin detection have used color. Skin detection using color information can be a sensitive technique under variety of conditions and factors such as illumination, background and camera characteristics. In addition, since color is orientation

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 471–480.
springrlink.com

invariant under certain conditions, motion estimating will be a very easy process by applying a simple translation model.

Using color based methods for skin detection has several disadvantages. For example color of skin in pictures taken by different cameras is not same even in same conditions. Furthermore the skin color gained from different pictures is different even in pictures taken by one camera. Another disadvantage of using skin color base methods is their sensitivity to illumination under which the image was taken especially in RGB colorspace [2].

AKUMANU P. and et al. [3] deal with the changing illumination conditions, by using illumination adaptation techniques along with skin-color detection. In their approach skin-color constancy and dynamic adaptation techniques were used to increase the efficiency of skin detection process in dynamically changing illumination and environmental conditions.

Two significant advantages of the skin detection methods based on color are that:

i. They are fast in training and usage
ii. They are independent to the distribution of neighbor color.

If, for example, we consider the RGB quantized to 8 bits per color, we'll need an array of 2^{24} elements to store skin probabilities. In our method we reduce the amount of memory needed by a statistical method proposed.

The breakthrough in hardware design for graphics processors and extensive use of parallelism has enabled graphic developers to achieve outstanding performance improvements in the past few years. Skin color based methods easily matched for parallel computing. Significant speed up can be seen through grid computing.

This paper attempts to improve the performance of color based skin detection methods along with low calculation cost. To employ simple yet effective approach, RGB vector space was used. We constructed an RGB vector as a condition of skin color that is compared with a predefined threshold and therefore we can distinguish the skin pixels. Also to improve the efficiency and gain less process time it is persuasive to consider multiple pixels instead of one. The new method overcomes the time problem of the previous method for extracting the whole data and vectors needed by mixing multiple pixels without affecting the accuracy. This will manifestly raise the efficiency of proposed method as it is clearly shown in a comparison made between our methods and other novel and successful skin detection systems. In addition an iterative technique is useful to attain the more accurate and efficient method. This can be done by changing the vectors in two phases.

Another important feature to be highlighted here is that skin color perceived by a camera can change by the changes in lighting. Therefore, for a robust skin pixel classifier, a dynamic skin color model that can handle the changes must be employed.

The organization of the paper is as follows. In Section 2 we present a survey on the existing methods. The proposed approach is discussed in Section 3. Section 4 presents the experimental results and finally we conclude the work and offer some future works in Section 5.

2 Pervious Works

Skin color segmentation problem can be solved by a variety of strategies depending strongly on type of image to be processed. Numerous works for skin color segmentation and detection have been reported in several past years and a few papers comparing different approaches have been published [4][5][1].

For example regarding to this point that skin surface reflect the light differently from other surfaces, data mining methods can be applied to detect skin pixels in an image. Decision rules and fuzzy clustering can be combined to detect skin colored regions in an image. Each pixel is represented in various color spaces such as HSV, YIQ, YCbCr and CMY to find the best result.

Sanjay Kr. Singh and et al. [6] integrated and compared 3 algorithms based on 3 color spaces RGB, YcbCr and HIS. They assessed advantages and disadvantages of using each color space. Then a high accuracy solution for skin detecting was produced.

Bruno Jedynak, and et. al. in [2] reported a work regarding applying maximum entropy models for skin detection. Three models from a large number of labelled images were considered. Each model was a maximum entropy model with respect to constraints concerning marginal distributions. First the baseline model was applied, and then Hidden Markov Model was examined and results were improved. Next, the color gradient was included. Finally an analytical expression for the coefficients of the associated maximum entropy model was obtained. The results were once more improved.

Benjamin D. Zarit and et al. in [2] suggested using color-histogram based approaches that were intended to work with a wide variety of individuals, lighting conditions, and skin tones. For example, one method was the widely-used lookup table method, the other made use of Bayesian decision theory. Also some spatial and texture analyses were used for enhancements.

Another related work reported by Peer [7] uses a simple skin classifier method using a number of rules. In this work (R, G, B) is classified as skin if:

$$\begin{aligned} & R > 95, G > 40, B > 20 \\ & \max\{R, G, B\} - \min\{R, G, B\} > 15 \\ & | R - G | > 15, R > G, R > B \end{aligned} \tag{1}$$

The most challenging part of this method is to find an appropriate color space and also some good decision rules practically which is complicated and few machine learning algorithms have been used to solve this problem.

Skin detection approaches can be divided into two main categories, nonparametric methods and parametric methods. Several nonparametric ones use histograms [8, 9, 10, 11]. The chrominance plane of color space is quantized into a number of bins, each corresponding to particular range of color component value pairs (in 2D case) or triads (in 3D case). These bins, forming a 2D or 3D histogram are referred to as the lookup table (LUT). Each bin stores the number of times this particular color occurred in the training skin images. After training, the histogram counts are normalized, converting histogram values to discrete probability distribution [1]:

$$P_{skin}(c) = \frac{skin[c]}{Norm} \tag{2}$$

Where skin[c] gives the value of the histogram bin, corresponding to color vector c and Norm is the normalization coefficient (sum of all histogram bin values [12], or maximum bin value present) [2].

Moreover, in nonparametric methods Bayes theorem was used:

$$P(skin \mid c) = \frac{P(c \mid skin)P(skin)}{P(c \mid skin)P(skin) + P(c \mid \neg skin)P(\neg skin)} \tag{3}$$

P(c|skin) and P(c|¬skin) are directly computed from skin and non-skin color histograms [13]. The prior probabilities P(skin) and P(¬skin) can also be estimated from the overall number of skin and non-skin samples in the training set [2,12,14].

The disadvantages of nonparametric methods are much storage space required and inability to interpolate or generalize the training data [1].

Parametric distributions are those methods that use a kind of specific distribution to estimate the skin pixels. One of distribution that has been used was Gaussian or normal distribution.

There are various methods working based on finding a threshold for skin classifier. For example there are known nonparametric methods, like histograms [15-17], semi-parametric ones, like the self-organizing map [18], neural networks [19], and parametricmethods, assuming a certain distribution, say a Gaussian or Gaussian mixture [16, 20, 21]. The Gaussian mixture is not a good choice in some conditions, however, and it will fail to discover true structure where the partitions are clearly non-Gaussian [22].

Nizar Bouguila and Djemel Ziou in [23] presented a generalization of the Dirichlet distribution which can be a very good choice to overcome the disadvantages of the Gaussian. The Dirichlet distribution is the multivariate generalization of the beta distribution, which offers considerable flexibility and ease of use [24].

3 Proposed Method

In the new method at first the RGB color space is used. A vector must be found to indicate the R, G and B of skin color [25]. However in the previous paper we used just one vector, using more vectors leads us to simulate vector colors better and so have a better result. Now to achieve this goal, we also use HSV color space, therefore a mixed vector space was used. The main problem is to locate the space that skin color vectors exist. This step should be done by checking the skin databases. This vector is found via training where the mean of a sample can be thought as indicator vector named, “the Basic Vector”. For clarifying the skin pixels, we have two steps. In the first step, if the pixel is a skin color pixel that is obtained by comparing norm and angle of pixels with Basic Vector (pixel color based phase) that now includes both HSV and RGB parameters. The second step is to find the relation between the pixel and neighbors. Since using an iterative model can lead us to a better result, in this work we first find skin color based on our database. Then we change the place of vector spaces to gain a better result. Altering vector spaces is based on first computation and it can be easily performed by the most repeated high probability skin pixels. In some cases this algorithm can even solve the problems that color based methods have in different illuminations or different races.

The formula of first step is shown below:

$$\cos\theta = \frac{\vec{a}.\vec{V}}{|\vec{a}|\times|\vec{V}|} \tag{4}$$

Where $\vec{V}$ is the Basic Vector and $\vec{a}$ is color vector and $\vec{a}\,.\vec{V}$ is dot product.

$$\theta < \beta \tag{5}$$

$$\|\vec{V} - \vec{v}_c\| < \kappa \tag{6}$$

The inequalities (5) and (6) are using threshold values (beta and kappa) estimating skinness of pixel. Beta and kappa are calculated via statistics. In this inequality $\vec{a}$ is the vector of current pixel.

In other hand we can use just (6) both for basic vector and the pixel that is going to be tested. We have tested both in two programs. This step can be considered as Color Slicing in some resources. Gonzales and Woods [13] have used these formulas for color slicing:

$$s_i = \begin{cases} 0.5 if\left[|r_j - a_j| > \dfrac{W}{2}\right]_{any\quad l\le j\le n} \\ r_i \end{cases}$$

Or (7)

$$s_i = \begin{cases} 0.5 \quad if \sum_{j=1}^{n}(r_j - a_j)^2 > R_0{}^2 \\ r_i \end{cases}$$

Where S_i is the transformed picture which slices a specific color, based on value of W or R_0. Here we tried to slice skin color.

In the next step we used the fact that a skin pixel is not by itself rather its neighbors should be skin pixel. The difference of eight vectors was calculated separately. Next a threshold has been applied on differences as shown in (8).

$$\|\vec{V} - \vec{v}_n\| < \xi \tag{8}$$

In this inequality $\vec{v}_n$ is the neighbor vector and ξ must be calculated through statistical techniques.

In the next iteration we change the place of Basic Vector based on the recognized skin areas in the earlier stage. Using this technique leads us to overcome the problem of illumination and race in certain situations. The result would be converged to the skin map, was our first estimate accurate enough.

4 Experimental Results

In our experiment database and masks found in [26] were used. We have applied our algorithm to several images. Most of the processing time of the classifier system is spent in the initial vector calculations which can be very fast by the parallel processing. The iterative technique would be needed when the initial Basic Vector isn't determined appropriately. The system was implemented using Maltab 704 running on a Core 2 Duo @ 2266 MHz processor.

Photo collections usually contain color images taken under changing lighting conditions and also have complex backgrounds. In addition, these images may have variations in color (also, related to race), illumination, and position.

In real situation multiple light sources impinge on the skin. Therefore, for robust skin pixel detection, a dynamic skin color model that can handle the changes was employed.

In Fig. 1 the skin color distribution is shown. By the first set of tests we just used the first step discussed in the last Section. The problem was there that some extra pixels in all pictures were detected. After correctness using the method as you see the results are promising.

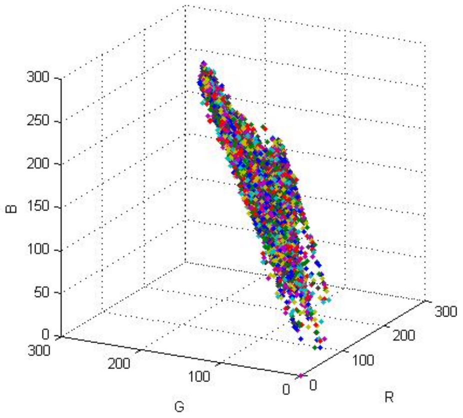


Fig. 1. Skin color distribution

Such an example of using our method is shown in figures 2, 3 and 4. The original image is shown in Fig. 2, whereas Fig. 3 shows the mask image, and Fig. 4 displays the skin area resulted by the developed method.

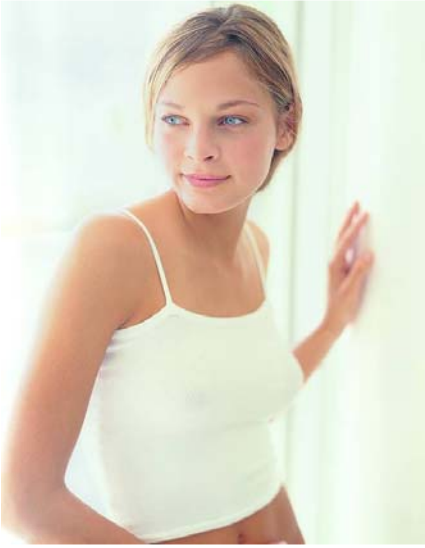


Fig. 2. The original Image

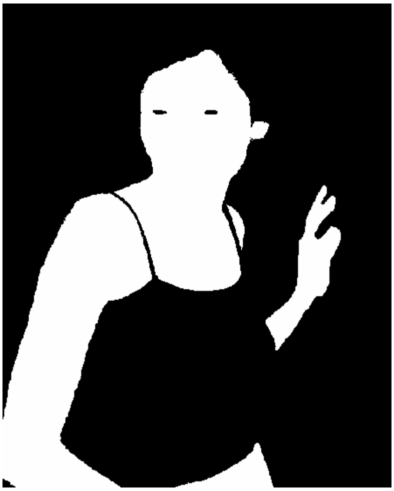


Fig. 3. Mask image

In table 1 we have computed different methods of skin detection. As you see, the results of comparing our two skin detection techniques and also new iterative one in terms of the TP and FP are presented.

As you see in the Table 1 the results can be easily compared. We believe that our new method is an appropriate model and performs well. It is clear that the performance of our detection method is comparable to highly efficient methods mentioned further.

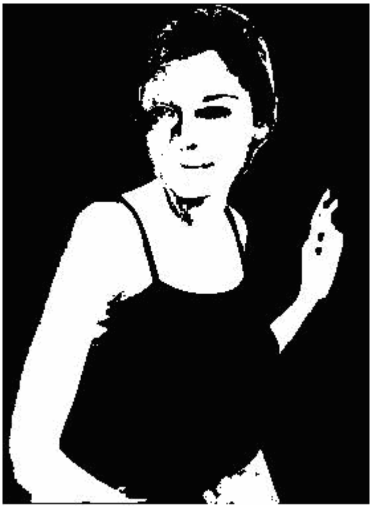


Fig. 4. Detected image

Table 1. Comparison of different methods

Method	TP	FP
Bayes SPM in RGB	80%	8.5%
[Jones and Rehg 1999]	90%	14.2%
Bayes SPM in RGB [Brand and Mason 2000]	93.4%	19.8%
Maximum Entropy Model in RGB [Jedynak et al. 2002]	80%	8%
Gaussian Mixture models	80%	~9.5%
in RGB [Jones and Rehg 1999]	90%	~15.5%
SOM in TS [Brown et al. 2001]	78%	32%
Elliptical boundary model in CIE-xy [Lee and Yoo 2002]	90%	20.9%
Single Gaussian in CbCr [Lee and Yoo 2002]	90%	33.3%
Gaussian Mixture in IQ [Lee and Yoo 2002]	90%	30.0%
Thresholding[Brand- and Mason 2000]	94.7%	30.2%
Our Method		
using inequality (4), (5),	83.3%	~15.6%
using inequality (6)	90.7%	~13.3%
Our iterative method	91.3%	~12.6%

5 Conclusions

Skin detection plays an important role in human motion analysis and face detection. It is widely needed in image processing applications ranging from face detection, face tracking, gesture analysis, content-based image retrieval systems and to various human computer interaction domains.

It is clear that the most popular approaches for skin detection are based on color information. Among the successful skin detection systems those systems that can have easy hardware implementation have gained the most success and our approach offers an implicit mathematical model [25]. This work uses mixed vector space of RGB and HSV and also neighbourhood pixels (vectors). This method overcomes the low accuracy of the methods for skin detection suggested before.

There are two phases, one is based on skin color and the other considers the neighbourhood pixels. In other words the first step determines skinness of each pixel considering HSV and RGB parameters, but another key required for accurate detection is applying a suitable threshold on difference of neighbourhood pixels [25]. In addition some iterative techniques were applied to attain a more efficient result. This can be achieved by changing the vectors in two phases. All in all to cope up with changes in the lighting of pictures taken by camera in different conditions a robust and dynamic method is required.

The proposed method has been implemented. The experimental data found in [26] has been used to check and evaluate this method. The experimental results obtained were encouraging and shows that our approach is superior over the previous one.

References

1. Vezhnevets, V., Sazonov, V., Andreeva, A.: A survey on pixel-based skin color detection techniques. In: Proc. Conference on GraphiCon, Moscow, Russia, pp. 85–92 (2003)
2. Zarit, B., Super, B., Quek, F.: Comparison of five color models in skin pixel classification. In: Proc. Internatonal Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, Corfu, Greece, pp. 58–63 (1999)
3. Kakumanu, P., Makrogiannis, S., Bourbakis, N.: A survey of skin-color modeling and detection methods. Elsevier Science, Oxford (1968)
4. Brand, J., Mason, J.S.: A comparative assessment of three approaches to pixel-level human skin detection. In: Proc. 15th International Conference on Pattern Recognition, pp. 1056–1059 (2000)
5. Terrillon, J.C., Shirazi, M.N., Fukamachi, H., Akamatsu, S.: Comparative performance of different skin chrominance and chrominance spaces for the automatic detection of human in colorimages. In: Proc. 4th IEEE International Conference Face and Gesture Recognition, Grenoble, France, pp. 54–61 (2000)
6. Singh, S.K., Chauhan, D.S., Vatsa, M., Singh, R.: A robust skin color based face detection algorithm. Tamkang Journal of Science and Engineering 6(4), 227–234 (2003)
7. Peer, P., Kovac, J., Solina, F.: Human skin color clustering for face detection. In: Proc. EUROCON Conference on Computer as a Tool, pp. 144–148 (2003)
8. Chen, Q., Wu, H., Yachida, M.: Face detection by fuzzy pattern matching. In: Proc. 5th International Conference on Computer Vision, Cambridge, MA, USA, pp. 591–597 (1995)

9. Schumeyer, R., Barner, K.: A color-based classifier for region identification in video. In: Visual Communications and Image Processing. SPIE, vol. 3309, pp. 189–200 (1998)
10. Soriano, M., Huovinen, S., Martinkauppi, Laaksonen, M.: Skin detection in video under changing illumination conditions. In: Proc. 15th International Conference on Pattern Recognition, Barcelona, Spain, pp. 839–842 (2000)
11. Birchfield, S.: Elliptical head tracking using intensity gradients and color histograms. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition, Santa Barbara, California, pp. 232–237 (1998)
12. Jones, M.J., Rehg, J.M.: Statistical color models with application to skin detection. In: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 274–280 (1999)
13. Gonzalez, R.C., Woods, R.E.: Digital Image Processing, 2nd edn. Prentice Hall, New York (2002)
14. Chai, D., Bouzerdoum, A.: Bayesian approach to skin color classification in YCbCr color space. In: Proc. IEEE Region Ten Conference, Kualu Lumpur, Malaysia, vol. 2, pp. 421–424 (2000)
15. Schiele, B., Waibel, A.: Gaze tracking based on face-color. In: Proc. International Workshop on Automatic Face- and Gesture-Recognition, Zurich, Switzerland, pp. 344–348 (1995)
16. Comaniciu, D., Ramesh, A.: Robust detection and tracking of human faces with an active camera. In: Proc. 3rd IEEE International Workshop on Visual Surveillance, Dublin, Ireland, pp. 11–18 (2000)
17. Jones, M.J., Rehg, J.M.: Statistical color models with application to skin detection. International Journal Computer Vision 46(1), 81–96 (2002)
18. Piirainen, T., Silvén, O., Tuulos, V.: Layered self organizing maps based video content classification. In: Proc. Workshop on Real-time Image Sequence Analysis, Oulu, Finland, pp. 89–98 (2000)
19. Son, L.M., Chai, D., Bouzerdoum, A.: A universal and robust human skin color model using neural networks. In: Proc. International Joint Conference on Neural Networks, Washington, DC, USA, vol. 4, pp. 2844–2849 (2001)
20. Terrillon, J.C., Shirazi, M.N., Fukamachi, H., Akamatsu, S.: Comparative performance of different skin automatic detection of human faces in color images. In: Proc. 4th IEEE International Conference on Automatic Face and Gesture Recognition, Grenoble, France, pp. 54–61 (2000)
21. Yang, M.H., Ahuja, N.: Detecting human faces in color images. In: Proc. International Conference on Image Processing, Chicago, IL, USA, vol. 1, pp. 127–130 (1998)
22. Raftery, A.E., Banfield, J.D.: Model-based Gaussian and non-Gaussian clustering. Biometrics 49, 803–821 (1993)
23. Bouguila, N., Ziou, D.M.I.: Dirichlet-based probability model applied to human skin detection. In: Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, Montreal, Canada, vol. 5, pp. 521–524 (2004)
24. Bouguila, N., Ziou, D., Vaillancourt, J.: Novel mixtures based on the dirichlet distribution: Application to data and image classification. In: Perner, P., Rosenfeld, A. (eds.) MLDM 2003. LNCS, vol. 2734, pp. 172–181. Springer, Heidelberg (2003)
25. Mahmoudi Aznaveh, M., Mirzaei, H.R., Roshan, E., Saraee, M.H.: A New Color Based Method for Skin Detection Based on RGB Vector Space. In: Proc. IEEE Conference on Human System Interaction, Krakow, Poland, pp. 932–935 (2008)
26. Ruiz-del-Solar, J., Verschae, R.: The skindiff algorithm, Web Faces Project (2007), http://agami.die.uchile.cl/skindiff/index.html (accessed September 25, 2008)

A Formal Model for Supporting the Adaptive Access to Virtual Museums

A. Dattolo[1] and F.L. Luccio[2]

[1] Dipartimento di Matematica e Informatica
Università di Udine, Udine, Italy
antonina.dattolo@uniud.it
[2] Dipartimento di Informatica
Università Ca' Foscari Venezia, Italy
luccio@unive.it

Abstract. Zz-structures are particular data structures capable of representing both hypertext data information and contextual interconnections among various chunks of information. We consider an extension of the standard zz-structure model in terms of computational agents. In particular, we propose a multi-agent adaptive system in which users store their personal information inside a zz-structure and agents collaborate in order to support them in the extraction of personalized navigational views, allowing creation of personalized tours, for example tours within virtual museums. The strength of this new model resides in the level of freedom users have for the dynamical choice, based on some present interest or necessity, of their navigational path inside the virtual museum.

1 Introduction

In the past culture was strictly related to a face-to-face model of interaction. With the diffusion of the Web this approach has deeply changed: while previously people would physically visit a museum, now the increased use of hypermedia to support access to museum information has modified this approach, thus allowing users to virtually navigate inside virtual museums. Users may have, however, various backgrounds, goals, preferences, hyperspace experiences, knowledge, thus the system has to match to their needs and support them during navigation [1].

Virtual marketing museums have been developed to promote real physical museums. On the other hand, learning museums are accessed by users that want to learn something while exploring the structured hyperspace with context-adapted narration, i.e. users that aim at interacting with a system that recreates a real life museum tour guided by a real human being [2]. However, virtual museums usually provide only very standard and impersonal guided tours by offering, e.g. replicas of the exhibition experience on-line (see, e.g., the open source toolkits created by Omeka [3]).

The creation of the so called *personalized views*, i.e. displaying of a limited and well defined and personalized sub-portion of the entire hyperspace is something that has already been considered in different settings. Traditional web browsers

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 481–492.
springrlink.com

implement this strategy through bookmarks or personalized site views such as, e.g. in My Yahoo and My Netscape. Some extensions to these examples are adaptive bookmarking systems such as WebTagger [4], Siteseer [5], and PowerBookmarks [6]. Finally, other solutions using information retrieval techniques to collect information for different users having various needs are Web-based information services [7, 8].

Some work towards the application of different techniques to design adaptable interfaces for museum web sites has been already done, e.g. in [1] where changes on the visitor user profile, and thus on the type of presentation of information, and of navigation are allowed. In [9] the Flamenco system is described, an interface for browsing and searching a large amount of information using similarities to sample images. Recently, the Delphi framework of semantic tools and community annotation for museum collections has been presented [10]. Delphi is an Open Source toolkit, that inherits some ideas of [9] and it is based on linguistic analysis tools, that can be used for browsing and searching in diverse museum collections.

Creating personalized views for the navigation inside virtual museum is thus a very interesting issue. In this paper we discuss a formal structure for the visualization of these views. The proposed visualization technique, supported by an agent-based technology, gives a certain level of freedom to the users, allowing them to interact with the system in order to (partially) choose the path to follow during their navigation.

Our system assumes the storage of the views inside zz-structures [11], particular data structures that store both hypertext data information and the contextual interconnections among different information. These structures are very flexible and have been used in different kind of applications, such as modeling, e.g., an information manager for mobile phones [12], bioinformatics workspaces [12, 13], and many others [14, 15]. Zz-structures have also been applied to the creation of an EAD (Encoded Archival Description) a hierarchical organization of archival collections, typically implemented in XML. This EAD finding aid has been transformed into a zz-structure for visualizing archival information.

Observe that, good navigational tools have to: 1) limit the navigational material by identifying a subset of interesting pages; 2) define adequate structures for the items' storage and create personalized user views; 3) define personalized and adaptive navigational paths for the users.

The paper is organized as follows: in Section 2 we first provide a formal description of the structures and views (in particular n-dimensional views) and we show how to use them to create personalized user views in the context of virtual museum tours. In Section 3 we show how to extend this new concept of views in a dynamically changing setting (point 3), i.e. in a setting where users may dynamically interact with the system in order to decide the path to be followed (limited to a restricted neighborhood of their actual position). Finally, we conclude in Section 4 where our on-going research can be directed.

2 Concept Space and Map

In order to define the concept space and map, we need some preliminary definitions of zz-structures and related views.

2.1 Zz-Structures

Zz-structures [11] introduce a new, graph-centric system of conventions for data and computing. A zz-structure can be thought of as a space filled with cells.

Cells are connected together with links of the same color into linear sequences called *dimensions*. A single series of cells connected in the same dimension is called *rank*, i.e., a rank is in a particular dimension. Moreover, a dimension may contain many different ranks. The starting and an ending cell of a rank are called, *headcell* and *tailcell*, respectively, and the direction from the starting (ending) to the ending (starting) cell is called *posward* (respectively, *negward*). For any dimension, a cell can only have one connection in the posward direction, and one in the negward direction. This ensures that all paths are non-branching, and thus embodies the simplest possible mechanism for traversing links.

Formally a zz-structure is defined as follows (see, [15]). Consider an *edge-colored multigraph* $ECMG = (MG, C, c)$ where: $MG = (V, E, f)$ is a multigraph composed of a set of *vertices* V, a set of *edges* E and a surjective function $f : E \rightarrow \{\{u,v\} \mid u,v \in V, u \neq v\}$. C is a set of colors, and $c : E \rightarrow C$ is an assignment of colors to edges of the multigraph. Finally, $deg(x)$ (respectively, $deg_k(x)$) denotes the number of edges incident to x, (respectively, of color c_k).

Definition 1. *A* **zz-structure** *is an edge-colored multigraph* $S = (MG, C, c)$, *where* $MG = (V, E, f)$, *and* $\forall x \in V$, $\forall k = 1,...,|C|$, $deg_k(x) = 0,1,2$. *Each vertex of a zz-structure is called* zz-cell *and each edge a* zz-link. *The set of isolated vertices is* $V_0 = \{x \in V : deg(x) = 0\}$.

An example of a zz-structure related to an art museum is given in Fig. 1. Vertices are paintings, in particular portraits. Normal (red), dashed (green) and thick (blue) lines group, respectively, operas of the same artist: in particular, $\{v_1, ..., v_{11}\}$ and $\{v_{12}, ..., v_{18}\}$ identify, respectively, Van Gogh's and Gauguin's portraits; dashed lines group self-portraits of the two artists ($\{v_1, ..., v_7\}$ of Van Gogh and $\{v_9, ..., v_{15}\}$ of Gauguin); finally thick lines group works of arts in the same museum: v_{12}, v_1, v_3 and v_6, are in the Van Gogh Museum in Amsterdam, while v_2, v_4, v_8, v_9, v_{10}, v_{11}, v_{15}, v_{17} and v_{18} are in Musée d'Orsay in Paris.

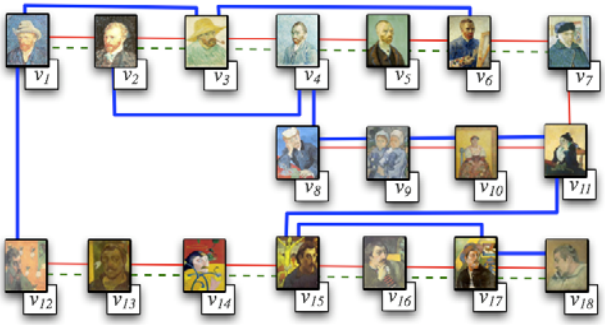


Fig. 1. Van Gogh - Gauguin: self-portraits

Dimensions

An alternative way of viewing a zz-structure is a union of subgraphs, each of which contains edges of a unique color.

Proposition 1. *Consider a set of colors* $C = \{c_1, c_2, ..., c_{|C|}\}$ *and a family of indirect edge-colored graphs* $\{D^1, D^2, ..., D^{|C|}\}$, *where* $D^k = (V, E^k, f, \{c_k\}, c)$, *with* $k = 1, ..., |C|$, *is a graph such that: 1)* $E^k \neq \emptyset$ *; 2)* $x \in V$, $deg_k(x) = 0,1,2$. *Then* $S = \bigcup_{k=1}^{|C|} D^k$ *is a zz-structure.*

Definition 2. *Given a zz-structure* $S = \bigcup_{k=1}^{|C|} D^k$, *then each graph* D^k, $k = 1, ..., |C|$ *is a distinct* **dimension** *of* S.

In Fig. 1, we may identify three dimensions: *artist*, *self-portraits* and *museum*, respectively represented by normal, dashed and thick lines.

Ranks

A rank is in a particular dimension and it must be a *connected* component.

Definition 3. *Consider a dimension* $D^k = (V, E^k, f, \{c_k\}, c)$, $k = 1, ..., |C|$ *of a zz-structure* $S = \bigcup_{k=1}^{|C|} D^k$. *Then, each of the* l_k *connected components of* D^k *is called a* **rank**.

A dimension can contain one (if $l_k = 1$) or more ranks. Moreover, the number l_k of ranks differs in each dimension D^k. In Fig. 1 each of the three dimensions contains two ranks; in particular, the dimension *artist* contains *van-gogh*, and g*auguin*. We note that the example in Fig. 1 is only a fragment of a larger zz-structure; for this reason, for example, in the dimension *museum* we identify only two ranks (related to the *Van Gogh Museum in Amsterdam* and to the Musée d'Orsay in Paris). In reality, this dimension contains all the ranks related to the museums in which the paintings are exhibited (such as the *National Gallery of Art in Washington DC*, the *MacNay Art Museum of San Antonio, Texas*, the *Fogg Art Museum in Cambridge*, etc.).

Given a rank R_i^k, an alternative way of viewing a dimension is a union of ranks: $D^k = \bigcup_{i=1}^{l_k} R_i^k \bigcup V_0^k$.

Head and tail cells

If we focus on a vertex *x*, $R_i^k = ...x^{-2}x^{-1}xx^{+1}x^{+2}...$ is expressed in terms of negward and posward cells of $x : x^{-1}$ is the negward cell of x and x^{+1} the posward cell. We also assume $x^0 = x$. In general $x^{-i}(x^{+i})$ is a cell at distance i in the negward (posward) direction.

Definition 4. *Given a rank* $R_i^k = (V_i^k, E_i^k, f, \{c_k\}, c)$*, a cell x is the* **headcell** *of* R_i^k *iff exists its posward cell* x^{+1} *and it does not exist its negward cell* x^{-1}*. Analogously, a cell x is the* **tailcell** *of* R_i^k *iff exists its negward cell* x^{-1} *and it does not exist its posward cell* x^{+1}*.*

Views

Personalized views are shown to the users, when they choose a vertex as a focus and a set of preferred topics, e.g., type of operas, artists and so on.

The classical ***guided tour*** is a suggested linear path, defined in terms of a specific dimension.

An example of guided tour on dimension *museum* is proposed in Fig. 2; in this case, a user is interested to view the portraits, accessing them museum by museum.

The focus is the Musée d'Orsay, and the related rank, composed by 9 cells, is visualized; referring to Fig. 1, the portraits contained in Fig. 2 are v_2, v_4, v_8, v_9, v_{10}, and v_{11} collocated in the room 35, v_{15} in the room 43, while v_{17} and v_{18} in the room 44. The user can choose to continue his/her tour by clicking on the forward button, or directly selecting one museum among the ranks present in the current dimension (*Van Gogh Museum of Amsterdam*, *National Gallery of Art of Washington DC*, *MacNay Art Museum of San Antonio, Texas*); in this case, the zoomed cell related to the Musée d'Orsay will substituted by the cell related to new museum, selected by the user.

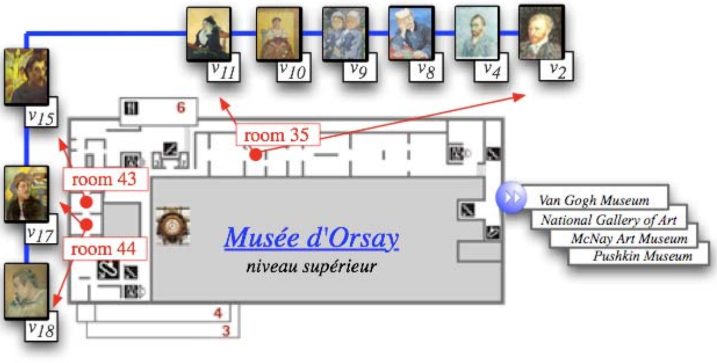


Fig. 2. A guided tour

A guided tour shows one dimension; now we formally present the concept of a 2-dimensional view. $x \in R_{(x)}^a$ denotes that $R_{(x)}^a$ is the rank related to *x* of color c_a.

Definition 5. *Given a zz-structure* $S = \bigcup_{k=1}^{|C|} D^k$*, where* $D^k = \bigcup_{i=1}^{l_k} R_i^k \bigcup V_0^k$*, and where* $R_i^k = (V_i^k, E_i^k, f, \{c_k\}, c)$*, the* ***H-view*** *of size* $l = 2m+1$ *and of focus* $x \in V = \bigcup_{i=0}^{l_k} V_i^k$*, on main vertical dimension* D^a *and secondary horizontal* *dimension* D^b $((a,b \in \{1,...,l_k\})$*, is defined as a tree whose embedding in the plane is a partially connected colored* $l \times l$ *mesh in which:*

- *the central vertex, in position ((m+1),(m+1)), is the focus x,*
- *the horizontal central path (the m+1-th row) from left to right, focused in vertex* $x \in R^b_{(x)}$ *is:* $x^{-g}...x^{-1}xx^{+1}...x^{+p}$ *where* $x^s \in R^b_{(x)}$, *for* $s=-g,...+p(g,p \le m)$,
- *for each cell* x^s, $s=-g,...+p$, *the related vertical path, from top to bottom, is:* $(x^s)^{-g_s}...(x^s)^{-1}x^s(x^s)^{+1}...(x^s)^{-p_s}$*,where* $(x^s)^t \in R^a_{(x^s)}$*,for* $t=-g_s,...,+p_s$ $(g_s,p_s \le m)$.

Intuitively, the *H*-view extracts ranks along the two chosen dimensions. Note that, the name *H*-view comes from the fact that the columns remind the vertical bars in a capital letter H. As example, consider Fig. 3 that refers to the zz-structure of Fig. 1. The chosen dimensions are *self-portraits* and *museum*. The view has size *l*=2*m*+1=5, the focus is v_3, the horizontal central path is $v_3^{-2}v_3^{-1}v_3v_3^{+1}v_3^{+2} = v_1v_2v_3v_4v_5(g,p=2)$. The vertical path related to $v_3^{-1} = v_2$ is $(v_3^{-1})(v_3^{-1})^{+1}(v_3^{-1})^{+2} = v_2v_4v_8(g_s=0,p_s=2)$, that is $v_3^{-1} = v_2$ is the head-cell of the rank as $g_s = 0 < m = 2$.

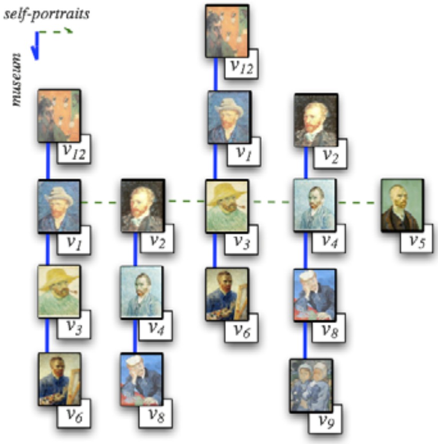


Fig. 3. An H-view related to Fig. 1

We now extend the known definition of *H* to a number $n > 2$ of dimensions. Intuitively, we will build $n-1$ different *H*-views, centered in the same focus, with a fixed main dimension and a secondary dimension chosen among the other $n-1$ dimensions. Formally:

Definition 6. *Given a zz-structure* $S=\bigcup_{k=1}^{|C|}D^k$ *, where* $D^k=\bigcup_{i=1}^{l_k}R_i^k\bigcup V_0^k$ *, and where* $R_i^k=(V_i^k,E_i^k,f,\{c_k\},c)$*, the* ***n*-dimensions *H*-view** *of size l=2m+1 and of* focus $x\in V=\bigcup_{i=0}^{l_k}V_i^k$ *, on dimensions* $D^1,D^2,...,D^n$ *is composed of* $n-1$ *rectangular H-views, of main dimension* D^1*,* $i=1,...,n$ *, all centered in focus x.*

2.2 A-Space and a-Map

A concept space provides an ontology for a given knowledge domain.

Definition 7. *An* **a-space** *a-CS is the representation of a concept space in terms of a multi-agent system composed of five types of agents: concept maps, dimensions, ranks, composite and atomic cells.*

These five agent classes represent five abstraction levels of the concept space. Concept maps know and directly manipulate dimensions and isolated cells; they include concepts and relationships between concepts that are organized in dimensions. Dimensions, uniquely identified by their colors, know and manipulate their connected components, i.e., their ranks. Ranks know and coordinate the cells and the links that connect them; composite cells contain concept maps related to more specific topics, and finally atomic cells are primary entities and directly refer to documents.

Agents can be used in order to model concurrent computations. They are organized as a universe of inherently autonomous computational entities, which interact with each other by sending messages and reacting to external stimuli by executing some predefined procedural skills. Various authors have proposed different definitions of agents. In our setting an agent is defined as follows:

Definition 8. *An* **agent** *is denoted A = (Ts, En, Re, Ac) where*

- *Ts represents its* ***topological structure****;*
- *En=\{η_1,η_2,...\} defines its local* ***environment****;*
- *Re=\{ρ_1,ρ_2,...\} is the finite set of incoming* ***requests****;*
- *Ac=\{α_1,α_2,...\} is the discrete, finite set of possible* ***actions****.*

Ts and *En* represent the passive part of the agent, while *Re* and *Ac* its active part. To give an idea of our agent classes, we define the concept map agent.

Definition 9. *An a-map is a* **concept map** *agent a-map=(Ts,En,Re,Ac) where*

- *Ts=S (see Definition 2.1 and Proposition 1);*
- *En= {dimensions, isolated-cells, colors, ranks, cells, links, ...};*
- *Re={∅}, initially;*
- *Ac={return-colors, return-ranks, return-cells, return-links, check-global-orientation, delete(cell$_1$,...,cell$_n$), ...}.*

dimensions, isolated-cells and the other data of *En* contain information on the structure. The first four actions of *Ac* are internal to the agent and enable it to derive the colors, ranks, cells and links of the zz-structure. These actions are performed by sending querying messages to dimensions and isolated-cells; ranks and used colors are obtained sending a request to the dimension agents, while cells and links references are requested from the dimensions to the rank agents. Other scripts, such as *check-global-orientation* that checks whether local orientation of neighboring cells are consistent, and *delete(cell$_1$, ..., cell$_n$)*, that deletes a chosen set of cells, are used in dynamic operations illustrated in the next section.

3 Displaying and Changing Views

In this section we will show how a session agent may interact with the users in order to first create and display *H*-views, and then neighboring views, i.e. views where the focus is at distance one from the previous one. While displaying these personalized views, the users will also be able to personalize their paths, by deciding to what neighboring view they will move, what dimension they would like to add/remove and so on. During this navigation process, the users can recreate a setting similar to the selected, one by storing the information they find interesting in an album, in order to create a personal, re-usable workspace.

More precisely, assume that *A* is the user-author, *SA* the session agent of *Z*, D^k is the dimension with color c_k, and for simplicity, $R^k_{(x)}$ is the rank cell *x* belongs to.

Whenever *SA* receives from *A* the message *h-view(x, l, D^a, D^b)*, regarding the displaying of an *H*-view, centered at *x*, of size *l*, using main vertical dimension D^a and secondary horizontal dimension D^b, it sends to *x* the message *focus*(*h-view, l, D^a, D^b)*, asking *x* to assume the role of the focus in the visualization of the *H*-view. The visualization operation may be divided into two different steps.

Step 1. *Wake-up of vertices in the horizontal m+1-th row, i.e., $x^{-g},...,x^{-1},x^o = x,x^{+1}...x^{+p}$. The focus activates rank $R^b_{(x)}$ that propagates the request to the vertices in the m+1-th row. These vertices are woken-up and are visualized horizontally, i.e., as* horiz *(see Fig. 4).*

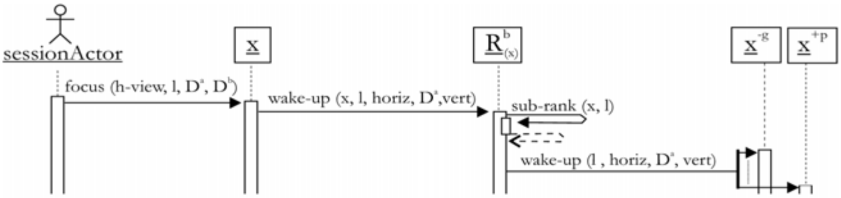

sessionActor
x
R^b_(x)
x^-g
x^+p
focus (h-view, l, D^a, D^b)
wake-up (x, l, horiz, D^a,vert)
sub-rank (x, l)
wake-up (l , horiz, D^a, vert)

Fig. 4. Computation related to D^b

Step 2. *Propagate the wake-up request from the vertices in the m+1-th row, to their rank of color c_a, and from the rank to its vertices at distance at most m from $x^{-g},...,x^{-1},x^o = x,x^{+1}...x^{+p}$, respectively. These vertices are visualized vertically, i.e., as* vert *(see Fig. 5).*

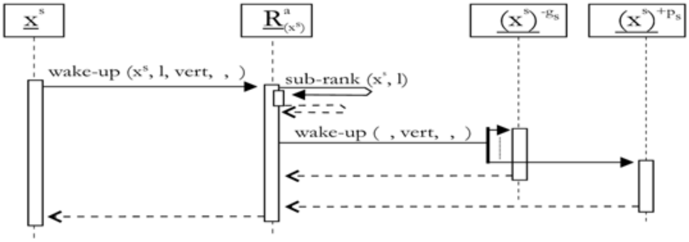


Fig. 5. Computation related to D^a

Another interesting issue is the visualization of a neighboring view, e.g., a view where the focus shifts by one position.

Agents may either apply the *sleep* procedure that turns off all the cells that have been woken up, and then start from scratch a new *wake-up* procedure, or try to optimize the computation and turn off only part of the cells and turn on part of others. E.g., whenever the focus moves vertically, the simplest procedure consists of turning off the whole old view and turning on the new one as it follows:

```
Posward-Vert-Shift-h-view(($x^0$), $(x^0)^{+1}$, $l$, $D^a$, $D^b$)
1. send-now(sleep(($x^0$), $l$, horiz, $D^a$, vert)) to $R^b_{(x^0)}$
2. send-now(focus(h-view, l, $D^a$, $D^b$)) to $(x^0)^{+1}$
```

Note that the focus is shifted from x^o to $(x^o)^{+1}$: the cells that do not change are the *(l-1)* placed vertically.

Now, we assume that the focus moves horizontally, i.e., from x to x^{+1}, as, e.g., in Fig. 3 from $x = v_3$ to $x^{+1} = v_4$ thus obtaining the new view of Fig. 6.

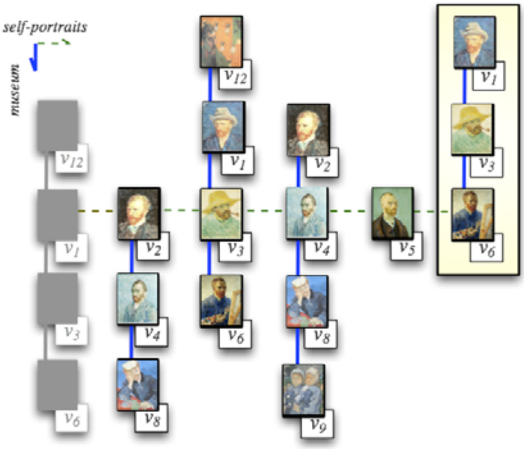


Fig. 6. Horizontal shift of focus

Posward-Horiz-Shift-h-View(x, x^{+1}, l, D^a, D^b)

```
1. if g=(l-1)/2 <vertex $x^{-((l-1)/2)}$ exists>
2. then send-now(sleep((l,horiz,$D^a$,vert)) to $x^{-g}$ <the (l-1)/2 -th column on the left is turned off>
3. if p=(l-1)/2 <vertex $x^{+((l-1)/2)}$ exists>
4. then send-now(wake-up((l, horiz, $D^a$, vert)) to $x^{+(p+1)}$ <the ((l-1)/2)+1-th column on the right is turned on>
5. update-view
```

In this case, an optimized procedure, called Posward-Horiz-Shift-h-view(x, x^{+1}, l, D^a, D^b), is activated by *SA*. The procedure has to turn off the vertical column related to vertex $x^{-(l-1)/2}$ (lines 1-2), if it exists, and has to turn on the column related to vertex $x^{+((l-1)/2)+1}$, given that vertex $x^{+((l-1)/2)}$ exists (lines 3-4). The first operation is analogous to the one of Fig. 5 with vertex x^s substituted by vertex x^{-g} and the script *wake-up* with the script *sleep*, with the same arguments. The second operation (lines 3-4) is similar but it has to be applied to vertex $x^{+(p+1)}$ using the script *wake-up*. Finally, *SA* has to update its view (line 5).

Displaying n-Dimensions H-Views

In this section we will show how to extend the technique of the previous section and thus how the *SA* may display n-dimensions *H*-views.

The technique is very similar. For example, for the *n*-dimensions *H*-view the *SA* contacts the focus *x* with message *focus*(*n-dim-h-view,l,D^1,...,D^n*). Since an n-dimensions *H-view* on main dimension D^1 and secondary dimensions $D^2, D^3, ..., D^n$ is composed of $n-1$ different *H-views* on main dimension D^1 and secondary dimension D^i, for i=2,…,n, the visualization operation is divided into two different steps.

Step 1. *Wake-up the vertices in the secondary dimensions of each H-view centered in x, i.e., of dimensions $D^2, D^3, ..., D^n$. Focus x multicasts message wake-up(x,l,plane,D^1,vert) to ranks $R^i_{(x)}$ (i=2,...,n). The output of this script is the computation of sub-ranks (x, l), i.e., the sub-ranks of size l centered in x in dimension D^i(x), for i=2,3,...,n. Then, each $R^i_{(x)}$, i=2, 3, ..., n sends message wake-up(x, l, plane, D^1, vert), to all its vertices.*

Step 2. *Similar to the one of the 2-dimensions case respectively applied to the vertices in ranks $R^i_{(x)}$, for i=2,3,..., n.*

The technique for the *n*-dimensions *I*-view is very similar. In the case of the *3*-dimensions extended *H*-view the SA contacts the focus x with message *focus (3-ext-h-view, l, D^1, D^2, D^3).* The focus, as in a *2*-dimensions H-view, activates rank $R^3_{(x)}$ in order to compute a 2-dimensions H-view with main dimension D^2 and secondary dimension D^3. The activated vertices x^{-g},…,x^{+p} of $R^3_{(x)}$, while executing the standard second step of the H-view wake-up procedure, will also act as focus and activate a new procedure for a new 2-dimensions H-view of size l on main dimension D^1 and secondary dimension D^2 .

4 Conclusion

In this paper we have proposed a multi-agent adaptive system to support tours of virtual museums. We have proposed an extended zz-structure model and we have shown how this new system can be used to display personalized user views and to create personalized and navigational adaptive paths for users. As future work we want to concentrate on the study of good techniques to identify the collection of "interesting" pages for specific users, based on their preferences and needs.

References

1. Paternò, F., Mancini, C.: Effective levels of adaptation to different types of users in interactive museum systems. J. Am. Society Information Science 51(1), 5–13 (2000)
2. Bowen, J.P., Filippini-Fantoni, S.: Personalization and the web from a museum perspective. In: Proc. Intern. Conference on Museums and the Web 2004: Archives & Museum Informatics, Washington, DC, pp. 63–78 (2004)
3. `http://omeka.org/` (accessed April 6, 2009)
4. Keller, R.M., Wolfe, S.R., Chen, J.R., Rabinowitz, J.L., Mathe, N.: A bookmarking service for organizing and sharing URLs. Computer Networks & ISDN Systems archive 29(8-13), 1103–1114 (1997)
5. Rucker, J., Polanco, M.J.: Siteseer: personalized navigation for the web. Communications of the ACM Archive 40(3), 73–75 (1997)
6. Li, W.S., Vu, Q., Agrawal, D., Hara, Y., Takano, H.: PowerBookmarks: a system for personalizable web information organization, sharing, and management. Computer Networks 31(11-16), 1375–1389 (1999)
7. Billsus, D., Pazzani, M.J.: Adaptive web site agents. Journal Agents & Multiagent Systems 5(2), 205–218 (2002)
8. Brusilovsky, P., Tasso, C.: User modeling for web information retrieval. Preface to special issue of User Modeling and User Adapted Interaction 14(2-3), 147–1571 (2004)
9. Yee, K.P., Swearingen, K., Li, K., Hearst, M.: Faceted metadata for image search and browsing. In: Proc. ACM Conference on Computer-Human Interaction, Florida, pp. 401–408 (2003)
10. Schmitz, P.L., Black, M.T.: The Delphi toolkit: enabling semantic search for museum collections. In: Proc. Intern. Conference on Museums and the Web 2001: Archives & Museum Informatics (2008), `http://www.informalscience.org/research/search/all?order_by=date&page=6` (accessed May 15, 2009)
11. Nelson, T.H.: A cosmology for a different computer universe: data model mechanism, virtual machine and visualization infrastructure. J. Digital Information 5(1), no. 298 (2004)

12. Moore, A., Goulding, J., Brailsford, T., Ashman, H.: Practical applitudes: case studies of applications of the ZigZag hypermedia system (2004), `http://portal.acm.org/citation.cfm?id=1012851` (accessed April 6, 2009)
13. Moore, A., Brailsford, T.: Unified hyperstructures for bioinformatics: escaping the application prison. J. Digital Information 5(1), no. 254 (2004)
14. Canazza, S., Dattolo, A.: Open, dynamic electronic editions of multidimensional documents. In: Proc. European Conference on Internet and Multimedia Systems and Applications, Chamonix, France, pp. 230–235 (2007)
15. Dattolo, A., Luccio, F.L.: Formalizing a model to represent and visualize concept spaces in e-learning environments. In: Proc. 4th WEBIST Intern. Conference, Funchal, Madeira, Portugal, pp. 339–346 (2008)

3D Molecular Interactive Modeling

M. Essabbah, S. Otmane, and M. Mallem

University of Evry, Evry, France
mouna.essabbah@ibisc.fr

Abstract. Initially, genomic sequences have been known by their linear form. However, they have also a three-dimensional structure, which can be useful for genomes analysis. This 3D structure representation brings a new point of view for the sequences analysis. Therefore, several studies have described the design of software for 3D molecular visualization. The search that we have carried out enabled us to see the various modeling systems and their principles, and we have to classify them in two great families: the class of the visualization systems and the class of 3D interactive modeling. Each one of these classes is divided into two common sub-classes: systems based prediction and systems based 3D data.

1 Introduction

The 3D molecular modeling is a new technology that is increasingly attracting scientists' interest. Its ability to simulate natural phenomena that are not exploitable experimentally offers great potential and opens doors to new research in the field. As a result, a large number of 3D modeling systems have emerged trying to be more precise and promising. However, most software for 3D molecular modeling is based on predictive methods. These methods often use local conformation tables – generated by statistics relative to local biological experiments – which restricts the 3D rending. In parallel, the advances made in the biology field and in 3D modeling, have made possible the implementation of 3D molecular modeling applications increasingly complex, dedicated to the study of molecular structures and molecular dynamics analysis. However 3D molecular modeling presents seve- ral scientific problems that are still the subject of intense researches, in particular on the fundamental interests of this modeling and numerical simulation, the accuracy of 3D models, and so on.

This article presents a recent overview on 3D molecular modeling [10]. It is divided into six sections: we begin with a focus on the importance of molecular spatial structure. The purpose of modeling is emphasized in the third section. Then, we will present different 3D molecular modeling systems. We conclude this paper with a short discussion.

2 Importance of the Molecular Spatial Structure

Initially scientists were interested in sequence analysis for the study of its rich syntax. Gradually (between 1984 and 1987) molecular biology lived a fulgurating progress of its technical means, which leads to automation and an increasingly

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 493–504.
springrlink.com

advanced and refined miniaturization. On the other hand, genomics allowed nowadays a global and complete approach for sequence analysis. It is no longer limited to the molecules study, but it also includes the study of the relationships between these elements, which causes the dynamic aspect of the whole. More generally, 3D visualization provides an overall view of the studied molecule. It allows modeling phenomenon or simulates a biological mechanism.

Currently, molecular modeling is an essential research area. It helps scientists to develop new drugs against diseases in general and particularly serious ones such as AIDS and cancer. This sector assists also the genomes analysis. As the interest that was accorded to it is extremely important, many researchers became specialized in molecular modeling.

2.1 When Biology Becomes Molecular

There are many researches in the literature that link molecular modeling in its biological context, being based on three-dimensional models. These studies present various concepts used in the development of models for biological structures (DNA, RNA, proteins, etc.). In the chemistry department at the University of Pennsylvania, Zou has studied the structure and dynamics of a three-dimensional protein (*Amphiphilic, Metalo-Porphyrin-Binding Protein*) via molecular dynamics simulations [37]. The simulation results match with the available experimental data, describing the structures in a lower resolution and to a limited size. In addition, a Polish group is working on forecasting (high resolution) of three-dimensional RNA structures (low resolution) to answer questions from the RNA molecular biology [29]. Their strategy is based on a program called the 3D-RNApredict, which implements and converts various structural data (RNA secondary structure, details of RNA motifs, experimental data) to create the input for the CYANA program (torsion-angle dynamics algorithm, TAD), which provides a fast engine for the 3D RNA structure calculation. The resultant RNA structure is refined using the X-PLOR program. The molecular structures are more interesting because of their complexity. Indeed, the molecular representation helped to make the connection between molecular complexity and its impact on the probability of discovering new drugs [14]. Structural analysis of Mu DNA transposition was successful using 3D reconstruction of images obtained by scanning transmission electron microscopy (TIGE) [36]. Moreover, a three-dimensional structure has served as a model for the replication study based on curl degree analysis along the DNA axis. The structure was built by cryo-electron microscopy and simple-particle reconstruction techniques [12]. The same technique has been adopted for the reconstruction of a three-dimensional complex DNA-protein [1]. There are also 3D reconstructions based on con-focal microscopy scanning laser [22]. Those researches represent a rich structural basis for different areas such as the biochemical function, the study of various phenomena (replication, transcription), etc.

2.2 Molecular Biology Is Rewarding

The award of several Nobel laureates promotes the importance of molecular modeling, particularly in the period between 1958 and 1969 that was successful for molecular biology. Indeed, the awarding of Nobel Prizes to Linus Pauling for his work on the structure of the chemical bond, and Watson and Crick for the DNA's double helix structure was recognition of the power of molecular models. Since this, six Nobel Prizes in physiology and medicine, and three prizes in chemistry were awarded in the molecular biology field, almost a Nobel Prize a year since 2000. The interest in molecular biology has grown steadily over the years until now.

3 Contributions of Molecular Modeling

Molecules play a key role in cellular processes and thus in life preservation. Therefore, it is essential to have a little knowledge of their behavior *in vivo*. This knowledge will enable us to better understand the biological phenomena and interpret them. However, the molecules are often not accessible to experimental studies because of their lack of stability and the difficulty in reproducing them and assembling them in their natural conformation *in vivo*. Thus, it was essential to develop an *in silico* approach to the problem. Therefore, the idea of creating molecular models is born. These models allow a better understand of the phenomena studied. The first models that emerged were the material models, such as those assembled from the ubiquitous wooden ball-and-stick and space-filling model kits. Molecular models have been used for over 125 years to represent chemistry rules in a very simple way (say, the DNA double helix of Watson and Crick [34]). Research is now turning to computer models and simulations, which leads to the design of molecules analyzing algorithms. Molecular modeling has become an independent science, adopting a set of techniques to model and imitate the molecules behavior.

***Contributions of Informatics in Biology*:** The development of the experience computerization has changed significantly the relationship between traditional theory and experiment. On the one hand, computer simulations have increased the requirement for models accuracy. Indeed, some tests are difficult to do just by the theoretical model; others were not even available in the past. Therefore, the simulation "brings to life" models, revealing critical properties and providing suggestions for improving them. On the other hand, the simulation has become an extremely powerful tool not only for understanding and interpreting experiments at the microscopic level, but also to study areas which are not accessible experimentally, or which involve very expensive experiments.

4 Three-Dimensional Structures: Visualization and Interaction

The 3D molecules visualization has always been an important chemistry chapter. As a result, molecular modeling has become a growing discipline, benefiting from the rapid development of new technologies. In addition, the 3D molecular structures analysis is a research field more mature than sequence analysis. Since the 70's beginning, it was essential to list the coordinates of macromolecules crystal structures such as protein (Protein Data Bank, PDB). Moreover, visualization of a three-dimensional structure is undeniably one of the first tools developed for structures analysis, but also one of the first tests that do wish biologist.

4.1 Three-Dimensional Modeling Approaches

DNA's 3D visualization is based mainly on two approaches. First, the observatory approach consists in reproducing the molecule's 3D model from real 3D data (eg. crystallographic data), which is not fairly well accepted in the case of DNA, because current techniques for the experimental study of DNA's 3D structure (crystallography, cryo-electron microscopy, AFM microscopy, etc.) present many limitations (size, molecule's deformations, etc.) and are very costly in time and price. Besides, most modeling techniques concerns especially proteins thanks to a great 3D data bank (PDB). On the other hand, the predictive approach, for molecular 3D modeling, aim to predict approximately the molecule's 3D structure from its textual sequence thanks to its spatial conformation model (obtained by statistical methods on small DNA's fragments). The two approaches remain considerably easier to implement, cheaper and faster than the experimental methods in vitro.

4.2 Molecular Visualization Systems

Due to the evolution of the molecular 3D visualization/analysis, several three-dimensional molecular viewers, simple and freely available, have emerged (say, ViewMol, MolMol, PyMol (Fig. 1), RasMol (Fig. 2), etc.).

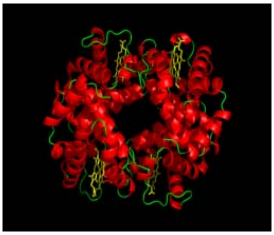


Fig. 1. Hemoglobin structure by PyMol

HIV Protease + Glaxo Wellcome Inhibitor
Roger Sayle with RasMol, 1995

Fig. 2. HIV Protease by RasMol

Other molecules viewers are online for immediate structure reconstruction (CBS-Metaserver, GENO3D, Swiss-Model, Biomer, etc.). Most of these systems are based on observatory approach, described before. Only few of them are based on the predictive approach used by the interactive systems.

4.3 Interactive Systems for Molecular Modeling

Gradually the visualization interests no longer stops at the molecules observation, but it extends to the structure analysis and functionality interpretation of molecules by their 3D structure, so new interactive viewers appear. In this section, we present a non-exhaustive list of these interactive systems.

Early molecular graphics: C. Levinthal at MIT built the first system for the interactive display of molecular structures in the mid-1960s. This system allowed a user to study interaction between atoms and the online manipulation of molecular structures. The speed and orientation of a molecular structure can be controlled by a globe-shaped device (a kind of trackball) (Fig. 3 shows the VR interface used) [21].

Moreover, it was possible to select from a menu, choose an object, or zoom into important parts of the molecule using a light pen on the computer display.

VMD (Visual Molecular Dynamics) [17] is a molecular visualization program for displaying, animation and analysis of large molecular systems using three-dimensional graphics (Fig. 4). It can also read standards PDB files and display structures. VMD provides a wide variety of methods to display and color molecules. It can also be used to animate and analyze the trajectory of molecular dynamics (MD) simulation. Its special feature is that it can be used as graphical interface for MD external program, displaying and animating a molecule that is simulated on a remote computer.

Fig. 3. The display terminal shows the molecule structure in wireframe fashion

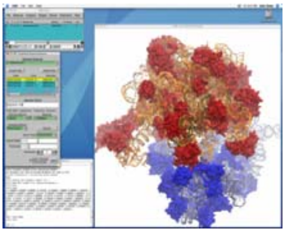


Fig. 4. Screenshot of VMD 1.8.5

ERNA-3D (Editor for 3D RNA) [26] is a molecular modeling system that has been specially developed for models creation of large RNA molecules. However, it is possible to manipulate proteins and other molecules. The difference between ERNA-3D and conventional molecular modeling systems is the ability to edit a molecule portion in a dynamic and realistic way (Fig. 6). ERNA-3D allows generation of various molecular abstraction degrees (one display per cylinder, tube or set of balls and sticks (Fig. 5)).

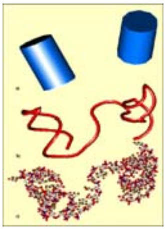


Fig. 5. Cylinder, tube, sets of balls and sticks representation

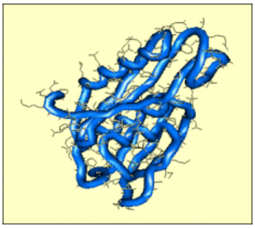


Fig. 6. Complex molecule representation by ERNA-3D

ADN-Viewer (DNA 3D modeling and stereoscopic visualization) [16] is particularly interested in DNA's spatial distribution. The system offers the 3D reconstruction of DNA's structure, which is based on the predictions approach (described in the section 4.1). ADN-Viewer offers several 3D DNA sequences representations;

the genomics representation (see Fig. 7) and the gene representation (see Fig. 8). It also provides the opportunity to explore and manipulate chromosomes' structures thanks to stereoscopic visualization and Virtual Reality (VR) devices (Fig. 9).

Researchers have developed other interactive systems [13, 24, 31] for molecular analysis employing immersive technologies.

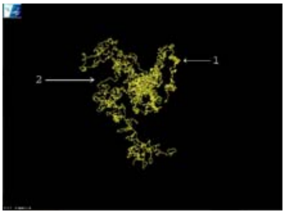


Fig. 7. Chromosome 3D model: compact (zone 1) and relaxed (zone 2) DNA areas

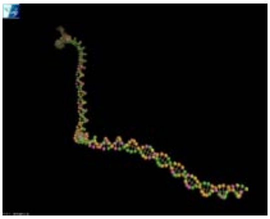


Fig. 8. Each colored sphere corresponds to a nucleotide on ADN-Viewer

Fig. 9. Stereoscopic visualization of chromosomes and immersive navigation with ADN-Viewer

3DNA (3D nucleic acid structures analysis and reconstruction) [23] was identified as American software for 3D nucleic acids structure analysis, reconstruction and visualization (see Fig. 10). 3DNA can locally handle double helices non-parallel and parallel, simple structures, triplex, and other quadruplex motifs found in complex DNA and RNA structures and this from a PDB file coordinates. The program uses also the predictive approach to build the structure. Tools are provided to locate base pairs and regions in a helical structure and to reorient structures for an effective visualization. This program can also handle helical regular models based on X-ray diffraction measurements of various repetition levels.

AMMP-Vis (a virtual environment for collaborative molecular modeling) [8] is an immersive system that offers to biologists and chemists the possibility of manipulating molecular models through a natural gesture. It allows receiving and displaying real-time molecular dynamics simulation results. It allows adapted views sharing and provides support for local and remote collaborative research (Fig. 11). It is based on the molecular visualization system AMMP described in the next section.

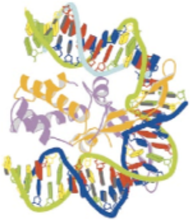


Fig. 10. 3DNA: local DNA's 3D modeling by a predictive method

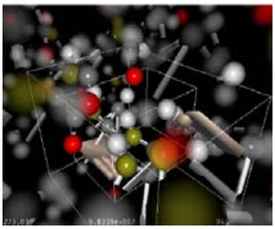


Fig. 11. Example of collaborative research

Fig. 12. 2D molecule manipulation (on the table) and 3D visualization (on the wall)

Other systems [5, 25, 32] take advantage of the benefits of collaboration in the modeling, visualization and analysis of molecules.

NAVRNA (Interactive system for structural RNA) [2] is an interactive system to visualize, explore and edit the RNA molecules. NAVRNA would visualize at the same time the three-dimensional structure (3D) projected onto a white wall and the secondary structure (2D) also projected on a table (Fig. 12). Both shows are strongly linked. It is a multi-surface collaborative system for RNA analysis.

Augmented reality with auto-made tangible models for molecular biology applications [11]: The evolution of auto-made computer technology (3D Printing) can now allows the production of physical models such as molecules and biological complex sets. It presents an application that demonstrates the use of tangible auto-made models and augmented reality (Fig. 13) for research and education in molecular biology, and to improve the environment for scientific collaboration and exploration. They have adapted an augmented reality system to allow 3D virtual representations to be overlaid on a real molecular model.

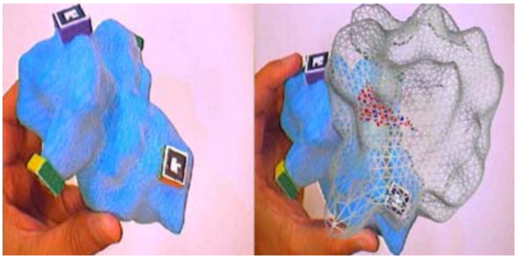


Fig. 13. Ribosome with (right) and without (left) RA

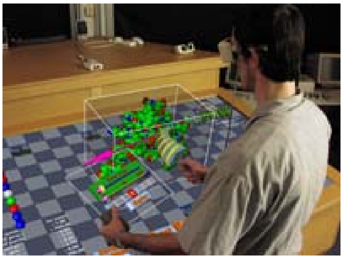


Fig. 14. MolDRIVE handling a 3D molecule

User can easily change this superposition of information, switching between different molecule representations, the display of molecular properties such as electrostatics, or dynamic information. The physical model provides a powerful and intuitive interface for manipulating computer models, improving the interface between the human intention, the physical model, and the computing activity.

MolDRIVE [18] represents a virtual environment for visualization and steering of real-time molecular dynamics simulations. MolDRIVE is also an interface to several MD simulation programs, which run in parallel on remote supercomputers. It uses DEMMPSI and also GROMACS (detailed in the next section). Remote simulations allow the high performance graphics workstation, to fully concentrate on the visualization part of the process. Therefore, larger systems can be visualized. It provides real-time visual force feedback by displaying a deformable spring stylus whose shape communicates a sense of the magnitude of force applied to a particle by a user (shown in Fig. 14).

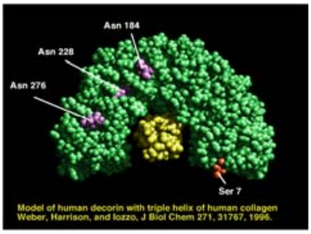
Recent research has highlighted the importance of using VR in 3D molecular modeling, combining immersive environment (efficient visual representation) and haptic devises (interactive manipulation) to offer a real-time sense of molecules flexibility, in education [30], ligand modeling [3,19], protein dynamic [6] and docking molecules [34].

4.4 Molecular Dynamics Modeling

Over the last five years, a significant increase has affected the number of publications describing an accurate and reliable molecular dynamics simulation, and especially for nucleic acids. Cheatham and Young wrote an article [9] describing successes limits and prospects in this area. The researchers are now studying the molecules behavior by spatial visualization and structures analysis. In reality the three-dimensional structure inform us about the molecule functionality. This approach allows a global view of the molecule spatial structure as well as possibilities of interaction with other molecules. In his book, Leach [20] describes the molecular modeling principles and applications. He asserts that modeling involves three essential steps; the first is the description of the intra and inter-molecular interactions systems. The second step is to calculate the molecules dynamics. Finally, the third step is to analyze these calculations to extract the properties and check the well functioning of the system. More generally, the molecular dynamics modeling allows calculating a particles system evolution over time. Therefore, the growing interest in the molecular dynamics simulation promoted the emergence of many modeling software. The description of some of the most popular is given in the following.

AMBER (Assisted Model Building with Energy Refinement) is a package developed from a program that was created in the end of the 70's. AMBER applies molecular mechanics, normal way analysis, molecular dynamics and calculating free energy to simulate the structural molecules and energetic properties. It now includes a program group representing a number of powerful tools of modern informatic chemistry, focused on the molecular dynamics and free calculations of proteins energy, nucleic acids and carbohydrates [7].

AMMP [15] is a program that models complete and modern dynamics and molecular mechanics. It can handle small molecules and macromolecules including proteins, nucleic acids and other polymers. In addition to the standard common features molecular modeling software, AMMP has a potential flexible choice and a simple and powerful capability to manipulate molecules and analyze various energy limits. A main advantage over many other programs is that it is easy to present the non-standard links between polymer unusual ligands, or non-standards residue. Furthermore, it is possible to add the hydrogen atoms missing and implement



Model of human decorin with triple helix of human collagen
Weber, Harrison, and Iozzo, J Biol Chem 271, 31767, 1996.

Fig. 15. Decorin model with fibrille collagen by AMMP

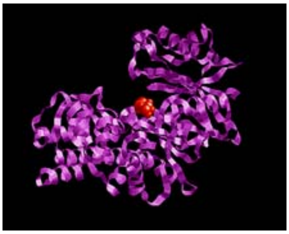


Fig. 16. Human glucokinase mo-del with glucose by AMMP

partial structures, which is difficult for many other modeling tools. In Fig. 15 and Fig. 16, two examples of AMMP simulations are shown.

CHARMM (Chemistry at Harvard, Molecular mechanics) is a package for molecular dynamics simulation and analysis [27]. The CHARMM project includes a developer network around the world working with Karplus and his group at Harvard University to develop and maintain the program CHARMM.

GROMACS is a set of codes running molecular dynamics, initially developed by Berendsen [4]. It also has a lot of analysis tools including a trajectories viewer. It was initially designed for biochemical molecules such as proteins and lipids that have rich and complex interactions. However, since GROMACS is extremely fast to calculate the interactions [33] many research groups use it for calculation on polymers.

TINKER is a complete and free program (with FORTRAN) for mechanics and molecular dynamics, especially polypeptides. TINKER has the option to use any common parameters set such as AMBER / OPLS, CHARMM22, MM2, MM3, ENCAD, MMFF, and the set specific to him. It implements an algorithms variety such as a new method for recognition the geometry of distance, an optimizer own local truncated Newton (TNCG) [28] and many other benefits.

5 Conclusions

The importance of the molecules spatial structure is clearly identified, and therefore its contributions in chemistry and biology research are considerable. So researchers have tried, through different ways, to study the molecular structure, whether static or dynamic, isolated or within a molecular complex. The main issue was to create models (either material or computing models) to imitate the studied molecules behavior.

In this paper we are particularly interested in computing molecular models and the bioinformatics contribution. The observation of these models can be local or global, although it is based on two main approaches: the observatory approach, which uses real 3D data, and predictive approach, which is based on approximations previously established by statistical study. The increasing interest accorded to this area has generated several systems for molecular analysis. We classified them into two main categories: visualization systems and interactive systems. Concerning the visualization systems, they offer a different molecule view, local or global. This type of observation can help biologists to have a first idea of the spatial architecture of certain molecules. However, these viewers are generally based on the observatory approach, so they are limited to a certain molecules category (having real 3D data). Moreover, these viewers are not powerful enough (important computational time) to model large molecules. Regarding interactive system, we saw that they offer a manipulation of the visualized molecules. Thus biologists can act on the structure to improve it tanks to their expertise. On the other hand, the expert may simply interact with the model to look over some assumptions. These systems can be immersive, collaborative and multimodal. However, exploration and interaction in existing molecular modeling virtual environments are often simply basic and limited to a single user. This is insufficient to really explore the ability of multimodal interaction in VR to improve the molecular analysis. In addition, scientists are often reticent to adopt these systems. Therefore, the main challenge is to show that a synergy between preferment visual modeling and immersive human/machine interactions is an interesting approach to help the biologist to analyze and to understand life complex phenomena.

References

1. Abu-Arish, A., et al.: Three-dimensional reconstruction of agrobacterium VirE2 protein with single-stranded DNA. Journal of Biological Chemistry 279(24), 25359–25363 (2004)
2. Bailly, G., Auber, D., Nigay, L.: From visualization to manipulation of RNA secondary and tertiary structures. In: Proc. 4th IEEE Conference on Information Visualization, Washington, DC, pp. 107–116 (2006)
3. Bayazit, O.B., Song, G., Amato, N.M.: Ligand binding with OBPRM and haptic user input. In: Proc. IEEE Intern. Conference on Robotics and Automation, Seoul, Korea, pp. 954–959 (2001)
4. Bekker, H., et al.: Gromacs: a parallel computer for molecular dynamics simulations (1993), http://www.gromacs.org/component/option,com_wrapper/Itemid,192/ (accessed April 15, 2009)
5. Bourne, P.E., et al.: A prototype molecular interactive collaborative environment (MICE). In: Altman, R., Dunker, K., Hunter, L., Klein, T. (eds.) Pacific Symp. on Biocomputing, pp. 118–129 (1998)
6. Bidmon, K., et al.: Time-based haptic analysis of protein dynamics. In: Proc. 2nd IEEE Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, pp. 537–542. IEEE Computer Society, Washington (2007)
7. Case, D.A., et al.: The Amber biomolecular simulation programs. J. Computational Chemistry 26, 1668–1688 (2005)
8. Chastine, J.W., et al.: AMMP-Vis: a collaborative virtual environment for molecular modeling. In: Proc. ACM Symposium on Virtual Reality Software and Technology, Monterey, CA, pp. 8–15. ACM Press, New York (2005)
9. Cheatham III, T.E., Young, M.A.: Molecular dynamics simulations of nucleic acids: successes, limitations and promise. Biopolymers Nuc. Acid Sci. 56(4), 232–256 (2001)
10. Essabbah, M., Otmane, S., Mallem, M.: 3D molecular modeling: from theory to applications. In: Proc. IEEE Conference on Human System Interaction, Cracow, Poland, pp. 350–355 (2008)
11. Gillet, A., et al.: Tangible interfaces for structural molecular biology structure. Structure 13(3), 483–491 (2005)
12. Gomez-Lorenzo, M.G., et al.: Large T antigen on the simian virus 40 origin of replication: a 3D snapshot prior to DNA replication. The EMBO Journal (2003), doi:10.1093/emboj/cdg612
13. Haase, H., Strassner, J., Dai, F.: VR techniques for the investigation of molecule data. Computers & Graphics 20(2), 207–217 (1996)

14. Hann, M.M., Leach, A.R., Harper, G.: Molecular complexity and its impact on the probability of finding leads for drug discovery. J. Chem. Information and Comp. Sci. 41(3), 856–864 (2001)
15. Harrison, R.W.: Integrating quantum and molecular mechanics. J. Computational Chemistry 20, 1618–1633 (1999)
16. Hérisson, J., Gherbi, R.: Model-based prediction of the 3D trajectory of huge DNA sequences interactive visualization and exploration (2001), http://portal.acm.org/citation.cfm?id=791313 (accessed April 15, 2009)
17. Humphrey, W., Dalke, A., Schulten, K.: VMD - visual molecular dynamics. J. Molecular Graphics 14, 33–38 (1996)
18. Koutek, M., et al.: Virtual spring manipulators for particle steering in molecular dynamics on the responsive workbench. In: Stürzlinger, W., Müller, S. (eds.) Proc. Intern Workshop on Virtual Environments, Barcelona, Spain. ACM International Conference Proceeding Series, vol. 23, pp. 53–62. Eurographics Association, Aire-la-Ville (2002)
19. Lai-Yuen, S.K., Lee, Y.: Interactive computer-aided design for molecular docking and assembly. Comput. Aided Des. App. 3(6), 701–709 (2006)
20. Leach, A.R.: Molecular modeling: principles and applications, p. 744. Prentice Hall (Pearson Education), Upper Saddle River (2001)
21. Levinthal, C.: Molecular model-building by computer. Sci. Am. 214(6), 42–52 (1966)
22. Liu, S., Weaver, D.L., Taatjes, D.J.: Three-dimensional reconstruction by confocal laser scanning microscopy in routine pathologic specimens of benign and malignant lesions of the human breast. Histochem. Cell Biol. 107(4), 267–278 (1997)
23. Lu, X.J., Olson, W.K.: 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. Nucleic Acids Research 31(17), 508–512 (2003)
24. Marchese, F.T.: A stereographic table for bio-molecular visualization. In: Proc. 6th Intern. Conference on Information Visualization, pp. 603–607. Computer Society, Washington (2002)
25. Mueller, F., et al.: A new model for the three-dimensional folding of escherichia coli 16 S ribosomal RNA. The topography of the functional centre. J. Mol. Biology 271, 566–587 (1997)
26. Patel, S., Mackerell, A.D., Brooks, C.L.: CHARMM fluctuating charge force field for proteins. Protein/solvent properties from molecular dynamics simulations using a non-additive electrostatic model. J. Computational Chemistry 25(12), 1504–1514 (2004)
27. Ponder, J.W.: TINKER - Software tools for molecular design version (2004), http://chem.skku.ac.kr/~wkpark/tutor/chem/summary.pdf (accessed May 15, 2009)
28. Popenda, M., Bielecki, Ł., Adamiak, R.W.: High-throughput method for the prediction of low-resolution, three-dimensional RNA structures. Nucleic Acids Symposium Series 50(1), 67–68 (2006)
29. Sato, M., et al.: A haptic virtual environment for molecular chemistry education. In: Pan, Z., Cheok, D.A.D., Müller, W., El Rhalibi, A. (eds.) Transactions on Edutainment I. LNCS, vol. 5080, pp. 28–39. Springer, Heidelberg (2008)
30. Stone, J.E., Gullingsrud, J., Schulten, K.: A system for interactive molecular dynamics simulation. In: Proc. ACM Symposium on Interactive 3D Graphics, New York, pp. 191–194 (2001)
31. Su, S., et al.: Distributed collaborative virtual environment: Pauling world. In: Proc. 10th International Conference on Artificial Reality and Telexistence, pp. 112–117 (2000)

32. Van der Spoel, D., et al.: GROMACS: fast, flexible and free. J. Computational Chemistry 26, 1701–1718 (2005)
33. Watson, J.D., Crick, F.H.C.: Molecular structure of nucleic acids. Nature 171, 737–738 (1953)
34. Wollacott, A.M., Merz Jr., K.M.: Haptic applications for molecular structure manipulation. J. Mol. Graphics and Modelling 25, 801–805 (2007)
35. Yuan, J.F., et al.: 3D reconstruction of the Mu transposase and the type 1 transpososome: a structural framework for Mu DNA transposition. Genes & Development 19, 840–852 (2005)
36. Zou, H., et al.: Three-dimensional structure and dynamics of a de novo designed, amphiphilic, metalo-porphyrin-binding protein maquette at soft interfaces by molecular dynamics simulations. J. Phy. Chemistry B 111(7), 1823–1833 (2007)

Shape Recognition of Film Sequence with Application of Sobel Filter and Backpropagation Neural Network

A. Głowacz and W. Głowacz

Faculty of Electrical Engineering, Automatics, Computer Science and Electronics,
AGH University of Science and Technology, Krakow, Poland
{adglow,wglowacz}@agh.edu.pl

Abstract. A new approach to shape recognition is presented. This approach is based on Sobel filter and backpropagation neural network. Investigations of the shape recognition were carried out for film sequences. The aim of this paper is analysis of a system which enables shape recognition.

1 Introduction

At present there are many methods of shape recognition [1, 5-8]. Most of them are based on data processing [9-15]. The aim of this paper is analysis of a system which enables shape recognition. Shape recognition is difficult problem [15-20]. Automatic application of shape recognition contains Sobel filter, thinning algorithm and backpropagation neural network. The system is based on modified "Leaves Recognition v1.0" application. Modifications include mechanism of automatic working and image processing. It could identify shapes. One of the tasks was to apply backpropagation neural network. Investigations were carried out for 96 frames (four seconds) of film sequence. System has five different categories which include specific shapes. Each category contains thirty pictures for the training process. There is the possibility of use of application as a system of automatic shape recognition.

2 Description of the System

System makes possible to realize following functions: video recording, splitting of film sequence on frames, definition of working area, recognition of category depending on configuration set.

After execution of shape recognition the results are presented on computer screen or written to the file.

3 Shape Recognition Process

Shape recognition process contains training process and identification process. At the beginning of the training process image edge detection algorithms (Sobel filter

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 505–516.
springrlink.com

and thinning algorithm) are used. Afterwards tokens of shape image are created. These tokens are used to train backpropagation neural network. At the end of the training process backpropagation neural network is trained. The training process contains following steps: filtration, thinning, creation of shape token, neural network training. Training process is shown in Fig. 1.

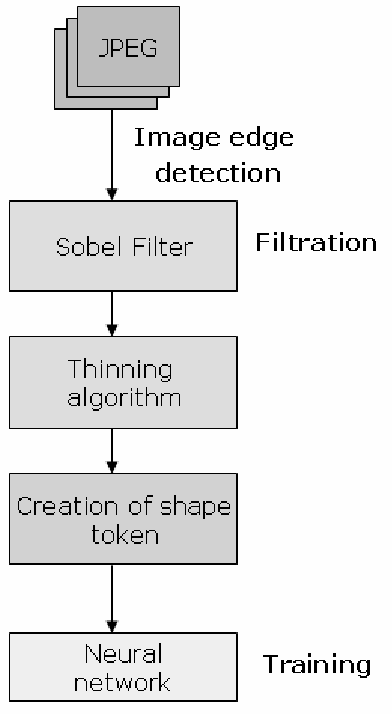


Fig. 1. Training process

New film sequence is used in the identification process. Next it is converted into digital video format (AVI, MPEG). Afterwards it converts into digital picture format (JPEG). Next image edge detection algorithms are used (Sobel filter and thinning algorithm). Afterwards tokens of shape are created. Backpropagation neural network is used as a classifier. The identification process contains following steps: video recording, conversion video recording into digital video format, conversion digital video format into frames, filtration, thinning, creation of shape token, neural network classification. The identification process is shown in Fig. 2.

3.1 Video Recording

Digital video camera records excellent quality video. It records video as a digital stream. After that data are written to computer disk for next calculations.

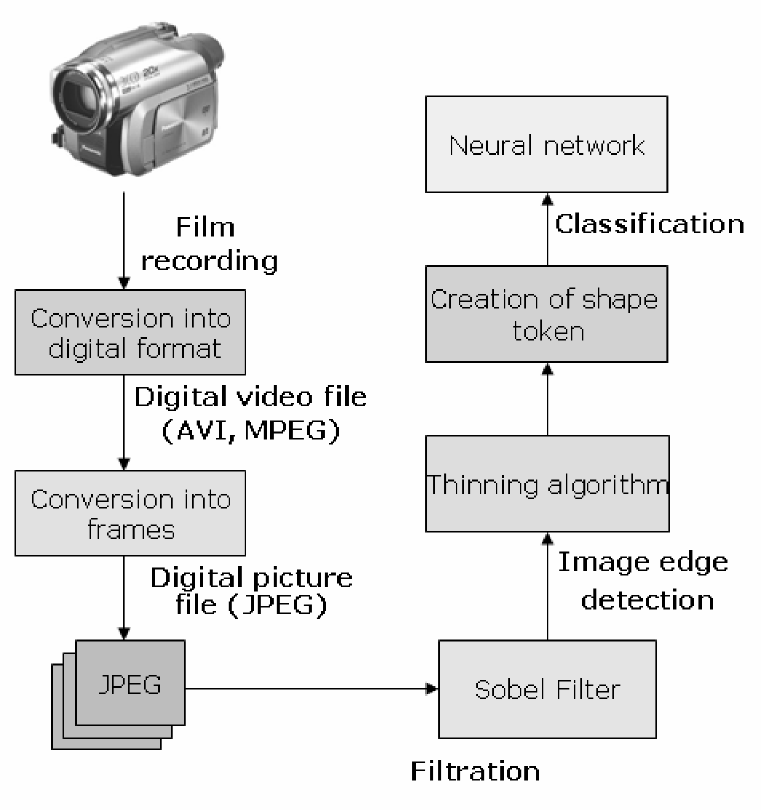


Fig. 2. Identification process

3.2 Digital Video Signal Conversion

Digital video signal is converted into frames (JPEG) by using a compression algorithm. In this aim perl scripts and mplayer library are used. The images have the resolution of 640 x 480 pixels (Fig. 3, 4).

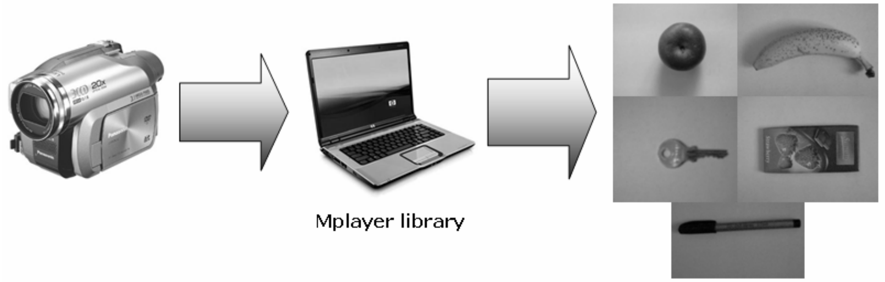


Fig. 3. Conversion digital video signal into frames

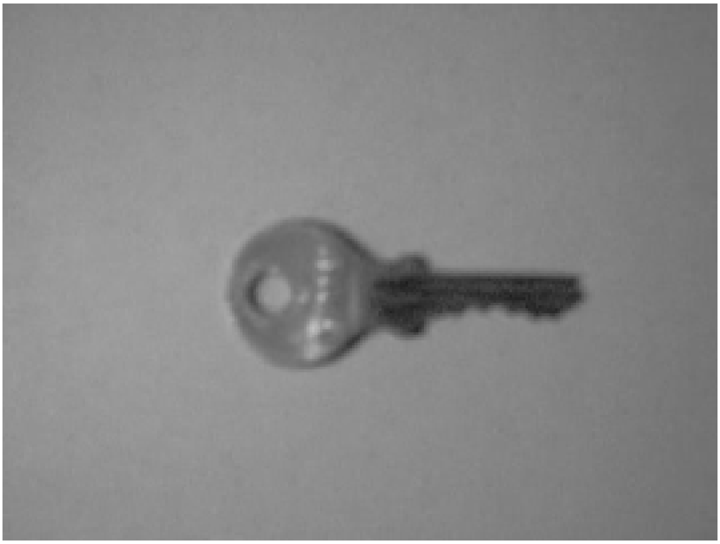


Fig. 4. Picture of the key before filtration

3.3 Filtration

Many methods are used in image processing. In this application Sobel edge detection filter was applied. Sobel edge detection filter creates an image where higher grey-level values indicate the presence of edge between two objects. The Sobel edge filter is used to detect edges (Fig. 5). Horizontal and vertical filters are used in sequence. Both filters are applied to the image and summed to form the final result. The Sobel edge detection filter computes the root mean square of two 3x3 templates. The two filters are basic convolution filters. Horizontal filter is defined as:

$$\mathrm{Y}=\begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \tag{1}$$

Vertical filter is defined as:

$$\mathrm{X}=\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \tag{2}$$

3x3 window is used as:

$$\begin{bmatrix} w_1 & w_2 & w_3 \\ w_4 & w_5 & w_6 \\ w_7 & w_8 & w_9 \end{bmatrix} \tag{3}$$

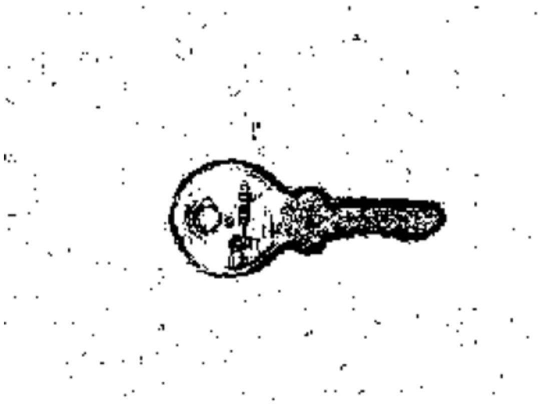


Fig. 5. Picture of the key after filtration

The horizontal and vertical components are used into the final form:

$$Z = \sqrt{(X \cdot X) + (Y \cdot Y)} \tag{4}$$

Here,
$X = (w_3+2w_6+w_9-w_1-2w_4-w_7)$
$Y = (-w_1-2w_2-w_3+w_7+2w_8+w_9)$
Z – modulus of Sobel gradient
$w_1, w_2, w_3, w_4, w_5, w_6, w_7, w_8, w_9$ - higher grey-level values which are clamped to the 0-255 range.

3.4 Thinning Algorithm

It is necessary to identify this outer frame exactly. The previously applied Sobel edge detection identify the edges with a specific threshold.

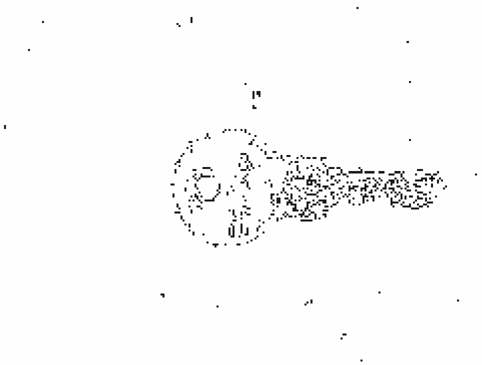


Fig. 6. Picture of the key after thinning algorithm

After that a thinning algorithm is performed to minimize this threshold-based edge to a one-line frame (Fig. 6). Thinning algorithm processed the image recursively and minimizes lines found to a one-pixel wide one, by comparing the actual pixel situation with specific patterns and then minimizes it.

3.5 Creation of Shape Token

The tokens of each shape image that are found after the image preprocessing. These tokens are important in shape recognition process. The tokens are the right-angled triangles (Fig. 7). These triangles represent tokens of shape image.

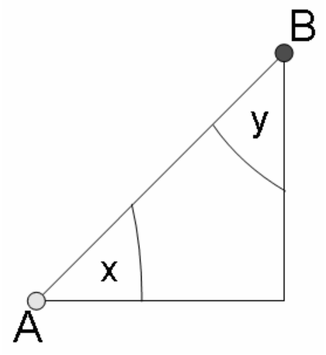


Fig. 7. Right-angled triangle represents a token of a shape image

The angles x and y are the two important parts which will be fit into the neural network input layer. These two angles represent the direction of the hypotenuse from point A to B. It is very important for the representation of a shape image (Fig. 8). Number of tokens depends on image and configuration set. After calculations each token is the input for backpropagation neural network.

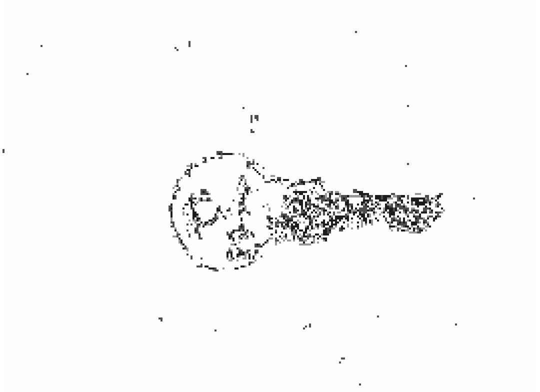


Fig. 8. Picture of the key after creation of shape tokens

3.6 Backpropagation Neural Network

The application of automatic shape recognition can classify tokens. Tokens are created as a result of preliminary data processing. Neural network usable form is when the cosine and sine angles of the shape represent inputs for backpropagation neural network (Fig. 9). The neural network consists of many neurons connected by synapses. The learning process in backpropagation neural network takes place in two phases. In the forward phase, the output of each neuron in each layer and the errors between the actual outputs from the output layer and the target outputs are computed, whereas in the backward phase weights are modified by the back-propagated errors that occurred in each layer of the network [2-4].

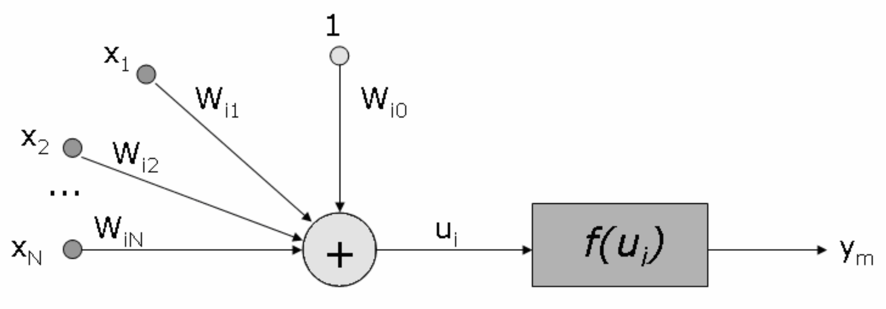


Fig. 9. Model of neuron

Formula to calculate a product of inputs and weights is defined as:

$$u_i = (\sum_{j=1}^{N} W_{ji} x_j + W_{i0}) \tag{5}$$

here,
x = $[x_1, x_2,\dots, x_N]^T$, – is an input vector,
$\mathrm{W_i} = [W_{i1}, W_{i2},\dots, W_{iN}]^T$ – is vector of weights of neuron with *i* index,
W_{i0} – threshold, and
$y_m=f(u_i)$ – a value which is obtained after usage of activation function.

Afterwards, it calculates the value which is on the output of next neuron. To realize this aim, it is necessary to apply the following equation:

$$y_m^{(j)} = f(\sum_{k \in M_i} w_k^{(m)(j)} y_k^{(j)}) \tag{6}$$

here,
M_i – is a set of neuron indexes which provide input signals to defined neuron,
$w_k^{(m)(j)}$ – the value of weights coefficient of synapse linking input of neuron with *m* index and output of neuron with *k* index during *j* learning step,
$y_m^{(j)}$ – the value on the output of neuron with *m* index in *j* step of learning,
$y_k^{(j)}$ – the value on the output of neuron with *k* index in *j* step of learning,

$f(x) = \dfrac{1}{(1+e^{-cx})}$ non-linear activation function (Fig. 10), c – constant.

Formula for error calculation is defined as:

$$\delta_m^{(j)} = \sum_{k\in M} w_k^{(m)(j)} \delta_k^{(j)} \tag{7}$$

Here,
M – is a set of neurons which obtain output signal from defined neuron of hidden layer,
$w_k^{(m)(j)}$ – the value of weights coefficient of synapse linking input of neuron with k index and output of neuron with m index during j learning step,
$\delta_m^{(j)}$ – is the error of neuron with m index in hidden layer, and
$\delta_k^{(j)}$ – is the error of neuron in output layer which obtains signal from neuron with m index.

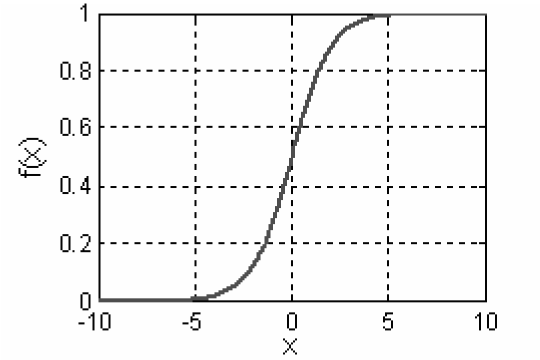


Fig. 10. Activation function, non-linear

The structure of backpropagation neural network is created (Fig. 11)

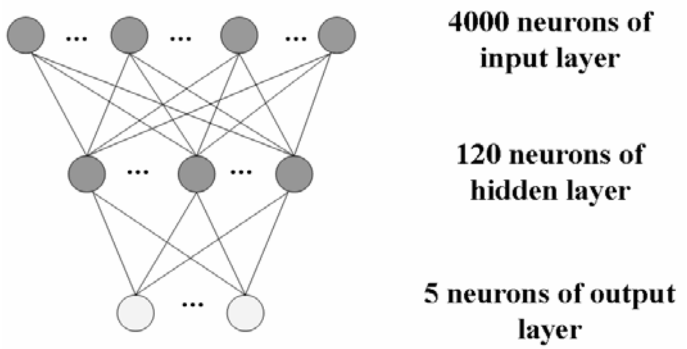

...
4000 neurons of input layer
120 neurons of hidden layer
5 neurons of output layer

Fig. 11. Neural network implemented in automatic application of shape recognition

After training of neural network it is necessary to perform the identification process. Image recognition efficiency is defined as:

$$I = 100 - (\frac{100T}{N})\,[\%] \tag{8}$$

here,
I – is efficiency of image recognition,
N – the number of output neurons, and
T– the output layer error.

In a case when efficiency of image recognition is more than 90%, shape is recognized properly (assumption). Shape recognition efficiency is defined as:

$$R = \frac{100\,P}{A}\,[\%] \tag{9}$$

here,
R – is efficiency of shape recognition,
P – the number of proper recognized images, and
A – the number of all images.

4 Shape Recognition Results

Shape recognition depends on many parameters. It is important to recognize shape correctly. The most impact on shape recognition is determined by following

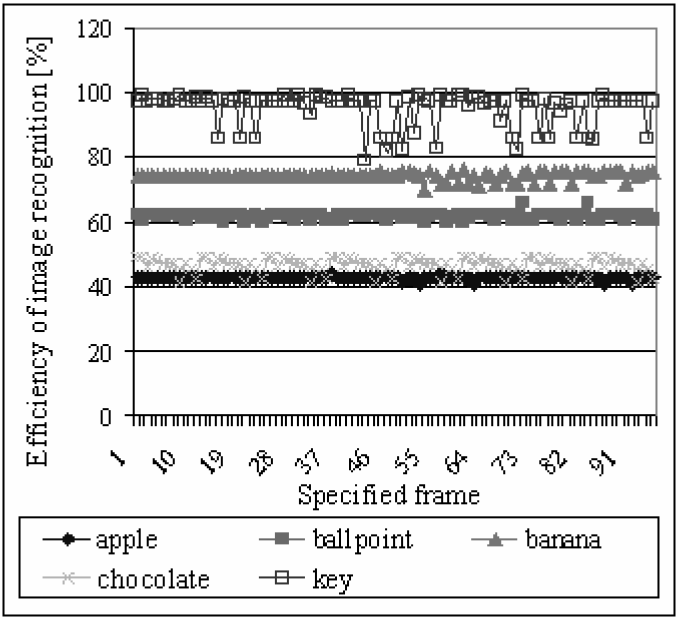


Fig. 12. Efficiency of image recognition depending on specified frame for film sequence with shape of key

parameters: quality of image, parameters of edge detection algorithm, neural network parameters. Investigations were carried out for five categories: shape of ballpoint, shape of chocolate, shape of banana, shape of apple, shape of key. For each category thirty images were used in the training process (30 frames). The identification process was carried out for 96 frames of specified shape (4 seconds). New unknown images were used in the identification process. The system specifies shape of image for each frame. Seventy eight of ninety six (81.25%) frames were recognized properly for film sequence with shape of key. The efficiency of image recognition for shape of key depending on specified frame is shown in Fig. 12. Eighty three of ninety six frames (86.45%) were recognized properly for film sequence with shape of apple. Eighty seven of ninety six frames (90.62%) were recognized properly for film sequence with shape of banana. Seventy two of ninety six frames (75%) were recognized properly for film sequence with shape of chocolate. Eighty of ninety six frames (83.33%) were recognized properly for film sequence with shape of ballpoint.

Efficiency of shape recognition for film sequence with shape of ballpoint is 83.33%. Efficiency of shape recognition for film sequence with shape of chocolate is 75%. Efficiency of shape recognition for film sequence with shape of apple is 86.45%. Efficiency of shape recognition for film sequence with shape of key is 81.25%.The best efficiency of shape recognition is 90.62% for film sequence with shape of banana (Fig. 13).

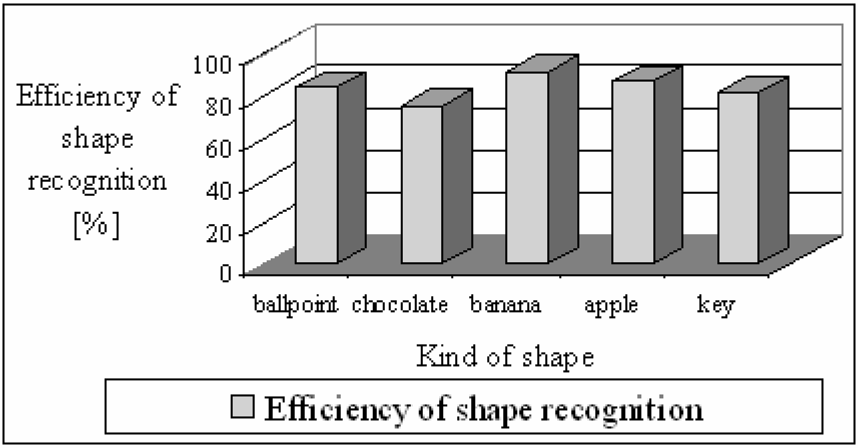


Fig. 13. Efficiency of shape recognition depending on kind of shape

5 Conclusions

We fund that artificial neural network can be very useful for shape recognition of various objects. However, it is important to train neural network properly. Investigations performer show that backpropagation neural network works correctly for different input data. Efficiency of shape recognition for film sequence with shape of ballpoint is 83.33%. Efficiency of shape recognition for film sequence with shape of chocolate is 75%. Efficiency of shape recognition for film sequence with shape of apple is 86.45%. Efficiency of shape recognition for film sequence with shape of key is 81.25%. The best efficiency of shape recognition is 90.62%, for film sequence with shape of banana. Time of the identification process of one sample is 0.97378 [s] for Intel Pentium M 730 processor.

The copyright for the program "Leaves Recognition v1.0" is held by Jens Langner.

References

1. Greene, E.: Simultaneity in the millisecond range as a requirement for effective shape recognition. Behav. Brain Funct. (2006), doi:10.1186/1744-9081-2-38
2. Golden, R.M.: Mathematical methods for neural network analysis and design. MIT Press, Cambridge (1996)
3. Anderson, J.A.: An introduction to neural networks. MIT Press, Cambridge (1995)
4. Fausett, L.V.: Fundamentals of neural networks: architectures, algorithms, and application. Prentice-Hall, Englewood Cliffs (1994)
5. Huang, K., Aviyente, S.: Wavelet feature selection for image classification. IEEE Transactions on Image Processing 17(9), 1709–1720 (2008)
6. Papari, G., Petkov, N.: Adaptive pseudo dilation for gestalt edge grouping and contour detection. IEEE Transactions on Image Processing 17(10), 1950–1962 (2008)
7. Burger, T., Urankar, A., Aran, O., Akarun, L., Caplier, A.: Cued speech hand shape recognition (2007), `http://www.lis.inpg.fr/pages_perso/burger/Publications/burger_aran_visapp.pdf` (accessed April 25, 2009)
8. Cao, F., Delon, J., Desolneux, A., Musé, P., Sur, F.: A unified framework for detecting groups and application to shape recognition. J. Math. Imaging and Vision 27(2), 91–119 (2007)
9. Kotsia, I., Pitas, I.: Facial expression recognition in image sequences using geometric deformation features and support vector machines. IEEE Transactions on Image Processing 16(1), 172–187 (2007)
10. Tsai, L.W., Hsieh, J.W., Fan, K.C.: Vehicle detection using normalized color and edge map. IEEE Transactions on Image Processing 16(3), 850–864 (2007)
11. Grzymala-Busse, J.W., Hippe, Z.S., Roj, E., Skowroński, B.: Applying expert systems to hop extraction monitoring and prediction. Polish Journal of Chemical Technology 8(4), 1–3 (2006)
12. Przytulska, M., Kulikowski, J.L., Wierzbicka, D.: Analysis of irregular shape's time-variations by serial contours enhancement. In: VIIP 2002 Conference on Visualization, Imaging, and Image Processing, Marbella, Spain (2002), `http://www.actapress.com/Content_of_Proceeding.aspx?proceedingID=367#` (accessed April 24, 2009)
13. Malik, J., Belongie, S., Leung, T.K., Shi, J.: Contour and texture analysis for image segmentation. Intern. J. Computer Vision 43(1), 7–27 (2001)
14. Belongie, S., Malik, J., Puzicha, J.: Shape context: a new descriptor for shape matching and object recognition. Adv. Neural Proc. Systems 13, 831–837 (2000)

15. Johnson, A., Bron, L., Brussee, P., Goede, I., Hoogland, S.: Learning from mistakes. In: Proc. Conf. on Human System Interaction, Maastricht, Netherlands, pp. 235–239 (2000)
16. Skaf, A., David, B., Descotes-Genon, B., Binder, Z.: General approach to man-machine system design: ergonomic and technical specification of actions. In: Proc. Conference on Human System Interaction, Maastricht, Netherlands, pp. 355–367 (2000)
17. Gavrila, D., Philomin, V.: Real-time object detection for smart vehicles. In: Intern. Conf. on Computer Vision, Corfu, Greece (1999), http://www.gavrila.net/Publications/./iccv99.pdf (accessed April 25, 2009)
18. Johnson, A.H., Hebert, M.: Recognizing objects by matching oriented points. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, San Juan, Puerto Rico, pp. 684–689 (1997)
19. He, Y., Kundu, A.: 2D shape classification using hidden Markov model. IEEE Transactions on Pattern Analysis and Machine Intelligence 13(11), 1172–1184 (1991)
20. Liu, H.C., Srinath, M.D.: Partial shape classification using contour matching in distance transformation. IEEE Transactions on Pattern Analysis and Machine Intelligence 12(11), 1072–1078 (1990)

Dynamic Changes of Population Size in Training of Artificial Neural Networks

A. Słowik and M. Białko

Department of Electronics and Computer Science,
Koszalin University of Technology, Koszalin, Poland
{aslowik,bialkomi}@ie.tu.koszalin.pl

Abstract. In this paper an adaptive differential evolution algorithm with dynamic changes of population size is presented. In proposed algorithm an adaptive selection of control parameters of the algorithm are introduced. Due to these parameters selection, the algorithm gives better results than differential evolution algorithm without this modification. Also, in presented algorithm dynamic changes of population size are introduced. This modification try to overcome limitations connected with premature convergence of the algorithm. Due to dynamic changes of population size, the algorithm can easier get out from local minimum. The proposed algorithm is used to train artificial neural networks. Results obtained are compared with those obtained using: adaptive differential evolution algorithm without dynamic changes of population size, method based on evolutionary algorithm, error back-propagation algorithm, and Levenberg-Marquardt algorithm.

1 Introduction

The differential evolution algorithm is a newly elaborated technique used in global optimization [1]. This algorithm has been developed to solve problems with continuous domain, but lately some its modifications have been elaborated to discrete problems, as for example in [2], where discrete differential evolution algorithm is used for the permutation flow-shop scheduling problem. In standard version of the differential evolution algorithm proposed by Storn and Price [1], it is necessary to set up at start three parameters such as: crossover rate *CR*, differential mutation factor *F*, and factor *NP* which represents the number of individuals in population. The optimal selection of the values of these parameters is a very complicated task. The values of parameters which are appropriate in one optimization problem are often ineffective in other situations, and can cause premature convergence of algorithm to local extremes. In some papers, the modifications of differential evolution algorithm are presented. Some of them concern adaptive selection of control parameters in differential evolution algorithm, as is mentioned in following papers [3-5]. In all these papers only values of *CR* and *F* parameter are adaptively selected; the values of parameter *NP* are constant during whole algorithm operation time. Therefore, in this article, the differential evolution algorithm with adaptive selection of *CR* and *F* parameters (similarly as in [5]), and with dynamic changes of population size is proposed. Due to introduction of dynamic changes of the number of individuals in the population, proposed algorithm can easier escape from local extremes.

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 517–527.
springrlink.com

Proposed in this paper adaptive differential evolution algorithm is used to train artificial neural networks. Artificial neural networks have many practical applications as is described in papers [6, 7]. The objective function describing the artificial neural network training problem is a multi-modal function, therefore the algorithms based on gradient methods can easily stuck in local extremes. In order to avoid this problem it is possible to use the technique of a global optimization, like for example the differential evolution algorithm. Also, the problem of training of large artificial neural networks is a very time consuming problem (especially for slow convergence methods as for example error back-propagation method), therefore it is necessary to search effective training methods, which can correctly train artificial neural networks in shorter time than standard training methods.

The adaptive differential evolution algorithm with dynamic changes of population size is named DPS-DE-ANNT (Dynamic Population Size – Differential Evolution for Artificial Neural Network Training), and is used to train artificial neural networks to classification of parity-p problem. Result obtained using the proposed method has been compared with results obtained using the error back-propagation algorithm [8, 9, 13-15], evolutionary EA-NNT method [10], Levenberg-Marquardt algorithm [11], and DE-ANNT method [12]. The paper arrangement is a follows: in section 2 the characteristics of standard differential evolution algorithm is presented, in section 3 the proposed algorithm is shown, in section 4 the structure of assumed neural networks and neuron model is described, in section 5 description of taken experiments and obtained results are presented, and in section 6 some conclusions are presented.

2 Differential Evolution Algorithm

The differential evolution algorithm has been proposed by Price and Storn [1]. Its pseudo-code form is as follows:

```
Create an initial population consisting of PopSize individuals
While (termination criterion is not satisfied)
Do Begin
    For each i-th individual in the population
    Begin
        Randomly generate three integer numbers:
        $r1; r2; r3 \in [1; PopSize]$; where $r1 \neq r2 \neq r3 \neq i$
        For each j-th gene in i-th individual ($j \in [1; n]$)
            Begin
                $v_{i,j} = x_{r1,j} + F \cdot (x_{r2,j} - x_{r3,j})$
                Randomly generate one real number $rand_j \in [0; 1)$
                If $rand_j < CR$ then $u_{i,j} := v_{i,j}$
                Else $u_{i,j} := x_{i,j}$
            End;
        If individual $u_i$ is better than individual $x_i$ then
        Replace individual $x_i$ by child individual $u_i$
        End;
    End;
```

The individual x_i is better than individual u_i when the solution represented by it has lower value of objective function (regarding to minimization tasks), or higher - (regarding to maximization tasks) than the solution stored in individual u_i. The algorithm shown in the pseudo-code optimizes the problem having n decision variables. The F parameter is scaling the values added to the particular decision variables, and the CR parameter represents the crossover rate. The parameters $F \in [0; 2)$, and $CR \in [0; 1)$ are determined by the user, and $x_{i,j}$ is the value of j - th decision variable stored in i-th individual in the population. This algorithm is a heuristic one for global optimization, and is operating with decision variables in real number form. The individuals occurring in this algorithm are represented by real number strings. Its searching space must be continuous [17, 18]. The differential evolution algorithm, by computation of difference between two randomly chosen individuals from the population, determines a function gradient in a given area (not in a single point), and therefore prevents sticking the solution in a local extreme of optimized function [18]. The other important property of this algorithm is a local limitation of selection operator only to the two individuals: parent (x_i) and child (u_i), and due to this property the selection operator is more effective and faster [18]. Also, to accelerate convergence of the algorithm, it is assumed that the index *r1* (occurring in the algorithm pseudo-code) points to the best individual in the population. In this paper such version of differential evolution algorithm has been used in experiments.

3 Proposed Method DPS-DE-ANNT

Proposed DPS-DE method is based on previously elaborated method [12], and operates in six following steps:

First Step
A population of individuals is randomly created. The number of individuals in the population is stored in parameter $PopSize \in [Pop_{min}, Pop_{max}]$. In this paper we assume experimentally, that Pop_{min}=20, and Pop_{max}=100, and at the start of the algorithm $PopSize$ is equal to Pop_{min}. Each individual x_i consists of k genes where k represents number of weights in trained artificial neural network). In Figure 1a a part of an artificial neural network with neurons from n to m is shown. Additionally, in Figure 1b the coding scheme for weights of the individual x_i connected to neurons from Figure 1a, is shown.

Second Step
The mutated individual v_i (vector) is created for each individual x_i in the population according to the formula:

$$v_i = x_{r1} + F \cdot \left(x_{r2} - x_{r3}\right) \tag{1}$$

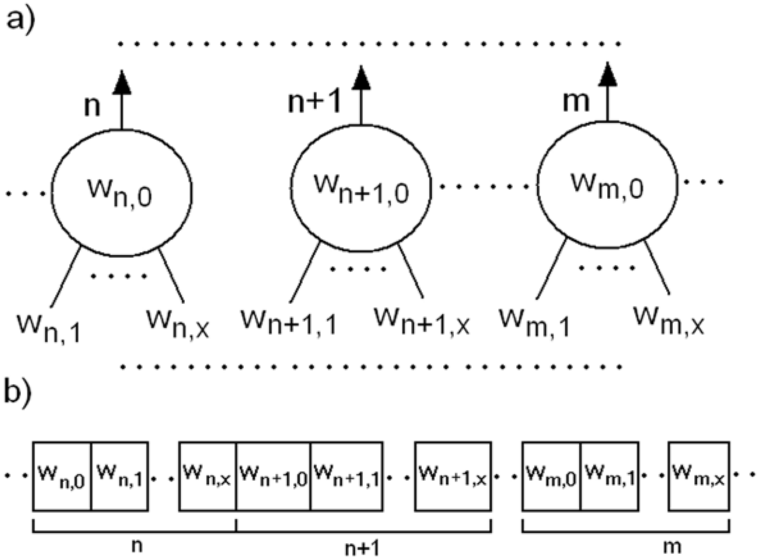


Fig. 1. Part of artificial neural network (a), corresponding to it chromosome containing the weight values (b); weights $w_{i,0}$ represent bias weights [12]

here,
$F \in [0, 2)$, and
r1, *r2*, *r3*, $i \in [1, PopSize]$ fulfill the constraint:

$$r1 \neq r2 \neq r3 \neq i \tag{2}$$

The indices *r2* and *r3* point at individuals randomly chosen from the population. Index *r1* points at the best individual in the population, which has the lowest value of the training error function *ERR*(.). This function is described as follows:

$$ERR() = \frac{1}{2} \cdot \sum_{i=1}^{T} \left(Correct_i - Answer_i\right)^2 \tag{3}$$

here,
I – is the actual number of training vector,
T – the number of all training vectors,
$Correct_i$ – required correct answer for *i*-th training vector,
$Answer_i$ – answer generated by the neural network for *i*-th training vector applied to its input. The DPS-DE-ANNT method is minimizing the value of the objective function *ERR*(.).

Third Step
In the third step, all individuals x_i are crossed-over with mutated individuals v_i. As a result of this crossover operation an individual u_i is created. The crossover operates as follows: for chosen individual $x_i = (x_{i,1}, x_{i,2}, x_{i,3}, ..., x_{i,j})$, and individual

$v_i = (v_{i,1}, v_{i,2}, v_{i,3}, ..., v_{i,j})$; for each gene $j \in [1; k]$ of individual x_i, randomly generate a number $rand_j$ from the range [0; 1), and use the following rule:

$$\text{if } rand_j < CR \text{ then } u_{i,j} = v_{i,j} \text{ else } u_{i,j} = x_{i,j} \tag{4}$$

here,
$CR \in [0; 1)$.

In this paper the adaptive selection of control parameter values F and CR are introduced (similarly as in paper [5]) according to the formulas:

$$A = \left| \frac{TheBest_i}{TheBest_{i-1}} \right| \tag{5}$$

$$F = 2 \cdot A \cdot random \tag{6}$$

$$CR = A \cdot random \tag{7}$$

here,
random – is a random number with uniform distribution in the range [0; 1),
$TheBest_i$ – the value of objective function for the best solution in *i*-th generation,
$TheBest_{i-1}$ – the value of objective function for the best solution in *i*-1-th generation.

Detailed discussion about properties of adaptive selection of control parameter values is described in paper [5].

Fourth Step
In the fourth step, a selection of individuals to the new population is performed according to following rule:

if $ERR(u_i) < ERR(x_i)$ then Replace x_i by u_i in the new population
else Leave x_i in the new population (8)

Fifth Step
The actual population size *PopSize* is modified using following code:

```
if $(TheBest_i/TheBest_{i-1}) = 1$ then
  begin
  $PopSize := PopSize+1$;
  if $(PopSize > Pop_{max})$ then $PopSize := Pop_{max}$;
  Randomly create one new individual and insert it at index $PopSize$ in population
  Evaluate the quality of newly created individual using objective function $ERR(.)$
  end
else
  begin
  $PopSize := PopSize-1$;
  if $(PopSize < Pop_{min})$ then $PopSize := Pop_{min}$;
  end;
```

Using above code, when the value of PopSize is decreasing, it can happen that the best individual will be lost. Therefore, in proposed algorithm the best individual is always written down in the index number one in the population to avoid this situation. But when after evaluation of individuals in population better solution than solution written down in index one will be found, then this particular solution is exchanged with the best solution in population. Due to dynamic changes of population size, new randomly generated solutions are added to population when the value of the best solution is not changed in successive generations. Therefore, the algorithm has higher chances to escape from local minimum, but obviously the computation time is also increasing. In other case, when the value of the best solution is changed in successive generations, then the last individual (solution having index number *PopSize*) from population is removed. Due to removal of the last solution form population, the algorithm computational time is decreasing, and therefore the algorithm can realize more iterations (generations) at the same time.

Sixth Step

In the sixth step, it is checked whether the value of $ERR(x_{r1}) < \varepsilon$, or the algorithm has reached the prescribed number of generations (index r1 points the best individual with the lowest value of objective function $ERR(.)$ in the population). If yes, then the algorithm is stopped, and result stored in individual x_{r1} is returned. Otherwise, the algorithm jumps to the second step.

4 Structure of Assumed Neural Networks and Neuron Model

Similarly as in paper [12], the proposed method has been tested using the same structure of artificial neural networks, which are shown in Figure 2. The input U0 is removed from each neuron for figure simplify.

The typical model of neuron including the adder of input signal values multiplied by corresponding values of weights - i.e. weighted sum, has been taken as a model of artificial neuron (as in paper [12]). The weighted sum WS_j of j-th neuron is defined as follows:

$$WS_j = \sum_{i=0}^{p} w_{j,i} \cdot U_j \tag{9}$$

here,

p – is the number of inputs in j-th neuron,
$w_{j,i}$ – value of weight in the connection between j-th neuron and its i-th input,
U_i – the value of signal occurring on i-th neuron input (U_0=1).

A bipolar sigmoid activation function has been assumed in the form:

$$U_j = f\left(WS_j\right) = \frac{1-\exp\left(-\lambda \cdot WS_j\right)}{1+\exp\left(-\lambda \cdot WS_j\right)} \tag{10}$$

here,
U_j – is the value of j-th neuron output,
λ – non-linearity coefficient of activation function; (assumed λ=1).

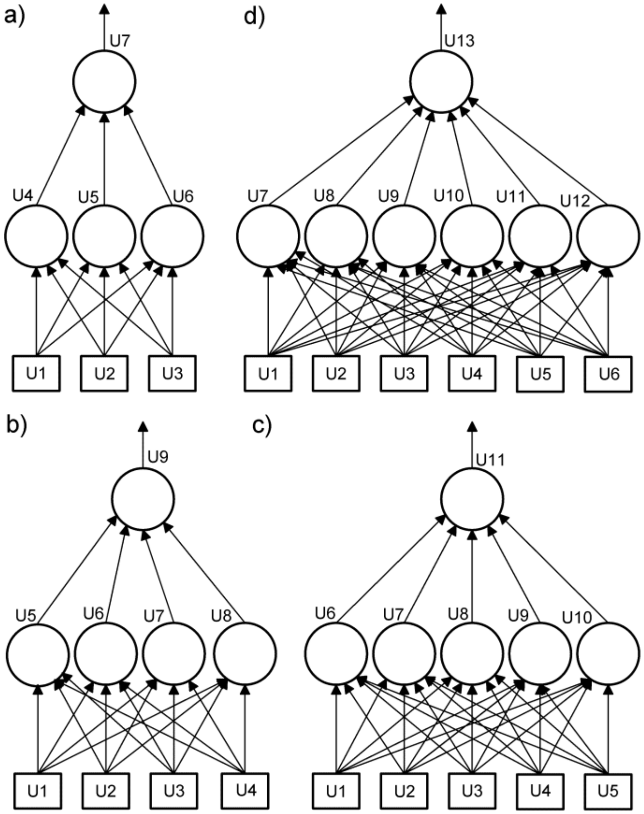


Fig. 2. Structures of artificial neural networks training for classification of problem: parity-3 (a), parity-4 (b), parity-5 (c), and parity-6 (d)

5 Experiments

Artificial neural networks having structures shown in Figure 2, to classification of parity-p problem ($p\in$ [3; 6]) have been trained using proposed and other methods. An example training set was equal to the testing set, and contained 2p vectors. Following values of parameters have been assumed: $PopSize=Pop_{min}$=20, ε=0.0001. Each of algorithms: DPS-DE-ANNT (DPS), DE-ANNT (DE) [12], EA-NNT (EA) [10], error back-propagation algorithm (EBP) [14, 15], and Levenbeg-Marquardt algorithm (LM) [11] have been executed 10-fold, and average values of results are presented in Tables 1-3. Identical termination criterion such as in DPS-DE-ANNT method described in the section 4 of this paper, have been assumed for all other algorithms. The learning coefficient ρ=0.2 has been assumed in the EBP algorithm. Also, the same maximal computation time has been assumed as in paper [12]: for parity-3 problem this time was equal to 1 [s], for parity-4 problem 3 [s], for parity-5 problem 9 [s], and for parity-6 problem 60 [s]. Each algorithm have been stopped when the value of training error of artificial neural network had lower value than ε (ε=0.0001) or when operation time exceeded the maximal computation time for each parity problem.

The results obtained using proposed method are presented in Tables 1-3 (the results for other methods are taken from [12]), in which the symbols used are as follows: *ME* – chosen training method, *NI* – number of iterations, *CC* – correct classification [%]. The values representing the correct *CC* were computed as follows:

$$CC = \frac{\sum_{i=1}^{M} C_i}{2^p} \cdot 100\% \tag{11}$$

here,
CC – means the correct classification [%],
M – is the number of testing vectors ($M \in [1, 2^p]$),
p – the number of inputs in artificial neural network,
C_i – a coefficient representing correctness of classification of *i*-th training vector, which is determined as follows:

$$C_i = \begin{cases} 1, & when\ U_{out} > \varphi\ for\ B_i = 1 \\ 1, & when\ U_{out} < -\varphi\ for\ B_i = -1 \\ 0, & otherwise \end{cases} \tag{12}$$

again here,
U_{out}=f(S_{out}) – is a value of the output signal of artificial neural network after application of *i*-th testing vector to its input,
φ – a threshold of training correctness,
B_i – a value expected at the output of the artificial neural network.

Table 1. Average values of results obtained after 10-fold repetition of each algorithm (φ=0)

Parity-3 Problem			Parity-4 Problem		
ME	NI	CC [%]	ME	NI	CC [%]
DSP	61.6	**100**	DSP	893.3	**97.50**
DE	23.5	**100**	DE	399.2	96.875
EA	57.7	95	EA	147.1	92.50
EBP	300	78.75	EBP	800	91.25
LM	20	**100**	LM	48.1	**97.50**
Parity-5 Problem			Parity-6 Problem		
ME	NI	CC [%]	ME	NI	CC [%]
DSP	1549.3	**97.1875**	DSP	5484.3	96.09375
DE	657.5	96.875	DE	1988	95.78125
EA	348.9	92.1875	EA	1486.6	94.53125
EBP	2250	94.6875	EBP	10800	96.09375
LM	64.1	**97.1875**	LM	391.2	**97.34375**

Table 2. Average values of results obtained after 10-fold repetition of each algorithm (φ=0.90)

Parity-3 Problem			Parity-4 Problem		
ME	NI	CC [%]	ME	NI	CC [%]
DSP	99.5	**100**	DSP	909.8	88.75
DE	24.9	**100**	DE	442.4	87.5
EA	56.4	81.25	EA	154.5	72.5
EBP	300	36.25	EBP	800	69.375
LM	19	**100**	LM	49.9	**97.5**
Parity-5 Problem			Parity-6 Problem		
ME	NI	CC [%]	ME	NI	CC [%]
DSP	1361.7	96.25	DSP	6136.53	87.84375
DE	420.3	96.5625	DE	2058.4	89.84375
EA	351	67.8125	EA	1505.4	78.75
EBP	2250	69.6875	EBP	10800	85.78125
LM	161.7	**96.5625**	LM	1044.2	**97.8125**

Table 3. Average values of results obtained after 10-fold repetition of each algorithm (φ=0.99)

Parity-3 Problem			Parity-4 Problem		
ME	NI	CC [%]	ME	NI	CC [%]
DSP	48.9	**97.5**	DSP	1191.4	**83.125**
DE	25.3	96.25	DE	485.4	81.25
EA	55.2	65	EA	153.3	60.625
EBP	300	6.25	EBP	800	2.5
LM	21.7	71.25	LM	33.9	81.875
Parity-5 Problem			Parity-6 Problem		
ME	NI	CC [%]	ME	NI	CC [%]
DSP	1145	**88.75**	DSP	5749.3	**90.9375**
DE	640.4	86.875	DE	2025.3	83.59375
EA	350.4	60.3125	EA	1498.9	65
EBP	2250	17.1875	EBP	10800	41.09375
LM	57.7	84.375	LM	154.6	85.9375

In Figure 3, graphical representations of φ parameter for its successive values: $\varphi = 0$ (Figure 3a), $\varphi = 0.9$ (Figure 3b), φ=0.99 (Figure 3c) are presented. In all Figures, dashed lines represents assumed sigmoidal activation functions.

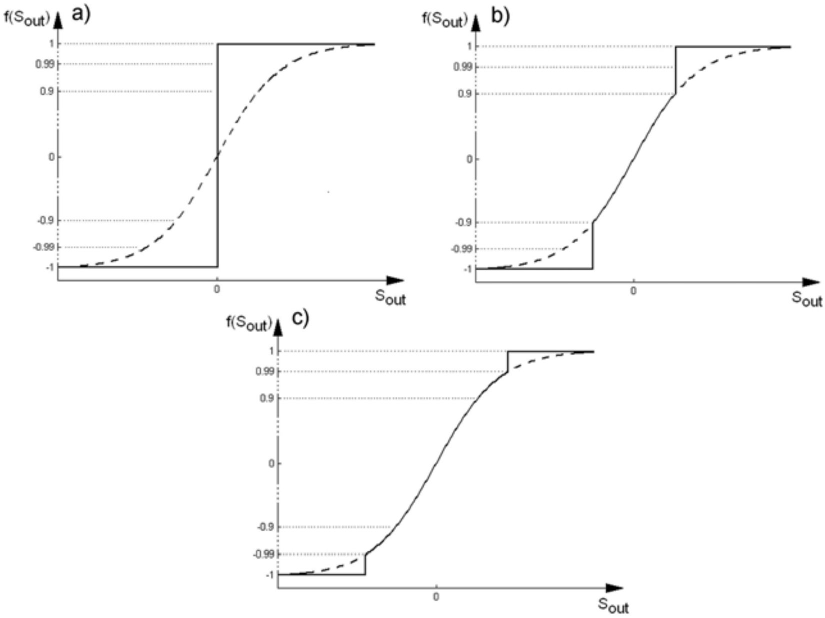


Fig. 3. Threshold of training correctness for: φ=0 (a), φ=0.9 (b), φ=0.99 (c) [12]

6 Conclusions

From Tables 1-3 we can see that application of adaptive differential evolution algorithm with dynamic changes of population size leads in many cases to better results, than results obtained using other methods. The results obtained using proposed DSP method are in 10 cases comparable or better (on 12 possible) than results obtained using DE method [12]. In comparison to EA, EBP, and LM method, results obtained using proposed method are in 32 cases better or comparable on 36 possible. It is necessary to point out that presented algorithm can also be easily used to train multi-output artificial neural networks, artificial neural networks with nonstandard architectures (for example the tower architecture [16]), and networks with non-differentiable neuron activation functions for which the application of gradient training methods as for example: EBP or LM algorithms is not possible. Also, proposed adaptive differential evolution algorithm with dynamic changes of population size can be easily used in any optimization task with continuous domain.

References

1. Storn, R., Price, K.: Differential evolution: A simple and efficient heuristic for global optimization over continuous spaces. J. Global Optimization 11, 341–359 (1997)
2. Pan, Q.Q., Tasgetiren, M.F., Liang, Y.C.: A discrete differential evolution algorithm for the permutation flow-shop scheduling problem. In: Proc. Genetic and Evolutionary Computation Conference, pp. 126–133 (2007)

3. Liu, J., Lampinen, J.: A fuzzy adaptive differential evolution algorithm: soft computing - a fusion of foundations. Methodologies and Applications 9(6), 448–462 (2005)
4. Ali, M.M., Torn, A.: Population set-based global optimization algorithms: some modifications and numerical studies. Computers and Operations Research 31(10), 1703–1725 (2004)
5. Słowik, A., Białko, M.: Adaptive selection of control parameters in differential evolution algorithms: computational intelligence: methods and applications. In: Zadeh, L., Rutkowski, L., Tadeusiewicz, R., Zurada, J. (eds.), pp. 244–253. Academic Publishing House Exit, Warsaw (2008)
6. Hudson, C.A., Lobo, N.S., Krishnan, R.: Sensor-less control of single switch-based switch- reluctance motor drive using neural network. IEEE Transactions on Industrial Electronics 55(1), 321–329 (2008)
7. Bose, B.K.: Neural network applications in power electronics and motor drives: an introduction and perspective. IEEE Transactions on Industrial Electronics 54(1), 14–33 (2007)
8. Fu, L.: Neural networks in computer intelligence. McGraw-Hill, New York (1994)
9. Masters, T.: Practical neural network recipes in C++. Academic Press, New York (1993)
10. Słowik, A., Białko, M.: Application of evolutionary algorithm to training of feed-forward flat artificial neural networks. In: Proc. 14th National Conference on Computer Application in Scientifics Research, Wrocław, pp. 35–40 (2007) (in Polish)
11. Osowski, A.: Neural networks in algorithmic use. WNT, Warsaw (1996) (in Polish)
12. Słowik, A., Białko, M.: Training of artificial neural networks using differential evolution algorithm. In: Proc. IEEE Conference on Human System Interaction, Cracow, Poland, pp. 60–65 (2008)
13. Hecht-Nielsen, R.: Neurocomputing. Addison-Wesley Publishing Company, Reading (1991)
14. Zurada, J.: Introduction to artificial neural systems. West Publishing Company, St. Paul (1992)
15. Białko, M.: Artificial intelligence and elements of hybrid expert systems. Publishing House of Koszalin University of Technology, Koszalin (2005) (in Polish)
16. Gallant, S.: Neural network learning and expert systems. MIT Press, Cambridge (1993)
17. Engelbrecht, A.P.: Computational intelligence – An introduction. Wiley & Sons Inc., New York (2007)
18. Becerra, R.L., Coello, C.A.: Cultured differential evolution for constrained optimization. Computer Methods in Applied Mechanics and Engineering 195(33-36), 4303–4322 (2006)

New Approach to Diagnostics of DC Machines by Sound Recognition Using Linear Predictive Coding

A. Głowacz and W. Głowacz

Faculty of Electrical Engineering, Automatics, Computer Science and Electronics,
AGH University of Science and Technology, Krakow, Poland
{adglow,wglowacz}@agh.edu.pl

Abstract. A new approach to determination of similarity of dc machine sounds is presented. This approach is based on Linear Predictive Coding (LPC) algorithm and some metrics of distance in multidimensional solution space. The aim of this paper is analysis of a system which enables sound recognition. Developed sound recognition system consists of preliminary data processing, feature extraction and classification algorithms. There were applied eight various distance metrics. Investigations were carried out for direct current machine because it produces characteristic sounds. It can be noticed that sound of faultless dc machine is different from sound of faulty dc machine, what can be used to determine the state of dc machine performance. Investigations of the sound recognition were carried out for faultless machine and machine with shorted rotor coils. The results of sound recognition are discussed in this paper.

1 Introduction

At present there are many methods of sound recognition [1, 4-7]. Most of them are based on data processing [8-12]. The aim of this paper is analysis of a system which enables sound recognition. Sound recognition system contains preliminary data processing, feature extraction and classification algorithms; together they make possible to identify sounds. Sound recognition application can classify feature vectors. Feature vectors are created as a result of preliminary data processing and linear predictive coding. One of the tasks was to use some metrics of distance in multidimensional solution space. There were applied eight various metrics of distance: Manhattan, Euclidean, Minkowski (with m=3, 4, 5, 6), then also cosine distance, and Jacquard distance. Investigations were carried out for direct current machine because it produces characteristic sounds. It can be emphasized that sound of faultless dc machine is different from sound of faulty dc machine. Difference of sounds depends on differences in ordered sequence [2, 3]. System has two different categories which include faultless dc machine sound and sound of dc machine with shorted rotor coils. It can determine the state of dc machine work. For recognition aim the mechanism of early detection of damages in dc machine was created.

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 529–540.
springrlink.com

2 Sound Recognition Process

Sound recognition process contains feature vectors creation process (Fig. 1) and identification process (Fig. 2). At the beginning of feature vectors creation process signals are sampled and normalized. Afterwards data are converted through the Hamming window. Next data are converted through the Linear Predictive Coding (LPC) algorithm. The LPC algorithm creates feature vectors. Feature vectors creation process and identification process are based on the same signal processing algorithms. The difference between them is a sequence of execution. In feature vectors creation process all feature vectors are averaged. Two averaged feature vectors are created. Feature vectors creation process contains following steps: sampling, quantization, normalization, filtration, windowing, feature extraction, averaging of feature vectors.

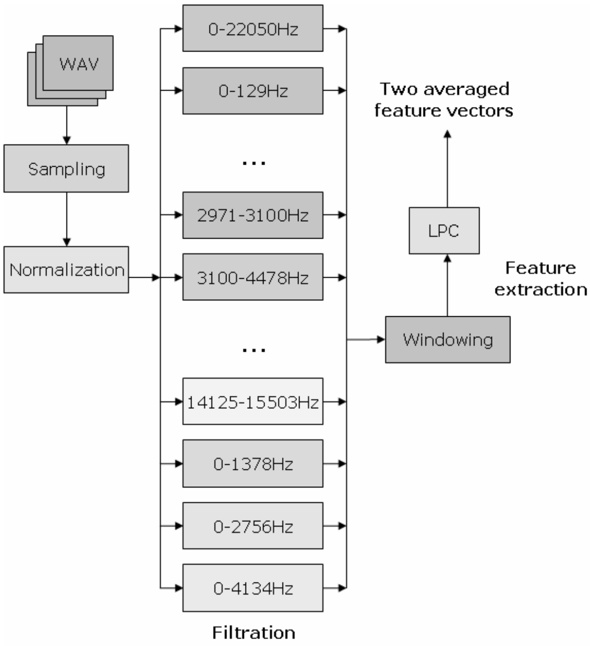


Fig. 1. Feature vectors creation process

In the identification process new acoustic signal is recorded. Afterwards it divides wave file. After that signals are sampled and normalized. Next data are converted through the Hamming window, and later converted through the LPC algorithm, in order to create so called feature vectors. These vectors are then applied in the identification in the identification process. To obtain results of recognition, the feature vector of a new sample is compared and averaged with feature vector of specific category. Identification process consists of the following steps: recording of acoustic signal, sound track dividing, sampling, quantization, normalization, filtration, windowing, feature extraction, and classification.

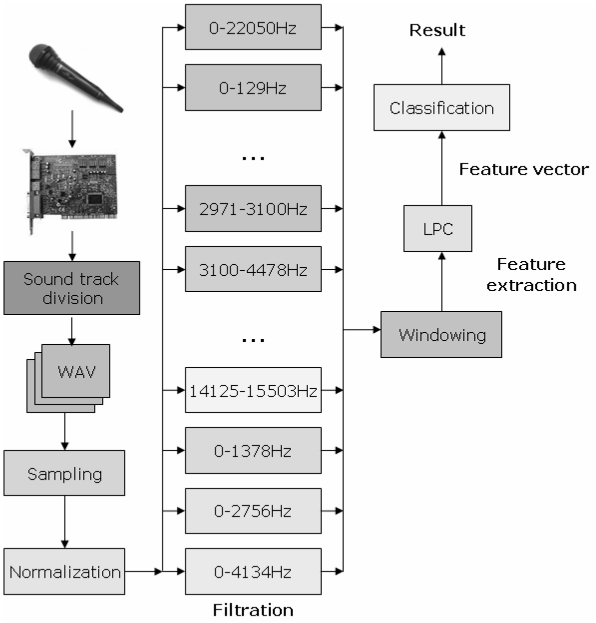


Fig. 2. Identification process

2.1 Acoustic Signal Recording

Sound card with analogue-digital converter is able to record, process and replay sound. Recording of the acoustic signal is the first part of the identification process. Acoustic signal is converted into digital data (wave format) by the microphone (OLYMPUS TP-7) and the sound card. This wave file contains following parameters: sampling frequency is 44100 Hz, number of bits is 16, and number of channels is 1 (mono). Investigated dc machine is shown in Fig. 3.

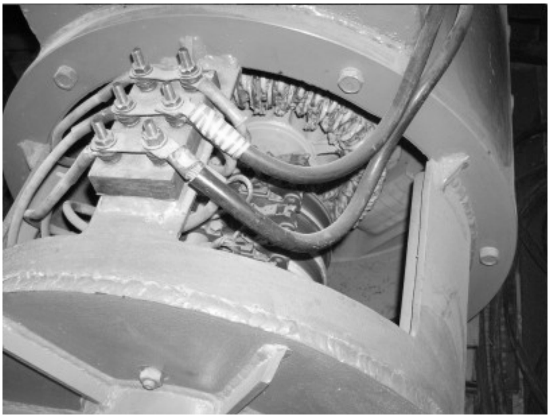


Fig. 3. Investigated dc machine

2.2 Sound Track Dividing

Application divides sound track into sound fragments. It divides data. Next it creates new wave header. Afterwards new wave header is copied. Then new wave header is added to each chunk of data. New wave files are obtained. These files (samples) are used in the identification process. There are following advantages of such solution: precise determination of sound appearing, precise sound identification, and additionally the application does not have to allocate so much memory in identification process.

2.3 Sampling

Sampling frequency is basic parameter. Sampling frequency is 44100 Hz in sound recognition application (Fig. 4 and 5).

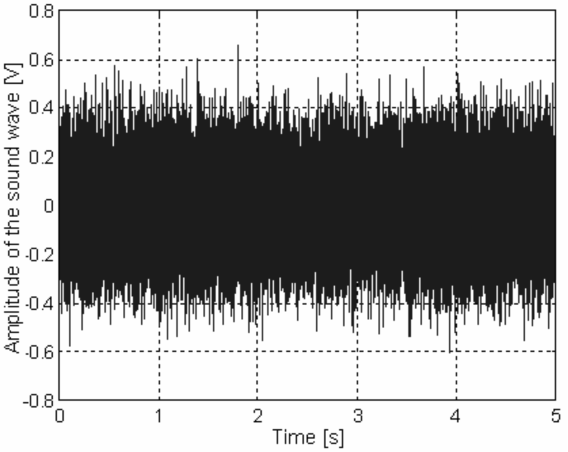


Fig. 4. Sound of faultless dc machine for five seconds before normalization

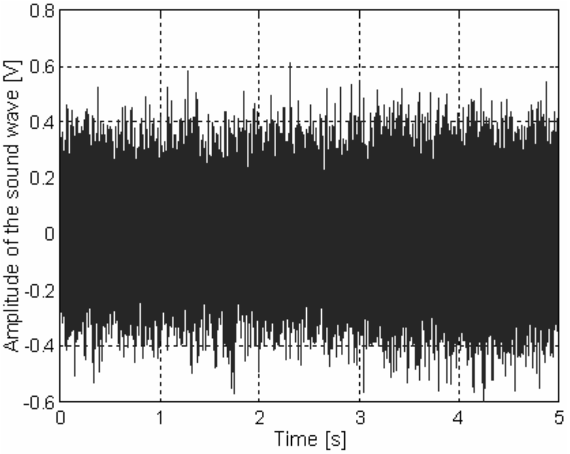


Fig. 5. Sound of dc machine with shorted rotor coils for five seconds before normalization

2.4 Quantization

Quantization is a technique to round intensity values to a quantum so that they can be represented by a finite precision. Precision of sample depends on number of bits. Common applied number of bits is 8 or 16. Sound recognition application uses 16 bits because it gives better precision. There is a choice of number of bits depending on quantity of input data and calculations speed in sound recognition process. The compromise is important to obtain good results in short time.

2.5 Amplitude Normalization

For sound recognition application, normalization is the process of changing of the amplitude of an audio signal. There is a possibility that some sounds aren't recorded at the same level. It is essential to normalize the amplitude of each sample in order to ensure, that feature vectors will be comparable. All samples are normalized in the range [−1.0, 1.0]. The method finds the maximum amplitude in the sample, and then scales down the amplitude of the sample by dividing each point by this maximum [12].

2.6 Filtration

Filtration is a very efficient way of removing the unwanted noise from the spectrum. Filtration is used to modify the frequency domain of the input sample. Filtration is not necessary to sound recognition. However use of it can improve the efficiency of the sound recognition. For investigations 36 filters were used: 0-129 Hz, 129-258 Hz, 258-387 Hz, 387-516 Hz, 516-645 Hz, 645-775 Hz, 775-904 Hz, 904-1033 Hz, 1033-1162 Hz, 1162-1291 Hz, 1291-1421 Hz, 1421-1550 Hz, 1550-1679 Hz, 1679-1808 Hz, 1808-1937 Hz, 1937-2067 Hz, 2067-2196 Hz, 2196-2325 Hz, 2325-2454 Hz, 2454-2583 Hz, 2583-2713 Hz, 2713-2842 Hz, 2842-2971 Hz, 2971-3100 Hz, 3100-4478 Hz, 4478-5857 Hz, 5857-7235 Hz, 7235-8613 Hz, 8613-9991 Hz, 9991-11369 Hz, 11369-12747 Hz, 12747-14125 Hz, 14125-15503 Hz, 0-1378 Hz, 0-2756 Hz, 0-4134 Hz.

2.7 Windowing

There are different types of window functions available, each with their own advantage. The Hamming window is used to avoid distortion of the overlapped window functions. It is defined as:

$$w(n) = 0.53836 - 0.46164 \cdot \cos(\frac{2\pi n}{m-1}) \tag{1}$$

where: w(n) – is new sample amplitude, n – index into the window, m – total length of the window.

2.8 Linear Predictive Coding

LPC analyzes the sound signal by estimating the formants, removing their effects from the sound signal, and estimating the intensity and frequency of the remaining buzz [13, 14]. It determines a set of coefficients approximating the amplitude versus frequency function. These coefficients create feature vectors which are used in calculations. The model of shaping filter is defined as:

$$H(z) = \frac{1}{1 - \sum_{k=1}^{p} a_k z^{-k}} \tag{2}$$

here, p is the order of the filter, a_k is prediction coefficient.

Prediction a sound sample is based on a sum of weighted past samples:

$$s'(n) = \sum_{k=1}^{p} a_k \cdot s(n-k) \tag{3}$$

here, s'(n) is the predicted value based on the previous values of the sound signal s(n).

LP analysis requires estimating the LP parameters for a segment of sound. Formula (3) provides the closest approximation to the sound samples. This means that s′(n) is closest to s(n) for all values of n in the segment. The spectral shape of s(n) is assumed to be stationary across the frame, or a short segment of sound. The error between the actual sample and the predicted one can be expressed as:

$$e(n) = s(n) - s'(n) \tag{4}$$

The summed squared error E over a finite window of length N is defined as:

$$E = \sum_{n} e^2(n) \tag{5}$$

where: 0≤n≤N+p-1

The minimum value of E occurs when the derivative is zero with respect to each of the parameters ak. By setting the partial derivatives of E, a set of p equations are obtained. The matrix form of these equations is:

$$\begin{bmatrix} r(0) & r(1) & \cdots & r(p-1) \\ r(1) & r(0) & \cdots & r(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ r(p-1) & r(p-2) & \cdots & r(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} r(1) \\ r(2) \\ \vdots \\ r(p) \end{bmatrix} \tag{6}$$

where r(i) is the autocorrelation of lag i computed as:

$$r(i) = \sum_{m=0}^{N-1-i} s(m) \cdot s(m+i) \tag{7}$$

where, N is the length of the sound segment s(n).

The Levinson-Durbin algorithm solves the n-th order system of linear equations.

$$R \cdot a = r \tag{8}$$

For the particular case where R is a Hermitian, positive definite, toeplitz matrix and r is identical to the first column of R shifted by one element.

The autocorrelation coefficients r(k) are used to compute the LP filter coefficients a_i, i=1,…p and k=1,…p, by solving the set of equations:

$$\sum_{i=1}^{p} a_i \cdot r(|i-k|) = r(k) \tag{9}$$

These coefficients are used in calculations (Fig. 6 and 7). The Levinson-Durbin algorithm is used to estimate linear prediction coefficients from a given sound waveform. This method is efficient, as it needs only the order of M2 multiplications to compute the linear prediction coefficients.

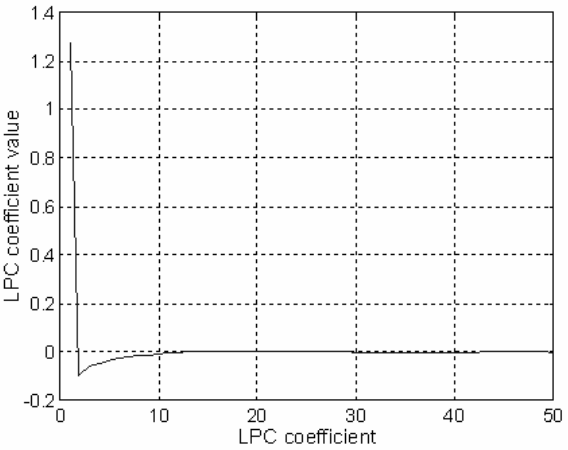


Fig. 6. LPC coefficients values for sound of faultless dc machine after normalization and use of low-pass filter which passes frequencies 0-1378 Hz

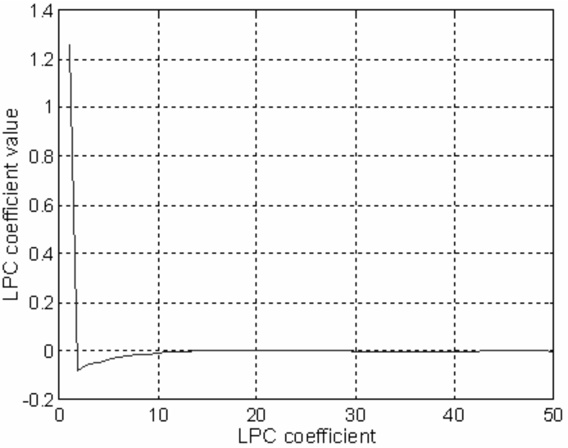


Fig. 7. LPC coefficients values for sound of dc machine with shorted rotor coils after normalization and use of low-pass filter which passes frequencies 0-1378 Hz

2.9 Classification

Difference between sounds depends on differences in ordered sequence. Classification uses feature vectors and distance metrics in the identification process. It compares different values of feature vectors. The least distance between feature vectors (feature vector of investigated sample, averaged feature vector of specific category) is chosen in the identification process. The nearest vector is the result of the identification.

2.9.1 Manhattan Distance

Manhattan distance is the measure of distance between two vectors. For vectors x and y with the same length n it is defined as:

$$d_m(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{n} (| x_i - y_i |) \tag{10}$$

here x and y are feature vectors with the same lengths, x=[x1,x2,…,xn], y=[y1,y2,…,yn].

2.9.2 Euclidean Distance

Euclidean distance is the measure of distance between two vectors. For vectors x and y with the same length n it is defined as:

$$d_e(\mathbf{x},\mathbf{y}) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \tag{11}$$

where: x and y are feature vectors with the same lengths, x=[x1,x2,…,xn], y=[y1,y2,…,yn].

2.9.3 Minkowski Distance

Minkowski distance is the measure of distance between two vectors. For vectors x and y with the same length n it is defined as:

$$d_{\min k}(\mathbf{x},\mathbf{y}) = \left(\sum_{i=1}^{n} (| x_i - y_i |)^m \right)^{\frac{1}{m}} \tag{12}$$

here x and y are feature vectors with the same lengths, x=[x1, x2,…, xn], y=[y1, y2,…, yn].

2.9.4 Cosine Distance

Cosine distance is the measure of distance between two vectors. For vectors x and y with the same length n it is defined as:

$$d_{\cos}(\mathbf{x},\mathbf{y}) = 1 - \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}} \tag{13}$$

here x and y are feature vectors with the same lengths, x=[x1, x2,…, xn], y=[y1, y2,…, yn].

2.9.5 Jacquard Distance

Jacquard distance is the measure of distance between two vectors. For vectors x and y with the same length n it is defined as:

$$d_j(\mathbf{x},\mathbf{y}) = 1 - \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} y_i^2 - \sum_{i=1}^{n} x_i y_i} \tag{14}$$

here x and y are feature vectors with the same lengths, x=[x1, x2,…, xn], y=[y1, y2,…, yn].

3 Sound Recognition Results

Investigations were carried out for sound of faultless dc machine and sound of dc machine with shorted rotor coils. Nine five-second samples were used for feature vectors creation process for each category. New unknown samples were used in the identification process. System should determine the state of dc machine correctly. Identification process was carried out for one-second, two-second, three-second, four-second, and five-second periods. Sound recognition efficiency depending on length of sample is presented in Fig. 8-11.

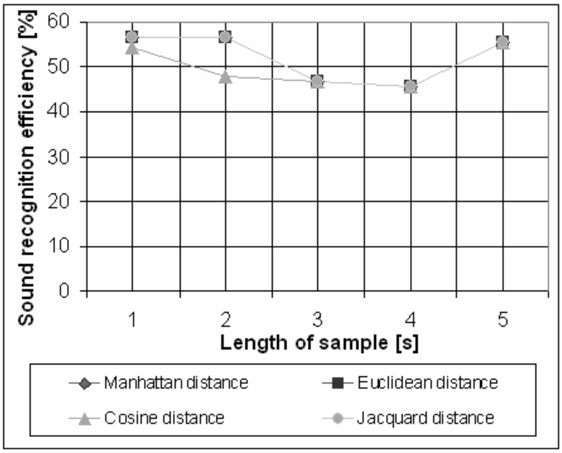


Fig. 8. Sound recognition efficiency of faultless dc machine depending on length of sample and distance metrics (normalization, 0-1378 Hz, LPC)

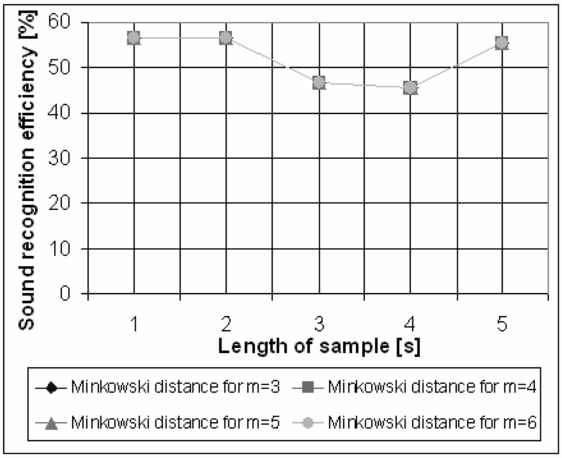


Fig. 9. Sound recognition efficiency of faultless dc machine depending on length of sample and Minkowski metric (normalization, 0-1378 Hz, LPC)

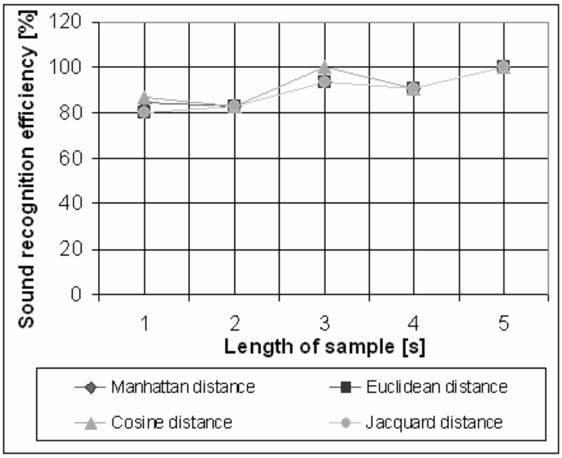


Fig. 10. Sound recognition efficiency of dc machine with shorted rotor coils depending on length of sample and distance metrics (normalization, 0-1378 Hz, LPC)

Sound recognition efficiency is defined as:

$$E = \frac{N_1}{N} \tag{15}$$

here, E – is sound recognition efficiency, N1 – number of correctly identified samples, N – number of all samples.

All investigated metrics gave very good results for five-second samples. Sound recognition efficiency was 55.55% for faultless dc machine. Sound recognition efficiency was 100% for dc machine with shorted rotor coils.

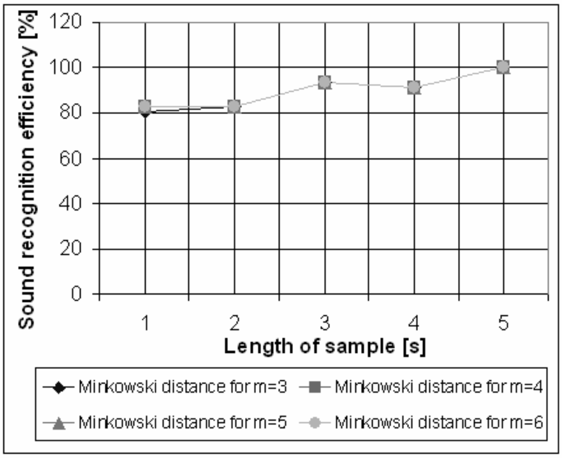


Fig. 11. Sound recognition efficiency of dc machine with shorted rotor coils depending on length of sample and Minkowski metric (normalization, 0-1378 Hz, LPC)

4 Conclusions

Sound recognition system was created. It identifies category which has the least distance between feature vectors. Algorithms of signal processing and distance metrics were used in the identification process. Investigations were carried out for different input data. Analysis shows the sensitivity of methods which are based on distance metrics depending on input data. The best results were obtained for five-second samples. It used distance metrics with the low-pass filter which passes frequencies 0-1378 Hz. Sound recognition efficiency was 55.55% for faultless dc machine. Sound recognition efficiency was 100% for dc machine with shorted rotor coils. Time of the identification process of one five-second sample was 1.672 s for Intel Pentium M 730 processor.

References

1. Lee, K.: Effective approaches to extract features and classify echoes in long ultrasound signals from metal shafts, PhD dissertation, Brisbane, Australia (2006)
2. Głowacz, A., Głowacz, W.: Sound recognition of dc machine with application of FFT and backpropagation neural network. Przegląd Elektrotechniczny 84(9), 159–162 (2008) (in Polish)
3. Głowacz, Z., Głowacz, W.: Mathematical model of DC motor for analysis of commutation processes. In: Proc. 6th IEEE Intern. Symposium on Diagnostics for Electric Machines, Power Electronics and Drives, Krakow, Poland, pp. 138–141 (2007)
4. Dubois, D., Guastavino, C.: Cognitive evaluation of sound quality: bridging the gap between acoustic measurements and meanings. In: Proc. 19th Intern. Congress on Acoustics, Madrid, Spain (2007)

5. Milner, B., Shao, X.: Prediction of fundamental frequency and voicing from mel-frequency cepstral coefficients for unconstrained speech reconstruction. IEEE Transactions on Audio, Speech and Language Processing 15(1), 24–33 (2007)
6. New, T.L., Li, H.: Exploring vibrato-motivated acoustic features for singer identification. IEEE Transactions on Audio, Speech and Language Processing 15(2), 519–530 (2007)
7. Kinnunen, T., Karpov, E., Fränti, P.: Real-time speaker identification and verification. IEEE Transactions on Audio, Speech, and Language Processing 14(1), 277–288 (2006)
8. Mitrovic, D., Zeppelzauer, M., Eidenberger, H.: Analysis of the data quality of audio features of environmental sounds. Journal of Universal Knowledge Management 1(1), 4–17 (2006)
9. Umapathy, K., Krishnan, S., Rao, R.K.: Audio signal feature extraction and classification using local discriminant bases. IEEE Transactions on Audio, Speech and Language Processing 15(4), 1236–1246 (2007)
10. Yoshii, K., Goto, M., Okuno, H.G.: Drum sound recognition for polyphonic audio signals by adaptation and matching of spectrogram templates with harmonic structure suppression. IEEE Transactions on Audio, Speech, and Language Processing 15(1), 333–345 (2007)
11. Alexandre, E., Cuadra, L., Álvarez, L., Rosa-Zurera, M., López-Ferreras, F.: Two-layers automatic sound classification system for conversation enhancement in hearing aids. Integrated Computer-Aided Engineering 15(1), 85–94 (2008)
12. The MARF Development Group. Modular audio recognition framework v.0.3.0-devel-20050606 and its applications. Application note, Montreal, Canada (2005)
13. ITU-T Recommendation G.729 - Coding of speech at 8 kbit/s using conjugate structure algebraic code excited linear prediction (CS-ACELP) (January 2007)
14. ITU-T Recommendation G.723.1 - Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s (May 2006)

Diagnosis Based on Fuzzy IF-THEN Rules and Genetic Algorithms

A. Rotshtein[1] and H. Rakytyanska[2]

[1] Department of Industrial Engineering and Management,
Jerusalem College of Technology, Jerusalem, Israel
rot@jct.ac.il
[2] Department of Soft Ware Design,
Vinnitsa National Technical University, Vinnitsa, Ukraine
h_rakit@ukr.net

Abstract. This paper proposes an approach for inverse problem solving based on the description of the interconnection between unobserved and observed parameters of an object (causes and effects) with the help of fuzzy IF-THEN rules. The essence of the approach proposed consists of formulating and solving the optimization problems, which, on the one hand, find the roots of fuzzy logical equations, corresponding to IF-THEN rules, and on the other hand, tune the fuzzy model on the readily available experimental data. The genetic algorithms are proposed for the optimization problems solving. The efficiency of the method is illustrated by computer experiment, and also by the example of the inverse diagnosis problem, which requires renewal of the causes (inputs) by the observed effects (outputs).

1 Introduction

The wide class of the problems, arising in engineering, medicine, economics and other domains, belongs to the class of the inverse problems [1]. The essence of the inverse problem consists in the following. The dependency ***Y=f(X)*** is known, which connects the vector ***X*** of the unobserved parameters with the vector ***Y*** of the observed parameters. It is necessary to ascertain the unknown values of the vector ***X*** through the known values of the vector ***Y***. The typical representative of the inverse problem is the problem of medical and technical diagnosis, which amounts to the restoration and the identification of the unknown causes of the disease or the failure through the observed effects, i.e. the symptoms or the external signs of the failure. The diagnosis problem, which is based on a cause and effect analysis and abductive reasoning can be formally described by neural networks [2] or Bayesian networks [3].

In the cases, when domain experts are involved in developing cause-effect connections, the dependency between unobserved and observed parameters can be modelled using the means of fuzzy sets theory [4] – [6]: fuzzy relations and fuzzy IF-THEN rules. The analytical [7] – [9] and numerical [10] – [12] methods of solving the inverse problems of diagnosis on the basis of fuzzy relations and

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 541–556.
springerlink.com

Zadeh's compositional rule of inference are the most developed ones. In this paper we propose an approach for solving diagnosis problem based on description of the cause-effect connections with the help of fuzzy IF-THEN rules. These rules enable to consider complex combinations in cause-effect connections simpler and more naturally, which are difficult to model with fuzzy relations. For example, the expert interconnection of the unobserved and the observed parameters (causes and effects) in the fuel pipe diagnosis problem can look as follows:

IF feed pressure is *high* and leakage is *low* and pipe resistance is *low,*
THEN delivery head is *high* and productivity is *high.*

This example has three input (unobserved) parameters and two output (observed) parameters. Each parameter is evaluated by the fuzzy term. The problem consists not only in solving system of fuzzy logical equations, which correspond to IF-THEN rules, but also in selection of such forms of the fuzzy terms membership functions and such weights of the fuzzy IF–THEN rules, which provide maximal proximity between model and real results of diagnosis.

The essence of the proposed approach consists in formulating and solving the optimization problems, which, on the one hand, find the roots of fuzzy logical equations, corresponding to IF-THEN rules, and, on the other hand, tune the fuzzy model on the readily available experimental data. The genetic algorithms are proposed for the formulated optimization problems solving.

2 Fuzzy Model of Diagnosis

Cause-effect interconnection can be represented with use of expert matrix of knowledge (Table 1) [5]. The fuzzy knowledge base below corresponds to this matrix:

$$\begin{gathered}\text{Rule } l : \text{IF } x_1 = a_{1l} \text{ and } x_2 = a_{2l} \text{ ... and } x_n = a_{nl} \\ \text{THEN } y_1 = b_{1l} \text{ and } y_2 = b_{2l} \text{ ... and } y_m = b_{ml} \text{ with weight } w_l,\ l = \overline{1,K},\end{gathered} \tag{1}$$

where: a_{il} is a fuzzy term for variable x_i evaluation in the rule with number l; b_{jl} is a fuzzy term for variable y_j evaluation in the rule with number l; w_l is a rule weight, i.e. a number in the range [0, 1], characterizing the measure of confidence of an expert relative to the statement with number l; K is the number of fuzzy rules.

Table 1. Fuzzy knowledge base

Rule	x_1	x_2	…	x_n	y_1	y_2	…	y_m	Weight
1	a_{11}	a_{21}	…	a_{n1}	b_{11}	b_{21}	…	b_{m1}	w_1
2	a_{12}	a_{22}	…	a_{n2}	b_{12}	b_{22}	…	b_{m2}	w_2
…	…	…	…	…	…	…	…	…	…
K	a_{1K}	a_{2K}	…	a_{nK}	b_{1K}	b_{2K}	…	b_{mK}	w_K

The problem of inverse logical inference is set in the following way: it is necessary to restore and identify the values of the input parameters $(x_1^*, x_2^*,..., x_n^*)$ through the values of the observed output parameters $(y_1^*, y_2^*,..., y_m^*)$.

The restoration of the inputs amounts to the solution of the system of fuzzy logical equations, which is derived from relation (1):

$$\begin{aligned} &\mu^{a_{1l}}(x_1) \wedge \mu^{a_{2l}}(x_2)... \wedge \mu^{a_{nl}}(x_n) = \\ &= w_l \cdot (\mu^{b_{1l}}(y_1) \wedge \mu^{b_{2l}}(y_2)... \wedge \mu^{b_{ml}}(y_m)). \end{aligned} \tag{2}$$

Here $\mu^{a_{il}}(x_i)$ is a membership function of a variable x_i to the fuzzy term (cause) a_{il}; $\mu^{b_{jl}}(y_j)$ is a membership function of a variable y_j to the fuzzy term (effect) b_{jl}.

Taking into account the fact that operation $\wedge$ is replaced by *min* in fuzzy set theory [4], system (2) is rewritten in the form

$$\min_{\mathrm{i=1,n}}\left[\mu^{a_{il}}(x_i)\right] = w_l \cdot \min_{\mathrm{j=1,m}}\left[\mu^{b_{jl}}(y_j)\right]$$

or

$$\min_{\mathrm{i=1,n}}\left[\mu^{a_{il}}(x_i)\right] = \mu^{B_l}(w_l, Y), \quad l = \overline{1, K}, \tag{3}$$

where $\mu^{B_l}(w_l, Y)$ is the measure of the effects combination significance in the rule with number l.

The use of fuzzy logical equations provides for the presence of the fuzzy terms membership functions included in the knowledge base. We use a bell-shaped membership function model in the form [13]:

$$\mu^T(u) = 1/(1 + ((u - \beta)/\sigma)^2), \tag{4}$$

where β is a coordinate of function maximum, $\mu^T(\beta) = 1$; σ is a parameter of concentration-extension.

Correlations (3) and (4) define the generalized fuzzy model of diagnosis as follows:

$$F_Y(\boldsymbol{X}, \boldsymbol{B}_C, \boldsymbol{\Omega}_C) = \mu^B(\boldsymbol{Y}, \boldsymbol{W}, \boldsymbol{B}_E, \boldsymbol{\Omega}_E), \tag{5}$$

where $\boldsymbol{X} = (x_1, x_2,..., x_n)$ is the vector of input variables; $\boldsymbol{Y} = (y_1, y_2,..., y_m)$ is the vector of output variables; $\mu^B = (\mu^{B_1}, \mu^{B_2},..., \mu^{B_K})$ is the vector of effects combinations significances measures in the IF–THEN rules; $\boldsymbol{W} = (w_1, w_2,..., w_K)$ is the vector of rules weights; $\boldsymbol{B}_C = (\beta^{C_1}, \beta^{C_2},..., \beta^{C_N})$ and $\boldsymbol{\Omega}_C = (\sigma^{C_1}, \sigma^{C_2},..., \sigma^{C_N})$

are the vectors of β - and σ -parameters for input variables membership functions to the fuzzy terms $C_1, C_2, \ldots, C_N$; $\boldsymbol{B}_E = (\beta^{E_1}, \beta^{E_2}, \ldots, \beta^{E_M})$ and $\boldsymbol{\Omega}_E = (\sigma^{E_1}, \sigma^{E_2}, \ldots, \sigma^{E_M})$ are the vectors of β- and σ-parameters for output variables membership functions to the fuzzy terms $E_1, E_2, \ldots, E_M$; *N* is the total number of fuzzy terms for input variables; *M* is the total number of fuzzy terms for output variables; F_Y is the operator of inputs–outputs connection, corresponding to formulae (3), (4).

3 Solving Fuzzy Logical Equations

The relational equations approach is the most developed one for solving inverse problem, and in most cases it is the final expression for other descriptions, e.g., when the relation between the input ***X*** and the output ***Y*** is described by fuzzy IF-THEN rules. We propose to use the fuzzy relational calculus theory [9] which provides relational equations resolution when the composition is *max-min* for fuzzy rules based inverse problem solving. In this case the system (3) can be considered as a system of fuzzy relational equations of the following form:

$$\boldsymbol{\mu}^C \circ \overline{\boldsymbol{R}} = \boldsymbol{\mu}^B$$

where: is the operation of *min-max* composition [9]; $\overline{\boldsymbol{R}}$ is the complement of the relational matrix ***R*** with elements $r_{kl} \in \{0,1\}$, $\mathrm{k} = \overline{1,\mathrm{N}}$, $\mathrm{l} = \overline{1,\mathrm{K}}$,

$$r_{kl} = \begin{cases} 1, \text{ if term } C_k \text{ is present in the } lth \text{ rule;} \\ 0, \text{ if term } C_k \text{ is absent in the } lth \text{ rule.} \end{cases}$$

The idea is to use the solvers for *max-min* composition when solving the systems with *min-max* composition, applying duality [9]. In the general case, system with *max-min* composition has a solution set, which is completely characterized by the unique greatest solution and a set of lower solutions [7] – [9]. In the dual case, we solve the inverse problem for *max-min* composition for the complement of $\boldsymbol{\mu}^B$ and ***R***, and then find the complement of its solutions [9].

A cornerstone of the approximate methods consists in the transformation of the fuzzy effects vector in a way leading to the exact solution of the modified equations [6]. In this paper a genetic algorithm transforms the initial fuzzy logical equations into solvable ones. Formation of the solution set for the modified equations is accomplished by exact analytical methods [7] – [9] supported by the free software [9].

3.1 Optimization Problem

Following the approach, proposed in [10]–[12], the problem of solving fuzzy logical equations (3) is formulated as follows. Vector $\boldsymbol{\mu}^C = (\mu^{C_1}, \mu^{C_2}, \ldots, \mu^{C_N})$ of the membership degrees of the inputs to fuzzy terms $C_1, C_2, \ldots, C_N$, should be found which satisfies the constraints $\mu^{C_k} \in [0,1]$, $\mathrm{k} = \overline{1,\mathrm{N}}$, and also provides the least distance between model and observed measures of effects combinations significances, that is between the left and the right parts of each system equation (3)

$$F = \sum_{l=1}^{K} \left[\underline{\min}_{\mathrm{i}=1,\mathrm{n}} \left[\mu^{a_{il}}(x_i) \right] - \mu^{B_l}(Y) \right]^2 = \min_{\mu^C} . \tag{6}$$

In accordance with [7] – [9], in the general case system (3) has a solution set $S(\boldsymbol{\mu}^B)$, which is completely characterized by the unique minimal solution $\underline{\boldsymbol{\mu}}^C$ and the set of maximal solutions $S^*(\boldsymbol{\mu}^B) = \left\{ \overline{\boldsymbol{\mu}}_t^C, \ t = \overline{1,T} \right\}$:

$$S(\boldsymbol{\mu}^B) = \bigcup_{\overline{\boldsymbol{\mu}}_t^C \in S^*} \left[\underline{\boldsymbol{\mu}}^C, \overline{\boldsymbol{\mu}}_t^C \right] . \tag{7}$$

Here $\underline{\boldsymbol{\mu}}^C = (\underline{\mu}^{C_1}, \underline{\mu}^{C_2}, \ldots, \underline{\mu}^{C_N})$, $\overline{\boldsymbol{\mu}}_t^C = (\overline{\mu}_t^{C_1}, \overline{\mu}_t^{C_2}, \ldots, \overline{\mu}_t^{C_N})$ are the vectors of the lower and upper bounds of the membership degrees of the inputs to the terms C_k, where the union is taken over all $\overline{\boldsymbol{\mu}}_t^C \in S^*(\boldsymbol{\mu}^B)$.

Formation of solution set (7) begins with the search for the null solution of optimization problem (6). As the null solution of optimization problem (6) we designate $\boldsymbol{\mu}_0^C = (\mu_0^{C_1}, \mu_0^{C_2}, \ldots, \mu_0^{C_N})$, where $\mu_0^{C_k} \geq \underline{\mu}^{C_k}$, $\mathrm{k} = \overline{1,\mathrm{N}}$. The modified vector of the effects combinations significances measures $\boldsymbol{\mu}_0^B = (\mu_0^{B_1}, \mu_0^{B_2}, \ldots, \mu_0^{B_K})$, which corresponds to the obtained null solution $\boldsymbol{\mu}_0^C$, provides the analytical solvability of the fuzzy logical equations (3). Formation of the solution set $S(\boldsymbol{\mu}_0^B)$ for the modified vector $\boldsymbol{\mu}_0^B$ is accomplished by exact analytical methods [7]–[9] supported by the free software [9].

3.2 Genetic Algorithm

The genetic algorithm is used for the null solution finding. We define the chromosome as the vector-line of binary solution codes μ^{C_k}, $\mathrm{k} = \overline{1,\mathrm{N}}$. The chromosomes of the initial population will be defined by: $\mu^{C_k} = \mathrm{RANDOM}([0,1])$. The crossover operation is carried out by way of exchanging genes inside each variable μ^{C_k}. The mutation operation implies random inversion of some bits. We used the selection procedure giving priority to the best solutions. The greater the fitness function of some chromosome the greater is the probability for the given chromosome to yield offsprings [14]. We choose criterion (6) as the fitness function.

While performing the genetic algorithm the size of the population stays constant. That is why after crossover and mutation operations it is necessary to remove the chromosomes having the worst values of the fitness function from the obtained population.

4 Fuzzy Model Tuning

It is assumed that the training data which is given in the form of L pairs of experimental data is known: $\left\langle \hat{\boldsymbol{X}}_p, \hat{\boldsymbol{Y}}_p \right\rangle$, where $\hat{\boldsymbol{X}}_p = \left(\hat{x}_1^p, \hat{x}_2^p, ..., \hat{x}_n^p\right)$ and $\hat{\boldsymbol{Y}}_p = \left(\hat{y}_1^p, \hat{y}_2^p, ..., \hat{y}_m^p\right)$ are the vectors of the values of the input and output variables in the experiment number p, $p = \overline{1, L}$.

The goal of the tuning stage is to find a fuzzy system that allows for particularly accurate solutions of the inverse problem. Inverse problem solving amounts to the search for the null solution $\boldsymbol{\mu}_0^C$, which directly produces the modified fuzzy effects vector $\boldsymbol{\mu}_0^B = F_Y(\boldsymbol{\mu}_0^C)$. The tuning stage guarantees finding such null solutions $\boldsymbol{\mu}_0^C(\hat{\boldsymbol{X}}_p)$ of the inverse problem, which minimize the criterion (6) for all points of the training data

$$\sum_{\text{p=1}}^{\text{L}} \left[F_Y(\boldsymbol{\mu}_0^{\text{C}}(\hat{\boldsymbol{X}}_{\text{p}})) - \hat{\mu}^B(\hat{\boldsymbol{Y}}_p)\right]^2 = \min.$$

Thus the essence of tuning of the fuzzy model (5) consists of finding such vector of fuzzy rules weights $\boldsymbol{W}$ and such vectors of membership functions parameters $\boldsymbol{B}_C$, $\boldsymbol{\Omega}_C$, $\boldsymbol{B}_E$, $\boldsymbol{\Omega}_E$, which provide the least distance between model and experimental vectors of the effects combinations significances measures:

$$\sum_{\text{p=1}}^{\text{L}} \left[F_Y(\hat{\boldsymbol{X}}_{\text{p}}, \boldsymbol{B}_{\text{C}}, \boldsymbol{\Omega}_{\text{C}}) - \hat{\mu}^B(\hat{\boldsymbol{Y}}_p, \boldsymbol{W}, \boldsymbol{B}_{\text{E}}, \boldsymbol{\Omega}_{\text{E}})\right]^2 = \min \cdot \tag{8}$$

The chromosome needed in the genetic algorithm for solving this optimization problem is defined as the vector-line of binary codes of parameters $\boldsymbol{W}$, $\boldsymbol{B}_C$, $\boldsymbol{\Omega}_C$, $\boldsymbol{B}_E$, $\boldsymbol{\Omega}_E$. Fitness function is built on the basis of criterion (8).

5 Computer Experiment

The aim of the experiment consists of checking the performance of the above proposed models and algorithms with the help of the target “inputs – outputs” model.

Some analytical functions $y_1 = f_1(x_1, x_2)$ and $y_2 = f_2(x_1, x_2)$ were approximated by the combined fuzzy knowledge base, and served simultaneously as training and testing data generator. The input values (x_1, x_2), restored for each output combination (y_1, y_2), were compared with the target level lines. The target model is given by the formulae: $y_1 = ((2z_1 - 0.9)\ (7z_1 - 1)\ (17z_2 - 19)\ (15z_2 - 2))/10$, $y_2 = -y_1 + 3.4$, where $z_1 = ((x_1 - 2.9)^2 + (x_2 - 2.9)^2)/39$, $z_2 = ((x_1 - 3.1)^2 + (x_2 - 3.1)^2)/41$.

The target model is represented in Fig. 1. The fuzzy IF-THEN rules correspond to this model:

Rule 1: IF x_1=*L* and x_2=*L* THEN y_1=*lA* and y_2=*hA*;
Rule 2: IF x_1=*A* and x_2=*L* THEN y_1=*hL* and y_2=*lH*;
Rule 3: IF x_1=*H* and x_2=*L* THEN y_1=*lA* and y_2=*hA*;
Rule 4: IF x_1=*L* and x_2=*A* THEN y_1=*hL* and y_2=*lH*;
Rule 5: IF x_1=*A* and x_2=*A* THEN y_1=*H* and y_2=*L*;
Rule 6: IF x_1=*H* and x_2=*A* THEN y_1=*hL* and y_2=*lH*;
Rule 7: IF x_1=*L* and x_2=*H* THEN y_1=*lA* and y_2=*hA*;
Rule 8: IF x_1=*A* and x_2=*H* THEN y_1=*hL* and y_2=*lH*;
Rule 9: IF x_1=*H* and x_2=*H* THEN y_1=*lA* and y_2=*hA*,

where the total number of the causes and effects consists of: C_1 *Low (L)*, C_2 *Average (A)*, C_3 *High (H)* for x_1, C_4 *(L)*, C_5 *(A)*, C_6 *(H)* for x_2; E_1=*higher than Low* (*hL*), E_2= *lower than Average (lA)*, E_3= *High (H)* for y_1; E_4=*Low (L)*, E_5=*higher than Average (hA)*, E_6=*lower than High (lH)* for y_2.

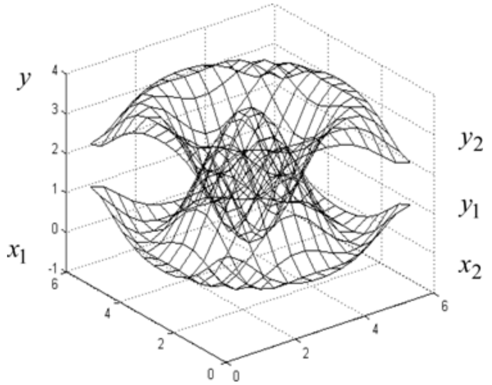


Fig. 1. «Inputs – outputs» model-generator

Fuzzy logical equations take the following form:

$$\begin{aligned}
\mu^{C_1} \wedge \mu^{C_4} &= w_1 \cdot (\mu^{E_2} \wedge \mu^{E_5}) \\
\mu^{C_2} \wedge \mu^{C_4} &= w_2 \cdot (\mu^{E_1} \wedge \mu^{E_6}) \\
\mu^{C_3} \wedge \mu^{C_4} &= w_3 \cdot (\mu^{E_2} \wedge \mu^{E_5}) \\
\mu^{C_1} \wedge \mu^{C_5} &= w_4 \cdot (\mu^{E_1} \wedge \mu^{E_6}) \\
\mu^{C_2} \wedge \mu^{C_5} &= w_5 \cdot (\mu^{E_3} \wedge \mu^{E_4}) \, . \\
\mu^{C_3} \wedge \mu^{C_5} &= w_6 \cdot (\mu^{E_1} \wedge \mu^{E_6}) \\
\mu^{C_1} \wedge \mu^{C_6} &= w_7 \cdot (\mu^{E_2} \wedge \mu^{E_5}) \\
\mu^{C_2} \wedge \mu^{C_6} &= w_8 \cdot (\mu^{E_1} \wedge \mu^{E_6}) \\
\mu^{C_3} \wedge \mu^{C_6} &= w_9 \cdot (\mu^{E_2} \wedge \mu^{E_5})
\end{aligned} \tag{9}$$

The training data $\left\langle \hat{\boldsymbol{X}}_p, \hat{\boldsymbol{Y}}_p \right\rangle$, $p = 1,2,...,1000$, was generated using the target model. The training data was used to evaluate criterion (8) during the evolutionary optimization of the fuzzy model parameters. The model was verified by renewal the testing data. In our experiment the testing data $\left\langle \hat{\boldsymbol{X}}_p, \hat{\boldsymbol{Y}}_p \right\rangle$, $p = 1,2,...,1000$, was generated using the target model. The input values $\boldsymbol{X}(\hat{\boldsymbol{Y}}_p)$, restored for each output $\hat{\boldsymbol{Y}}_p$, were compared with the target values $\hat{\boldsymbol{X}}_p$. We evaluated the quality of the model using the following root mean-squared errors $RMSE_{x_1}$ and $RMSE_{x_2}$:

$$RMSE_{x_i} = \sqrt{\frac{1}{1000} \sum_{p=1}^{1000} \left[x_i(\hat{\boldsymbol{Y}}_p) - \hat{x}_i^p \right]^2} \, .$$

In each run, the value of criterion (8) for the training data and the values of $RMSE_{x_i}$ for the testing data were evaluated. Dependence of the number of generations, necessary to obtain optimal solutions of the optimization problems (8) and (6), on the population size (*V*), cross-over (p_c) and mutation (p_m) ratios was studied in the course of the computer experiment. It was determined that population size of V =10 is sufficient for tuning fuzzy model and solving fuzzy logical equations. To exclude hitting the local minimum the experiment was carried out for large values of p_c and p_m. Under conditions of $p_c = 0.7$ and $p_m = 0.02$ about 15000 generations were required to grow optimal solution of optimization problem (8) and 1500 generations were required to grow optimal solution of optimization problem (6). To cut time losses in unpromising fields studies some parameters of the main genetic operations were experimentally selected. Setting of cross-over ratio at the level of 0.6 allowed to cut the number of generations on the average to 12000 for solving optimization problem (8) and to 1200 for solving optimization problem (6). Reduction of the mutation ratio to 0.01 allowed to cut the number of generations to 10000 and 1000 for optimization problems (8) and (6), respectively. Criterion (8) takes the values of 10.1248 and 3.1274 before and after tuning, respectively. The $RMSE_{x_1}$ and $RMSE_{x_2}$ take the values of 0.7684 and 0.7219 before tuning; 0.2163 and 0.2044 after tuning, respectively.

The parameters of the fuzzy model after tuning are given in Tables 2, 3. The results of solving the problem of inverse inference after tuning are shown in Fig. 2. The same figure depicts the causes and effects membership functions after tuning.

Table 2. Fuzzy rules weights after tuning

w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9
0.93	0.97	0.92	0.95	1.00	0.97	0.92	0.96	0.93

Table 3. Parameters of the causes and effects membership functions after tuning

	C_1	C_2	C_3	C_4	C_5	C_6	E_1	E_2	E_3	E_4	E_5	E_6
β -	0.03	3.04	5.97	0.02	3.05	5.96	0.52	0.91	3.35	0.10	2.57	3.03
σ -	0.41	0.82	0.39	0.43	0.90	0.41	0.28	0.16	1.95	1.93	0.14	0.26

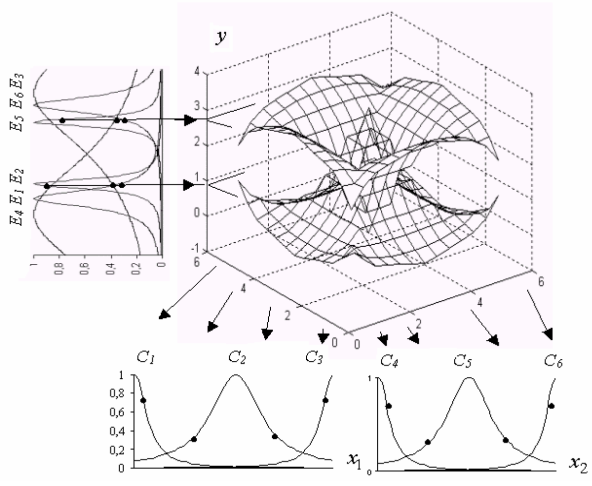


Fig. 2. Solution to the problem of inverse inference

Let the specific values of the output variables consists of y_1^*=0.95 and y_2^*=2.65. The degrees of membership of the outputs to the fuzzy terms $E_1 \div E_6$ for these values can be defined with the help of the membership functions in Fig.2

$\mu^{E_1}(y_1^*)$ =0.30; $\mu^{E_2}(y_1^*)$ =0.94; $\mu^{E_3}(y_1^*)$ =0.40; $\mu^{E_4}(y_2^*)$ =0.36; $\mu^{E_5}(y_2^*)$ =0.75; $\mu^{E_6}(y_2^*)$ =0.32. Taking into account the weights of rules (Table 2), the vector of the effects combinations significances measures takes the following form:

$$\boldsymbol{\mu}^B(Y^*) = (\mu^{B_1} = 0.70, \mu^{B_2} = 0.29, \mu^{B_3} = 0.70, \mu^{B_4} = 0.29, \mu^{B_5} = 0.36,$$
$$\mu^{B_6} = 0.29, \mu^{B_7} = 0.70, \mu^{B_8} = 0.29, \mu^{B_9} = 0.70).$$

The null solution was obtained with the help of the genetic algorithm

$$\boldsymbol{\mu}_0^C = (\mu_0^{C_1} = 0.9, \mu_0^{C_2} = 0.3, \mu_0^{C_3} = 0.8, \mu_0^{C_4} = 0.7, \mu_0^{C_5} = 0.3, \mu_0^{C_6} = 0.7),$$

for which the modified fuzzy effects vector corresponds

$$\boldsymbol{\mu}_0^B = (\mu_0^{B_1} = 0.7, \mu_0^{B_2} = 0.3, \mu_0^{B_3} = 0.7, \mu_0^{B_4} = 0.3, \mu_0^{B_5} = 0.3,$$
$$\mu_0^{B_6} = 0.3, \mu_0^{B_7} = 0.7, \mu_0^{B_8} = 0.3, \mu_0^{B_9} = 0.7).$$

The optimization criterion (6) takes the value of F=0.0040.

This modified vector allows us to use the implemented in MATLAB Fuzzy Relational Calculus Toolbox [9] for finding the solution set $S(\boldsymbol{\mu}_0^B)$. Using the standard solver ***solve_flse*** [9] we obtain the following results. The solution set $S(\boldsymbol{\mu}_0^B)$ for the modified vector $\boldsymbol{\mu}_0^B$ is completely determined by the unique minimal solution

$$\underline{\boldsymbol{\mu}}^C = (\underline{\mu}^{C_1} = 0.7, \underline{\mu}^{C_2} = 0.3, \underline{\mu}^{C_3} = 0.7, \underline{\mu}^{C_4} = 0.7, \underline{\mu}^{C_5} = 0.3, \underline{\mu}^{C_6} = 0.7)$$

and the two maximal solutions $S^* = \{\overline{\boldsymbol{\mu}}_1^C, \overline{\boldsymbol{\mu}}_2^C\}$

$$\overline{\boldsymbol{\mu}}_1^C = (\overline{\mu}_1^{C_1} = 0.7, \overline{\mu}_1^{C_2} = 0.3, \overline{\mu}_1^{C_3} = 0.7, \overline{\mu}_1^{C_4} = 1.0, \overline{\mu}_1^{C_5} = 0.3, \overline{\mu}_1^{C_6} = 1.0);$$
$$\overline{\boldsymbol{\mu}}_2^C = (\overline{\mu}_2^{C_1} = 1.0, \overline{\mu}_2^{C_2} = 0.3, \overline{\mu}_2^{C_3} = 1.0, \overline{\mu}_2^{C_4} = 0.7, \overline{\mu}_2^{C_5} = 0.3, \overline{\mu}_2^{C_6} = 0.7).$$

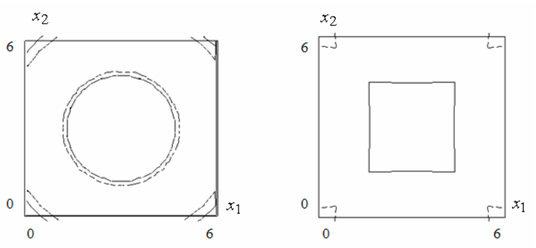
Thus the solution of the system (9) of fuzzy logical equations can be represented in the form of intervals:

$$\begin{gathered} S(\boldsymbol{\mu}^B) = \{\mu^{C_1} = 0.7, \mu^{C_2} = 0.3, \mu^{C_3} = 0.7, \mu^{C_4} \in [0.7, 1.0], \\ \mu^{C_5} = 0.3, \mu^{C_6} \in [0.7, 1.0]\} \cup \\ \cup \{\mu^{C_1} \in [0.7, 1.0], \mu^{C_2} = 0.3, \mu^{C_3} \in [0.7, 1.0], \mu^{C_4} = 0.7, \\ \mu^{C_5} = 0.3, \mu^{C_6} = 0.7\}. \end{gathered} \tag{10}$$

The intervals of the values of the input variable for each interval in solution (10) can be defined with the help of the membership functions in Fig. 2:

$$x_1^* = 0.3 \text{ or } x_1^* \in [0, 0.3] \text{ for } C_1;$$
$$x_1^* = 1.8 \text{ or } x_1^* = 4.3 \text{ for } C_2;$$
$$x_1^* = 5.7 \text{ or } x_1^* \in [5.7, 6] \text{ for } C_3;$$
$$x_2^* \in [0, 0.3] \text{ or } x_2^* = 0.3 \text{ for } C_4;$$
$$x_2^* = 1.7 \text{ or } x_2^* = 4.4 \text{ for } C_5;$$
$$x_2^* \in [5.7, 6] \text{ or } x_2^* = 5.7 \text{ for } C_6.$$

The restoration of the input set for $y_1^*=0.95$ and $y_2^*=2.65$ is shown in Fig. 2. The values of the membership degrees of the inputs to fuzzy terms $C_1 \div C_6$ are marked. The comparison of the target and restored level lines for $y_1^*=0.95$ and $y_2^*=2.65$ is shown in Fig. 3.



a) b)

Fig. 3. Comparison of the target (a) and restored (b) level lines for $y_1^*=0.95$ (___) and $y_2^*=2.65$ (_ _ _)

6 Example of Technical Diagnosis

We shall consider faults causes diagnosis of the hydraulic elevator (for dump truck body, excavator ladle etc.). Input parameters of the hydro elevator are (variation ranges are indicated in parentheses): x_1 – engine speed (30–50 r/s); x_2 – inlet pressure (0.02–0.15 kg/cm^2); x_3 – clearance of the feed change gear (0.1–0.3 mm); x_4 – oil leakage (0.5–2.0 cm^3/min). Output parameters of the elevator are: y_1 – productivity (17–22 l/min); y_2 – force main pressure (13–24 kg/cm^2); y_3 – consumed power (2.1–3.0 kw); y_4 – suction conduit pressure (0.5–1 kg/cm^2).

Fuzzy knowledge base is presented in Table 4, where the total number of the causes and effects consists of: C_1 *Decrease* (*D*), C_2 *Increase* (*I*) for x_1; C_3 (*D*), C_4 (*I*) for x_2; C_5 (*D*), C_6 (*I*) for x_3; C_7 (*D*), C_8 (*I*) for x_4; E_1 (*D*), E_2 (*I*) for y_1; E_3 (*D*), E_4 (*I*) for y_2; E_5 (*D*), E_6 (*I*) for y_3; E_7 (*D*), E_8 (*I*) for y_4.

Table 4. Fuzzy knowledge base for hydro elevator diagnosis

Rule	x_1	x_2	x_3	x_4	y_1	y_2	y_3	y_4
1	D	D	D	I	D	D	D	I
2	D	I	D	D	D	D	I	D
3	I	D	D	I	D	I	D	I
4	I	D	D	D	I	I	D	D
5	I	I	D	I	D	I	I	D
6	I	I	I	D	I	D	I	I
7	I	I	I	I	D	D	I	I

Fuzzy logical equations take the following form:

$$\begin{aligned}
\mu^{C_1} \wedge \mu^{C_3} \wedge \mu^{C_5} \wedge \mu^{C_8} &= w_1 \cdot (\mu^{E_1} \wedge \mu^{E_3} \wedge \mu^{E_5} \wedge \mu^{E_8}) \\
\mu^{C_1} \wedge \mu^{C_4} \wedge \mu^{C_5} \wedge \mu^{C_7} &= w_2 \cdot (\mu^{E_1} \wedge \mu^{E_3} \wedge \mu^{E_6} \wedge \mu^{E_7}) \\
\mu^{C_2} \wedge \mu^{C_3} \wedge \mu^{C_5} \wedge \mu^{C_8} &= w_3 \cdot (\mu^{E_1} \wedge \mu^{E_4} \wedge \mu^{E_5} \wedge \mu^{E_8}) \\
\mu^{C_2} \wedge \mu^{C_3} \wedge \mu^{C_5} \wedge \mu^{C_7} &= w_4 \cdot (\mu^{E_2} \wedge \mu^{E_4} \wedge \mu^{E_5} \wedge \mu^{E_7})\,. \\
\mu^{C_2} \wedge \mu^{C_4} \wedge \mu^{C_5} \wedge \mu^{C_8} &= w_5 \cdot (\mu^{E_1} \wedge \mu^{E_4} \wedge \mu^{E_6} \wedge \mu^{E_7}) \\
\mu^{C_2} \wedge \mu^{C_4} \wedge \mu^{C_6} \wedge \mu^{C_7} &= w_6 \cdot (\mu^{E_2} \wedge \mu^{E_3} \wedge \mu^{E_6} \wedge \mu^{E_8}) \\
\mu^{C_2} \wedge \mu^{C_4} \wedge \mu^{C_6} \wedge \mu^{C_8} &= w_7 \cdot (\mu^{E_1} \wedge \mu^{E_3} \wedge \mu^{E_6} \wedge \mu^{E_8})
\end{aligned} \tag{11}$$

For the fuzzy model tuning we used the results of diagnosis for 220 hydro elevators. The training set was used to evaluate criterion (8) during the evolutionary optimization of the fuzzy model parameters. To test the fuzzy model we used the results of diagnosis for 210 elevators with different kinds of faults. The goal was to identify the possible fault causes and evaluate the average percentage of correct diagnosis. In each run, the value of criterion (8) for the training data and the accuracy rate for the testing data were evaluated. The parameters of the main genetic operations were experimentally selected. It was determined that population size of V =10 is sufficient for solving optimization problems (8) and (6). In the experiments, crossover and mutation ratios were set to 0.7 and 0.01, respectively. As the number of generations increased, the value of criterion (8) decreased, starting with the value of 12.1783 before tuning and finally converging to 4.5516 after 10000 iterations of the genetic algorithm. The fault causes diagnosis started with an average accuracy of 80% and obtained an accuracy rate of 96% after 10000 iterations of the genetic algorithm (200 min on Celeron 700).

The results of the fuzzy model tuning are given in Tables 5-7 and Fig. 4.

Table 5. Fuzzy rules weights after tuning

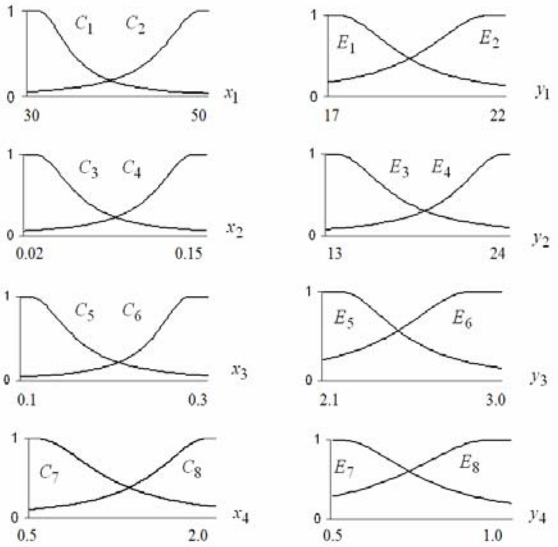
w_1	w_2	w_3	w_4	w_5	w_6	w_7
0.80	0.65	0.99	0.95	0.98	0.92	0.53

Table 6. Parameters of the membership functions for the causes after tuning

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8
β	31.20	49.15	0.03	0.14	0.11	0.28	0.55	1.94
σ	3.87	4.82	0.03	0.03	0.05	0.04	0.59	0.50

Table 7. Parameters of the membership functions for the effects after tuning

	E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8
β	17.27	21.52	13.50	22.65	2.19	2.84	0.52	0.94
σ	1.88	2.10	3.22	2.84	0.33	0.41	0.24	0.28



a) b)

Fig. 4. Membership functions of the causes (a) and effects (b) fuzzy terms after tuning

Let us represent the vector of the observed parameters for a specific elevator: $\boldsymbol{Y}^*$=(y_1^*=18 l/min; y_2^*=21.5 kg/cm^2; y_3^*=2.35 kw; y_4^*=0.8 kg/cm^2).

The degrees of membership of the outputs to the effects $E_1 \div E_8$ for these values can be defined with the help of the membership functions in Fig. 4,b: $\mu^{E_1}(y_1^*)$=0.87; $\mu^{E_2}(y_1^*)$=0.26; $\mu^{E_3}(y_2^*)$=0.14; $\mu^{E_4}(y_2^*)$=0.86; $\mu^{E_5}(y_3^*)$=0.81; $\mu^{E_6}(y_3^*)$=0.41; $\mu^{E_7}(y_4^*)$=0.42; $\mu^{E_8}(y_4^*)$=0.80. Taking into account the weights of rules (Table 5), the vector of the effects combinations significances measures takes the following form:

$$\boldsymbol{\mu}^{B}(Y^{*})=(\mu^{B_1}=0.11,\mu^{B_2}=0.09,\mu^{B_3}=0.80,\mu^{B_4}=0.25,$$
$$\mu^{B_5}=0.40,\mu^{B_6}=0.13,\mu^{B_7}=0.07).$$

The null solution was obtained with the help of the genetic algorithm

$$\boldsymbol{\mu}_0^{C}=(\mu_0^{C_1}=0.1,\mu_0^{C_2}=0.9,\mu_0^{C_3}=1.0,\mu_0^{C_4}=0.4,\mu_0^{C_5}=0.8,$$
$$\mu_0^{C_6}=0.1,\mu_0^{C_7}=0.25,\mu_0^{C_8}=0.9),$$

for which the modified fuzzy effects vector corresponds

$$\boldsymbol{\mu}_0^{B}=(\mu_0^{B_1}=0.1,\mu_0^{B_2}=0.1,\mu_0^{B_3}=0.8,\mu_0^{B_4}=0.25,$$
$$\mu_0^{B_5}=0.4,\mu_0^{B_6}=0.1,\mu_0^{B_7}=0.1).$$

The optimization criterion (6) takes the value of F=0.0020.

Using the standard solver ***solve_flse*** [9] we obtain the following results. The solution set $S(\boldsymbol{\mu}_0^{B})$ for the modified vector $\boldsymbol{\mu}_0^{B}$ is completely determined by the unique minimal solution

$$\underline{\boldsymbol{\mu}}^{C}=(\underline{\mu}^{C_1}=0.1,\underline{\mu}^{C_2}=0.8,\underline{\mu}^{C_3}=0.8,\underline{\mu}^{C_4}=0.4,\underline{\mu}^{C_5}=0.8,$$
$$\underline{\mu}^{C_6}=0.1,\underline{\mu}^{C_7}=0.25,\underline{\mu}^{C_8}=0.8)$$

and the four maximal solutions $S^{*}=\{\overline{\boldsymbol{\mu}}_1^{C},\overline{\boldsymbol{\mu}}_2^{C},\overline{\boldsymbol{\mu}}_3^{C},\overline{\boldsymbol{\mu}}_4^{C}\}$

$$\overline{\boldsymbol{\mu}}_1^{C}=(\overline{\mu}_1^{C_1}=0.1,\overline{\mu}_1^{C_2}=0.8,\overline{\mu}_1^{C_3}=1.0,\overline{\mu}_1^{C_4}=0.4,\overline{\mu}_1^{C_5}=1.0,$$
$$\overline{\mu}_1^{C_6}=0.1,\overline{\mu}_1^{C_7}=0.25,\overline{\mu}_1^{C_8}=1.0)$$
$$\overline{\boldsymbol{\mu}}_2^{C}=(\overline{\mu}_2^{C_1}=0.1,\overline{\mu}_2^{C_2}=1.0,\overline{\mu}_2^{C_3}=0.8,\overline{\mu}_2^{C_4}=0.4,\overline{\mu}_2^{C_5}=1.0,$$
$$\overline{\mu}_2^{C_6}=0.1,\overline{\mu}_2^{C_7}=0.25,\overline{\mu}_2^{C_8}=1.0)$$
$$\overline{\boldsymbol{\mu}}_3^{C}=(\overline{\mu}_3^{C_1}=0.1,\overline{\mu}_3^{C_2}=1.0,\overline{\mu}_3^{C_3}=1.0,\overline{\mu}_3^{C_4}=0.4,\overline{\mu}_3^{C_5}=0.8,$$
$$\overline{\mu}_3^{C_6}=0.1,\overline{\mu}_3^{C_7}=0.25,\overline{\mu}_3^{C_8}=1.0)$$
$$\overline{\boldsymbol{\mu}}_4^{C}=(\overline{\mu}_4^{C_1}=0.1,\overline{\mu}_4^{C_2}=1.0,\overline{\mu}_4^{C_3}=1.0,\overline{\mu}_4^{C_4}=0.4,\overline{\mu}_4^{C_5}=1.0,$$
$$\overline{\mu}_4^{C_6}=0.1,\overline{\mu}_4^{C_7}=0.25,\overline{\mu}_4^{C_8}=0.8).$$

Thus the solution of the system (11) of fuzzy logical equations can be represented in the form of intervals

$$
\begin{aligned}
S(\boldsymbol{\mu}^B)=\{&\mu^{C_1}=0.1,\mu^{C_2}=0.8,\mu^{C_3}\in[0.8,1.0],\mu^{C_4}=0.4,\mu^{C_5}\in[0.8,1.0],\\
&\mu^{C_6}=0.1,\mu^{C_7}=0.25,\mu^{C_8}\in[0.8,1.0]\}\cup\\
\cup\{&\mu^{C_1}=0.1,\mu^{C_2}\in[0.8,1.0],\mu^{C_3}=0.8,\mu^{C_4}=0.4,\mu^{C_5}\in[0.8,1.0],\\
&\mu^{C_6}=0.1,\mu^{C_7}=0.25,\mu^{C_8}\in[0.8,1.0]\}\cup\\
\cup\{&\mu^{C_1}=0.1,\mu^{C_2}\in[0.8,1.0],\mu^{C_3}\in[0.8,1.0],\mu^{C_4}=0.4,\mu^{C_5}=0.8,\\
&\mu^{C_6}=0.1,\mu^{C_7}=0.25,\mu^{C_8}\in[0.8,1.0]\}\cup\\
\cup\{&\mu^{C_1}=0.1,\mu^{C_2}\in[0.8,1.0],\mu^{C_3}\in[0.8,1.0],\mu^{C_4}=0.4,\mu^{C_5}\in[0.8,1.0],\\
&\mu^{C_6}=0.1,\mu^{C_7}=0.25,\mu^{C_8}=0.8\}.
\end{aligned}
\tag{12}
$$

Following the resulting solution (12), the causes C_2, C_3, C_5 and C_8 are the causes of the observed elevator state, so that $\mu^{C_2}>\mu^{C_1}$, $\mu^{C_3}>\mu^{C_4}$, $\mu^{C_5}>\mu^{C_6}$, $\mu^{C_8}>\mu^{C_7}$. The intervals of the values of the input variables for these causes can be defined with the help of the membership functions in Fig. 4,a: $x_1^* \in [47, 50]$ for C_2; $x_2^* \in [0.020, 0.043]$ for C_3; $x_3^* \in [0.100, 0.135]$ for C_5; $x_4^* \in [1.69, 2.00]$ for C_8. Thus, the causes of the observed elevator state should be located and identified as the increase of the engine speed to 47-50 r/s, the decrease of the inlet pressure to 0.02-0.04 kg/cm^2, the decrease of the feed change gear clearance to 100-135 mk, and the increase of the oil leakage to 1.69-2.00 cm^3/min. The tuning algorithm efficiency characteristics for the testing data are given in Table 8.

Table 8. Tuning algorithm efficiency characteristics

Cause (diagnose)	Number of cases in the data sample	Probability of the correct diagnose before tuning	Probability of the correct diagnose after tuning
C_1	56	47 / 56 = 0.84	54 / 56 = 0.96
C_2	154	125 / 154 = 0.81	147 / 154 = 0.95
C_3	100	76 / 100 = 0.76	98 / 100 = 0.98
C_4	110	80 / 110 = 0.72	105 / 110 = 0.95
C_5	167	132 / 167 = 0.79	162 / 167 = 0.97
C_6	43	38 / 43 = 0.88	41 / 43 = 0.95
C_7	92	74 / 92 = 0.80	89 / 92 = 0.97
C_8	118	98 / 118 = 0.83	115 / 118 = 0.97

7 Conclusions and Future Work

This paper proposes an approach for inverse problem solving based on the description of the interconnection between unobserved and observed parameters of an object with the help of fuzzy IF-THEN rules. The restoration and identification of the inputs through the observed outputs is accomplished by way of solving system of fuzzy logical equations, which correspond to IF-THEN rules, and tuning the fuzzy model on the readily available experimental data. The genetic algorithms are proposed for the optimization problems solving. The future work consists of the use of the genetic algorithm at the level of the initialization of the gradient-based learning schemes. Such an adaptive approach envisages the development of a hybrid genetic and neuro algorithms for the fuzzy rules based inverse problem solving. The approach proposed can find application not only in engineering but also in medicine, economics, military affairs and other domains, in which the necessity of interpreting the experimental observations arises.

References

1. Tihonov, N., Arsenin, V.Y.: Methods of solving incorrect problems. Science, Moscow (1974)
2. Abdelbar, A.M., Andrews, E., Wunsch, D.: Abductive reasoning with recurrent neural networks. Neural Netw. 16(5-6), 665–673 (2003)
3. Li, H.L., Kao, H.Y.: Constrained abductive reasoning with fuzzy parameters in Bayesian networks. Computers & Operations Research 32(1), 87–105 (2005)
4. Zadeh, L.: The concept of linguistic variable and it's application to approximate decision making. Mir, Moscow (1976)
5. Yager, R.R., Filev, D.P.: Essentials of fuzzy modelling and control. Wiley & Sons Inc., New York (1994)
6. Pedrycz, W.: Inverse problem in fuzzy relational equations. Fuzzy Sets and Systems 36(2), 277–291 (1990)
7. Di Nola, A., Sessa, S., Pedrycz, W., Sanchez, E.: Fuzzy relation equations and their applications to knowledge engineering. Kluwer Academic Press, Dordrecht (1989)
8. De Baets, B.: Analytical solution methods for fuzzy relational equations. In: Dubois, D., Prade, H. (eds.) Fundamentals of fuzzy sets. Kluwer Academic Publishers, Dordrecht (2000)
9. Peeva, K., Kyosev, Y.: Fuzzy relational calculus theory, applications and software. World Scientific, New York (2004), http://mathworks.net (accessed May 15, 2009)
10. Rotshtein, A., Rakytyanska, H.: Genetic algorithm of fuzzy logic equations solving in expert systems of diagnosis. In: Monostori, L., Váncza, J., Ali, M. (eds.) IEA/AIE 2001. LNCS (LNAI), vol. 2070, pp. 349–358. Springer, Heidelberg (2001)
11. Rotshtein, A., Posner, M., Rakytyanska, H.: Cause and effect analysis by fuzzy relational equations and a genetic algorithm. Reliability Engineering & System Safety 91(9), 1095–1101 (2006)
12. Rotshtein, A., Rakytyanska, H.: Diagnosis problem solving using fuzzy relations. IEEE Transactions on Fuzzy Systems 16(3), 664–675 (2008)
13. Rotshtein, A.: Design and tuning of fuzzy rule-based systems for medical diagnosis. In: Teodorescu, N.H., Kandel, A., Gain, L. (eds.) Fuzzy and neuro-fuzzy systems in medicine, pp. 50–54. CRC Press, London (1998)
14. Gen, M., Cheng, R.: Genetic algorithms and engineering design. Wiley & Sons Inc., New York (1997)

Necessary Optimality Conditions for Fractional Bio-economic Systems: The Problem of Optimal Harvest Rate

D.V. Filatova[1] and M. Grzywaczewski[2]

[1] Analytical Centre, Russian Academy of Sciences, Moscow, Russian Federation
daria_filatova@rambler.ru
[2] University of Technology, Radom, Poland
mgrzyw@interia.pl

Abstract. We consider the task of bioeconomic optimal control. The biomass dynamics is given by fractional stochastic differential equation. The discounted multiplicative production function describes net revenue and takes into account elasticity coefficients. Stochastic control problem is converting to non-random one. Necessary optimality conditions with respect to fractional terms are formulated as theorems and present the main result of this paper.

1 Introduction

Theory of system analysis allows classifying biological systems as reflexive one as far as they react on the changes of existence conditions, explicitly on environment actions and own states. In order to keep the completeness of the system under environmental variability and internal transformations considered biological system has to be in some dynamic equilibrium, which guarantees the existence of the entire system. The communities of endangered animals and plants are some of the biological systems examples, where human control factor plays very important role. In this case optimal control solution depends not only on a priori information about systems dynamic movement but also on its evolution and inter-connections with environment. This requires the control, which only corrects the system development and does not affect its natural behavior. Unfortunately solution of this problem is strongly connected with mathematical model selection.

Bio-economical models usually contain two main components. First component defines biological system description (usually one or more renewable resources) and second one characterizes the policy of this system exploitation [1]. The problem of *optimal* harvest rate, which is still widely studied [2-6], can be given as an example of this class models. So, in this case a renewable resource stock dynamics (or population growth) can be given as growth model of type

$$dX(t) = \left(1 - u(t)\right)\theta_1 X(t)\,dt\,, \tag{1}$$

Z.S. Hippe and J.L. Kulikowski (Eds.): Human-Computer Sys. Intera., AISC 60, pp. 557–567.
springerlink.com

here,
$X(t) \geq 0$ is size population at time t with given initial conditions $X(t_0) = x_0$,
$s(X(t))$ is a function, which describes population growth.

Model selection depends on the purpose of the modeling, characteristics of biological model and observed data [7]. Usually one takes $s(X(t)) = \theta_1 X(t)$ or $s(X(t)) = \theta_1 X(t)\left(1 - \frac{1}{K} X(t)\right)$, where $\theta_1 > 0$ is the intrinsic growth rate, $K > 0$ is the carrying capacity.

If in the first case population has unlimited growth, in the second case we can also show that the population biomass $X(t)$ will increase whenever $X(t) < K$, will decrease for $X(t) > K$ and is in a state of equilibrium if $X(t) \to K$ as $t \to \infty$. In the next reasoning we will use the model of unlimited growth.

Modification of the model (1) allows introducing of continuous harvesting at variable rate $u(t)$

$$dX(t) = (1 - u(t))\theta_1 X(t) dt \,. \tag{2}$$

It is clear that the harvest rate has to be controlled and constrained

$$0 \leq u(t) \leq u_{\max} \tag{3}$$

in order to guarantee the existence of the ecosystem under environmental variability and internal transformations.

An economical component of this bioeconomic model can be introduced as discounted value of utility function or production function (which may involve three types of input, namely labor $L(t)$, capital $C(t)$ and natural resources $X(t)$):

$$F(t, X(t), u(t)) = e^{-\delta t} \Pi\left(L^{\gamma_L}(t), C^{\gamma_C}(t), X^{\gamma}(t)\right), \tag{4}$$

where:
$\Pi\left(L^{\gamma_L}(t), C^{\gamma_C}(t) X^{\gamma}(t)\right)$ is the multiplicative Cobb-Douglas function with γ_L, γ_C and γ constants of elasticity, which correspond to the net revenue function at time t from having a resource stock of size $X(t)$ and harvest $u(t)$, δ is the annual discount rate (other production function models can be found, for an example in [2] or [8]).

Our optimal harvest problem on time interval $[t_0, t_1]$ becomes

$$\max_u \int_{t_0}^{t_1} F(t, X(t), u(t)) dt \tag{5}$$

subject to (2) and (3).

Mentioned problem could be easily solved by means of maximum principle. Unfortunately this model specification does not take into account fluctuations of stocks of renewable resource as well as changes in the marine ecosystem, migrations or spawning patterns and thus it is very difficult to verify the optimality of the solution in real life. For that reason in order to take into account stochastic effect of different factors, the population growth has to be considered as fractal stochastic variable [9] with dynamics given by fractional stochastic differential equation

$$dX(t)=\left[(1-u(t))\theta_1 X(t)\right]dt+\theta_2 X(t)\omega(t)(dt)^H \tag{6}$$

here,

$(1-u_t)\theta_1 X(t)$ is the growth of the biomass,

$\theta_2 X(t)$ is the diffusion term,

$\omega(t)$ is Gaussian random variable $\mathcal{N}(0,1)$, and

$H\in\,]0,1[$ is the fractional differencing parameter (the term $\omega(t)(dt)^H$ is fractional white noise, but it can be also considered as fractional Brownian motion with self-similarity parameter H [10]).

When the population growth is stochastic, the objective of the management is to maximize the expected utility subject to (3) and (6), to be exact

$$\mathcal{J}\left(X^\gamma(t),u(t)\right)=\max_u\left[\mathbb{E}\int_{t_0}^{t_1} e^{-\delta t}\Pi\left(L^{\gamma_L}(t),C^{\gamma_C}(t),X^\gamma(t)\right)dt\right], \tag{7}$$

where $\mathbb{E}[\cdot]$ is mathematical expectation operator.

There are several approaches, which allow finding optimal control. First group operates in terms of stochastic control [11] and [12], second one is based on converting the task (7) to non-random fractional optimal control [13]. It is also possible to use system of moment equations instead of equation (6) as it was proposed in [14] and [15].

In this work we will use transformation to non-random task, having minded that optimal solution depends on value of elasticity coefficient. So, if $\gamma=1$, then cost function (7) does not contain any fractional term, otherwise, if $\gamma\in(0,1)$, then cost function is fractional. We will focus our attention on the last case in order to get necessary optimality conditions for the task (3), (6) – (7).

2 Some Required Transformations

To transform stochastic problem to non-random one we introduce new state variable

$$y(t)=\mathbb{E}\left[X^\gamma(t)\right] \tag{8}$$

Using fractional difference filter [19] rewrite equation (6) with respect to (8)

$$dy(t) = \gamma(1-u(t))\theta_1 y(t)\,dt + \frac{\gamma(\gamma-1)}{2}\theta_2^2 y(t)(dt)^{2H} \tag{9}$$

To get rid of fractional term $(dt)^{2H}$ and for the convenience of the results formulation we replace ordinary fractional differential equation (9) by integral one

$$y(t)-y(t_0) = \int_{t_0}^{t} \Phi_0(u(\tau))\,y(\tau)\,d\tau + \int_{0}^{t} \Phi y(\tau)(d\tau)^{2H}\,, \tag{10}$$

where: $\Phi_0(u(t)) = \gamma(1-u(t))\theta_1$, $\Phi = \frac{\gamma(\gamma-1)}{2}\theta_2^2$.

Following reasoning is strongly dependent on H value as far as it changes the role of integration with respect to fractional term, namely as in [16], [17], denoting the kernel by $\kappa(\tau)$, one has for $0 < H < 1/2$

$$\int_{t_0}^{t} \kappa(\tau)(d\tau)^{2H} = 2H\int_{t_0}^{t} (t-\tau)^{2H-1}\kappa(\tau)\,d\tau\,, \tag{11}$$

and for $1/2 < H < 1$

$$\int_{t_0}^{t} \kappa(\tau)(d\tau)^{2H} = H^2\left[\int_{t_0}^{t} (t-\tau)^{H-1}\kappa^{1/2}(\tau)\,d\tau\right]^2. \tag{12}$$

So, if $0 < H < 1/2$, then denoting $\beta = 2H$ equation (10) can be rewritten as

$$y(t)-y(t_0) = \int_{t_0}^{t} \Phi_0(u(\tau))\,y(\tau)\,d\tau + \int_{t_0}^{t} \frac{\beta}{(t-\tau)^{1-\beta}}\Phi y(\tau)\,d\tau\,, \tag{13}$$

for $1/2 < H < 1$ equation (10) takes a form

$$y(t)-y(t_0) = \int_{t_0}^{t} \Phi_0(u(\tau))\,y(\tau)\,d\tau + \left[H\int_{t_0}^{t} \frac{\sqrt{\Phi y(\tau)}}{(t-\tau)^{1-H}}\,d\tau\right]^2. \tag{14}$$

3 Necessary Optimality Conditions

3.1 Statement of the Problem

Let the time interval $-\infty < t_0 < t_1 < \infty$ ($t \in \mathbb{R}$) is fixed, $y \in \mathbb{R}$ is a state variable, $u \in \mathbb{R}$ is a control. We rewrite the cost function (7) as

$$\mathcal{J}(y(\cdot),u(\cdot)) = \max_{u}\left[\int_{t_0}^{t_1} F(t,y(t),u(t))\,dt\right], \tag{15}$$

it is subjected to two groups of constraints.

First group of constraints (3) presents inequality constrains, which can be written as

$$\varphi(t, y(t), u(t)) \leq 0, \tag{16}$$

where vector function $\varphi(t, y(t), u(t))$ of the dimension m and its derivatives with respect to y and u are continuous functions. We also assume that the gradients $\varphi_{iu}(t, y(t), u(t))$, $i \in I(t, y(t), u(t))$, of the active constraints are positively independent at each point $(t, y(t), u(t))$ such that $\varphi_i(t, y(t), u(t)) \leq 0$. Here $I(t, y(t), u(t)) = \{i \in \{1, ..., m\} | \varphi_i(t, y(t), u(t)) = 0\}$ is the set of active indexes at the point $(t, y(t), u(t))$. Second group of constraints is the equality constraints given as the object equation (13) for $0 < H < 1/2$ and as equation (14) for $1/2 < H < 1$.

Our goal is to get the necessary optimality conditions for the problem (13), (15) and (16) and for the problem (14) – (16), solving them by maximum principle [18].

3.2 Solution for $0 < H < 1/2$ Case

Let $(y(t), u(t))$ be an optimal process. Thus we consider the operator $\mathrm{P} : (y, u) \in C \times L_\infty \to C \times \mathbb{R}$, $\mathrm{P} : (y, u) \to (z, \xi)$, where

$$z(t) = y(t) - y(t_0) - \int_{t_0}^{t} \left[\Phi_0(u(\tau))\, y(\tau) + \frac{\beta \Phi\, y(\tau)}{(t-\tau)^{1-\beta}} \right] d\tau, \ \xi = y(t_0) - a.$$

This operator is correctly defined. In our task $\mathrm{P}(y, u) = 0$ is the equality constrain. Frechét derivate of operator P in point (y, u) exists and is an operator $\mathrm{P}'(y, u) \circ (\bar{y}, \bar{u}) = (\bar{z}, \bar{\xi})$, such that

$$\bar{z}(t) = \bar{y}(t) - \int_{t_0}^{t} \Big[\Phi_0(u(\tau))\, \bar{y}(\tau) + y(\tau)\, \Phi_{0u}(u(\tau))\, \bar{u}(\tau) + \frac{\beta}{(t-\tau)^{1-\beta}} \bar{y}(\tau)\, \Phi \Big] d\tau, \ \ \bar{\xi} = \bar{y}(t_0),$$

and $\mathrm{P}'(y, u) : C \times L_\infty \to C \times \mathbb{R}$ is bounded linear operator. We are interested in general form of linear functional, vanishing on the kernel of operator $\mathrm{P}'(y, u)$. To get it we introduce following theorem [18].

Theorem 1. *If $A:Y\to Z$ is limited linear operator with closed image, then common form of linear functional $\ell:Y\to\mathbb{R}$ disappearing on* $\ker A$*, i.e. $\ell(\ker A)=0$, is $\ell(y)=z^*AY$, where $z^*\in Z^*$ (Y,Z are Banach spaces).*
Application of this theorem gives following results

$$\ell(\bar{y},\bar{u})=c_0\bar{y}(t_0)+\int_{t_0}^{t_1}\bar{y}(t)d\mu(t)-\int_{t_0}^{t_1}\int_{t_0}^{t}\left(\bar{y}(\tau)\Phi_0(u(\tau))\right)d\tau d\mu(t)$$
$$-\int_{t_0}^{t_1}\int_{t_0}^{t}\left(y(\tau)\Phi_{0u}(u(\tau))\bar{u}(\tau)\right)d\tau d\mu(t)+\int_{t_0}^{t_1}\int_{t_0}^{t}\frac{\beta}{(t-\tau)^{1-\beta}}\Phi\bar{y}(\tau)d\tau d\mu(t)$$

where $d\mu$ is the measure of Lebesgue–Stieltjes on $[t_0,t_1]$, respectively $\mu(t)$ is function of limited variation on $[t_0,t_1]$, $c_0\in\mathbb{R}$.

We denote $[\mu](t)=\mu(t+0)-\mu(t-0)$ is a jump. If the measure $d\mu$ is given, then $[\mu](t)$ $\forall t$ are known, particularly $[\mu](t_0)$ and $[\mu](t_1)$. This means, that $\mu(t_0-0)$ and $\mu(t_1+0)$. We transform $\ell(\bar{y},\bar{u})$, using following formula [13]

$$\int_{t_0}^{t_1}\left\{\int_{t_0}^{t}A(t,\tau)d\tau\right\}d\mu(t)=\int_{t_0}^{t_1}\left\{\int_{\tau}^{t_1}A(t,\tau)d\mu(t)\right\}d\tau$$

with $A(t,\tau)=\bar{y}(\tau)\Phi_0(u(\tau))+y(\tau)\Phi_{0u}(u(\tau))\bar{u}(\tau)+\dfrac{\beta}{(t-\tau)^{1-\beta}}\Phi\bar{y}(\tau)$.

Thus, we obtain

$$\ell(\bar{y},\bar{u})=c_0\bar{y}(t_0)+\int_{t_0}^{t_1}\bar{y}(t)d\mu(t)-\int_{t_0}^{t_1}\bar{y}(t)\int_{t}^{t_1}\left\{\Phi_0(u(t))+\frac{\beta}{(\tau-t)^{1-\beta}}\Phi\right\}d\mu(\tau)dt$$
$$-\int_{t_0}^{t_1}\left[\bar{u}(t)\int_{t}^{t_1}y(t)\Phi_{0u}(u(t))d\mu(\tau)\right]dt.$$

Now we can write the Euler equation for optimal process $(y(t),u(t))$

$$-\alpha_0\int_{t_0}^{t_1}\left[F_y(t,y(t),u(t))\bar{y}(t)+F_u(t,y(t),u(t))\bar{u}(t)\right]dt+$$
$$+(\bar{y},\bar{u})+\lambda(\varphi_y\bar{y}+\varphi_u\bar{u})=0,\ \forall\bar{u}\in L_\infty,\ \bar{y}\in C,$$

where $\lambda\in L_\infty^*$, $\lambda=(\lambda_1,...,\lambda_m)$, λ_i is concentrated on set $\{t|\varphi_i(X(t),u(t))=0\}$ $\forall i$, additionally, $\alpha_0\ge 0$ and $\alpha_0+\|\lambda\|+|c_0|+\|\mu\|>0$.

Taking into account that transversality conditions are given as $\psi(t)=\int_t^{t_1} d\mu(\tau)=\int_t^{t_1}\sigma(\tau)d\tau$, (where $t>t_0$, $\sigma\in L_1$, $\psi(t_1)=0$) we formulate necessary optimality conditions for the problem (13), (15) and (16) as follows.

Theorem 2. *Let* $(y(t),u(t))$ *be the optimal process on the interval* $[t_0,t_1]$ *(where* $y(\cdot)\in C[t_0,t_1]$, $u(\cdot)\in L_\infty[t_0,t_1]$ *). Then there is a set of Lagrange multipliers* $(\alpha_0,\sigma(\cdot),\lambda(\cdot))$ *such that* α_0 *is the number,* $\sigma:[t_0,t_1]\to\mathbb{R}$ *is an integrable function,* $\lambda:[t_0,t_1]\to\mathbb{R}^m$ *is integrable function, and the following conditions are fulfilled:*

- *nonnegativity condition* $\alpha_0\geq 0$, $\lambda(t)\geq 0$ *on* $[t_0,t_1]$*;*
- *nontriviality condition* $\alpha_0+\int_{t_0}^{t_1}|\sigma(t)|dt+\int_{t_0}^{t_1}|\lambda(t)|dt>0$*;*
- *complementary* $\lambda(t)\cdot\varphi(y(t),u(t))=0$ *a.e. on* $[t_0,t_1]$*;*
- *adjoint equation*

$$\sigma(t)=\alpha_0F_y(t,y(t),u(t))-\lambda(t)\varphi_y(t,y(t),u(t))$$
$$+\Phi_0(u(t))\int_t^{t_1}\sigma(t)d\tau+\Phi\int_t^{t_1}\frac{\beta}{(\tau-t)^{1-\beta}}\sigma(t)d\tau.$$

- *the local maximum principle is*

$$y(t)\Phi_{0u}u(t)\int_t^{t_1}\sigma(\tau)d\tau-\alpha_0F_u(t,y(t),u(t))-\lambda(t)\varphi_u(y(t),u(t))=0\,.$$

3.3 Solution for $1/2<H<1$ Case

Now we consider the problem (14) – (16). Let us introduce a new state variable

$$z(t)=\int_{t_0}^{t}\frac{g(\tau,y(\tau))}{(t-\tau)^{1-H}}d\tau\,, \tag{17}$$

where $g(\tau,y(\tau))=H\sqrt{\Phi y(\tau)}$, then equation (14) can be rewritten as

$$y(t)=y(t_0)+\int_{t_0}^{t}f(\tau,y(\tau),u(\tau))d\tau+z^2(t)\,, \tag{18}$$

where $f(\tau,y(\tau),u(\tau))=\gamma(1-u(\tau))\theta_1y(\tau)$.

Thus we get the system of integral equations ($t \in \left[t_0, t_1\right]$)

$$y(t) = \Xi\left(z(t)\right) + \int_{t_0}^{t} f\left(\tau, y(\tau), u(\tau)\right) d\tau ,$$

$$z(t) = z(t_0) + \int_{t_0}^{t} \frac{g\left(\tau, y(\tau)\right)}{\left(t-\tau\right)^{1-H}} d\tau ,$$

where $\Xi\left(z(t)\right)$ is an arbitrary smooth function.

In this case the necessary optimality conditions can be formulated as follows.

Theorem 3. *Let* $\left(y(t), z(t), u(t)\right)$ *be the optimal process on the interval* $\left[t_0, t_1\right]$ *(where* $y(\cdot), z(\cdot) \in C[t_0, t_1]$*,* $u(\cdot) \in L_\infty[t_0, t_1]$ *). Then there is a set of Lagrange multipliers* $\left(\alpha_0, \sigma(\cdot), \lambda(\cdot)\right)$ *such that* α_0 *is the number,* $\psi(\cdot):[t_0, t_1] \to \mathbb{R}^*$ *is an absolutely continuous function,* $\lambda:[t_0, t_1] \to \mathbb{R}^m$ *is integrable function, and the following conditions are fulfilled:*

- *nonnegativity condition* $\alpha_0 \geq 0$*,* $\lambda(t) \geq 0$ *on* $\left[t_0, t_1\right]$*;*
- *nontriviality condition* $\alpha_0 + \int_{t_0}^{t_1} \left|\lambda(t)\right| dt > 0$ *;*
- *complementary* $\lambda(t) \cdot \varphi\left(y(t), u(t)\right) = 0$ *a.e. on* $\left[t_0, t_1\right]$*;*
- *adjoint equation*

$$-\psi'(t) = \psi(t) f_y\left(t, y(t), u(t)\right) + \alpha_0 F_y\left(t, y(t), u(t)\right)$$
$$- g_y\left(t, y(t)\right) \int_{t}^{t_1} \frac{\psi'(t)\Xi'\left(z(\tau)\right)}{\left(\tau - t\right)^{1-H}} d\tau = 0$$

- *transversality condition* $\psi(t_1) = 0$ *;*
- *the local maximum principle is*

$$\psi(t) f_u\left(t, y(t), u(t)\right) + \alpha_0 F_u\left(t, y(t), u(t)\right)$$
$$- \lambda(t) \varphi_u\left(t, y(t), u(t)\right) = 0.$$

3.4 Solution for *Limited Growth*

In this subsection we will take into account all previous results and notations and consider the limited growth model for biomass introducing the parameter $\theta_{12} = 1/K$ (it corresponds to the carrying capacity). This allows on rewriting (6) as

$$dX(t) = \left[\left(1 - u(t)\right)\theta_1 X(t)\left(1 - \theta_{12} X(t)\right)\right] dt + \theta_2 X(t) \omega(t) \left(dt\right)^H .$$

Taking here into account only the case, where $0 < H < 1/2$, the object equation (13) is

$$y(t) - y(t_0) = \int_{t_0}^{t} \Phi_0(u(\tau))(1 - \theta_{12} y(\tau)) y(\tau) d\tau + \int_{t_0}^{t} \frac{\beta}{(t-\tau)^{1-\beta}} \Phi d\tau \ .$$

Using the same reasoning as before and applying theorem 1 we get

$$\begin{aligned} \ell(\bar{y},\bar{u}) = c_0 \bar{y}(t_0) + \int_{t_0}^{t_1} \bar{y}(t) d\mu(t) \\ - \int_{t_0}^{t_1} \int_{t_0}^{t} \Big[\Phi_{0y}(u(\tau)) \bar{y}(\tau) - 2\Phi_{0y}(u(\tau)) \theta_{12} y(\tau) \bar{y}(\tau) \\ + y(\tau) \Phi_{0u}(u(\tau)) \bar{u}(\tau) - y^2(\tau) \Phi_{0u}(u(\tau)) \theta_{12} \bar{u}(\tau) \Big] d\tau d\mu(t) \\ - \int_{t_0}^{t_1} \int_{t_0}^{t} \frac{\beta}{(t-\tau)^{1-\beta}} \Phi_y \, \bar{y}(\tau) d\tau d\mu(t). \end{aligned}$$

After the measure transformation we have

$$\begin{aligned} \ell(\bar{y},\bar{u}) = c_0 \bar{y}(t_0) + \int_{t_0}^{t_1} \bar{y}(t) d\mu(t) \\ - \int_{t_0}^{t_1} \bar{y}(\tau) \int_{t}^{t_1} \left[\Phi_{0y}(u(\tau)) - 2\Phi_{0y}(u(\tau)) \theta_{12} y(\tau) + \frac{\beta}{(\tau - t)^{1-\beta}} \Phi_y \right] d\mu(\tau) dt \\ - \int_{t_0}^{t_1} \bar{u}(\tau) \left[\int_{t}^{t_1} \left(y(\tau) \Phi_{0u}(u(\tau)) - y^2(\tau) \Phi_{0u}(u(\tau)) \theta_{12} \right) \right] d\mu(\tau) dt. \end{aligned}$$

Thus, the Euler equation for optimal process $(y(t), u(t))$ now is

$$\begin{aligned} -\alpha_0 \int_{t_0}^{t_1} \Big[F_y(t, y(t), u(t)) \bar{y}(\tau) + F_y(t, y(t), u(t)) \bar{u}(\tau) \Big] dt \\ + \ell(\bar{y},\bar{u}) + \lambda(\phi_y \bar{y} + \phi_u \bar{u}) = 0. \end{aligned}$$

The results of Euler equation analysis with respect to $\bar{y} = 0$ and $\bar{u} = 0$ can be formulated as theorem 2 with only difference, namely the local maximum principle is

$$y(t) \Phi_{0u}(u(\tau)) \int_{t}^{t_1} \sigma(\tau) d\tau + \alpha_0 F_y(t, y(t), u(t)) - \lambda(t) \varphi_y(t, y(t), u(t)) = 0 \ .$$

(One can get optimal solution to the problem for limited growth for the $1/2 < H < 1$ case using the same reasoning as in subsection 3.3.)

4 Conclusion

In this work we studied stochastic harvest problem, where the production function was presented by multiplicative Cobb-Douglass model with elasticity coefficients and the population was described by stochastic logarithmic growth model with fractional white noise. This formulation could not be solved by classical methods and required some additional transformations. We used fractional filtration and got the integral object equation, which did not contain stochastic term. As a result stochastic optimization problem was changed to non-random one. Using maximum principle we got necessary optimality conditions, which can be used for numerical solution of the problem.

References

1. Clark, C.W.: Bio-economic modeling and resource management. In: Levin, S.A., Hallam, T.G., Gross, L.J. (eds.) Applied mathematical ecology, pp. 11–57. Springer, New York (1989)
2. Kugarajh, K., Sandal, L.K., Berge, G.: Implementing a stochastic bioeconomic model for the north-east arctic cod fishery. Journal of Bioeconomics 8(1), 35–53 (2006)
3. Alvarez, L.H.R., Koskela, E.: Optimal harvesting under resource stock and price uncertainity. J. Economic Dynamics and Control 31(7), 2461–2485 (2007)
4. Nostbakken, L.: Regime switching in a fishery with stochastic stock and price. Environmental Economics and Management 51(2), 231–241 (2006)
5. Kulmala, S., Laukkanen, M., Michielsens, C.: Reconciling economic and biological modeling of migratory fish stock: optimal management of the atlantic salmon fishery in the baltic sea. Ecological Economics 64, 716–728 (2008)
6. Gasca-Leyva, E., Hernandez, J.M., Veliov, V.M.: Optimal harvesting time in a size-heterogeneous population. Ecological Modeling 210(1-2), 161–168 (2008)
7. Drechsler, M., Grimm, V., Mysiak, J., Watzold, F.: Differences and similarities between ecological and economic models for biodiversity conservation. Ecological Economics 62(2), 232–241 (2007)
8. Gonzalez-Olivares, E., Saez, E., Stange, E., Szanto, I.: Topological description of a non-differentiable bio-economic models. Rocky Mountain J. Mathematics 35(4), 1133–1145 (2005)
9. Jumarie, G.: Stochastics of order n in biological systems applications to population dynamics, thermodynamics, nonequilibrium phase and complexity. J. Biological Systems 11(2), 113–137 (2003)
10. Mandelbrot, B.B., van Ness, J.W.: Fractional brownian motions, fractional noises and applications. SIAM Review 10(4), 422–437 (1968)
11. Yong, J., Zhou, X.Y.: Stochastic control: hamiltonian systems and HJB equations. Springer, Heidelberg (1999)
12. Biagini, F., Hu, Y., Oksendal, B., Sulem, A.: A stochastic maximum principle for the processes driven by fractional brownian motion. Stochastic Processes and their Applications 100(1-2), 233–253 (2002)
13. Jumarie, G.: Merton's model of optimal portfolio in a black-scholes market driven by a fractional brownian motion with short-range dependence insurance. Mathematics and Economics 37(3), 585–598 (2005)

14. Krishnarajaha, I., Cooka, A., Marionb, G., Gibsona, G.: Novel moment closure approximations in stochastic epidemics. Bulletin of Mathematical Biology 67, 855–873 (2005)
15. Lloyd, A.L.: Estimating variability in models for recurrent epidemics: assessing the use of moment closure techniques. Theoretical Population Biology 65, 49–65 (2004)
16. Tarasov, V.E.: Liouville and Bogoliubov equations with fractional derivatives. Modern Physics Letters B 21, 237–248 (2006)
17. Jumarie, G.: Lagrange mechanics of fractional order, Hamilton-Jacobi fractional PDE and Taylor's series of non-differentiable functions. Chaos, Solutions and Fractals 32, 969–987 (2007)
18. Milyutin, A.A., Dmitruk, A.V., Osmolovskii, N.P.: Maximum principle in optimal control. MGU, Moscow (2004) (in Russian)
19. Hosking, J.R.M.: Fractional differencing. Biometrica 68(1), 165–176 (1981)

Author Index